



ELSEVIER

Contents lists available at [ScienceDirect](#)

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych



Response-time evidence for mixed memory states in a sequential-presentation change-detection task



Robert M. Nosofsky^{a,*}, Chris Donkin^b

^a Indiana University, United States

^b University of New South Wales, Australia

ARTICLE INFO

Article history:

Accepted 11 November 2015

Keywords:

Change detection
Response-time models
Memory search
Visual memory

ABSTRACT

Response-time (RT) and choice-probability data were obtained in a rapid visual sequential-presentation change-detection task in which memory set size, study-test lag, and objective change probabilities were manipulated. False “change” judgments increased dramatically with increasing lag, consistent with the idea that study items with long lags were ejected from a discrete-slots buffer. Error RTs were nearly invariant with set size and lag, consistent with the idea that the errors were produced by a stimulus-independent guessing process. The patterns of error and RT data could not be explained in terms of encoding limitations, but were consistent with the hypothesis that long retention lags produced a zero-stimulus-information state that required guessing. Formal modeling of the change-detection RT and error data pointed toward a hybrid model of visual working memory. The hybrid model assumed mixed states involving a combination of memory and guessing, but with higher memory resolution for items with shorter retention lags. The work raises new questions concerning the nature of the memory representations that are produced across the closely related tasks of change detection and visual memory search.

© 2015 Elsevier Inc. All rights reserved.

* Corresponding author at: Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, United States.

E-mail address: nosofsky@indiana.edu (R.M. Nosofsky).

1. Introduction

Visual working memory (VWM) is the short-term memory system that maintains representations of visual inputs. The nature of VWM is often investigated by presenting observers with brief visual displays and then testing the observers' memory for objects from different locations of the displays. A fundamental result is that VWM is highly limited in its capacity: As the number of to-be-remembered objects increases, the observer's ability to remember individual objects within the display gets worse.

A longstanding theoretical debate concerns the underlying basis for this limited capacity. According to the "discrete-slots" view (e.g., Awh, Barton, & Vogel, 2007; Barton, Ester, & Awh, 2009; Cowan, 2001; Luck & Vogel, 1997, 2013; Pashler, 1988; Rouder et al., 2008; Vogel, Woodman, & Luck, 2006), VWM makes available some number of slots for storing to-be-remembered items. The slot-based memories are conceptualized as being all-or-none: When memory is probed, if the test item occupies one of the slots, then the observer can judge its presence with no loss in resolution, regardless of the number of other items in the set of to-be-remembered objects. By contrast, if the object has not been stored in one of the discrete slots, then there is a complete loss of resolution. That is, no information about the presented object remains and the observer is forced to guess with regard to its identity. Although there are various versions of the discrete-slots view, the key aspect that we emphasize in this article is that the process gives rise to *mixed states* of memory: responding is presumed to reflect a mixture of memory-based responding and "zero-stimulus-information" responding that requires guessing.

An alternative view proposes that VWM consists of a pool of resources that is allocated in continuous fashion through sharing of the resources (e.g., Alvarez & Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011; Bays & Husain, 2008; Keshvari, van den Berg, & Ma, 2013; Ma, Husain, & Bays, 2014; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004). Thus, if the number of memory items is small, then the observer can maintain high-resolution representations of all of them. By contrast, if a large number of items must be maintained, then the continuous sharing of resources leads the observer to have lower-resolution representations of the individual items. Again, there are wide varieties of shared-resources models. The key aspect that we emphasize in this article is that the process gives rise to *continuously varying* states of memory with at least partial information retained for all to-be-remembered objects. Except for very occasional spurious trials involving inattention, there is no "zero-information" state that requires a true guessing process.

Finally, a variety of hybrid models have also been proposed that combine key elements of the discrete-slots and shared-resources approaches (e.g., Cowan & Rouder, 2009; Sims, Jacobs, & Knil, 2012; Zhang & Luck, 2008). Furthermore, various forms of evidence suggest that item limits and resource limits may be mediated by separate neural and cognitive systems (e.g., Awh et al., 2007; Unsworth, Fukuda, Awh, & Vogel, 2014; Xu & Chun, 2006).

These different views have led to an extensive and influential debate about the properties of VWM. Perhaps the most dominant modern paradigms used for contrasting the theories are "continuous recall" paradigms (Wilken & Ma, 2004; Zhang & Luck, 2008). In these paradigms, observers attempt to recall and reproduce a continuous value from a probed location of a briefly studied visual display (e.g., the color of a square). The discrete-slots models predict that the distribution of reproduced values is a mixture of responses from VWM and components that are purely guessing. In contrast, strong versions of the shared resource models predict that a single-component distribution becomes more diffuse as VWM resources are increasingly taxed.

The continuous-recall paradigms have yielded important insights. However, there are also some limitations associated with the approach. One limitation is that the properties of the paradigm itself may interfere with the information stored in VWM. For example, in the typical version of the paradigm, the response is produced by indicating a location along a continuous-valued response device. Note that presentation of the device itself (in which all continuous values are simultaneously present) may be highly interfering of the original memory, thereby leading to underestimates of the amount of information that was immediately available when memory was probed (e.g., Souza, Rerko, Lin, &

Oberauer, 2014). A second limitation is that modern versions of shared-resources models make allowance for the possibility that the pool of resources is shared in highly uneven fashion across sets of to-be-remembered items (e.g., Fougnie, Suchow, & Alvarez, 2012; Sims et al., 2012; van den Berg et al., 2012; van den Berg, Awh, & Ma, 2014). In extreme cases, the memory distribution for an item may become so diffuse as to essentially mimic a guessing distribution. Under such conditions, it can be exceedingly difficult to tease apart mixed-state from shared-resources models (but see Rouder, Thiele, Province, Cusumano, & Cowan, 2015, for one novel approach).

In recent research, Donkin, Nosofsky, Gold, and Shiffrin (2013) pursued a new theoretical approach and set of associated methods for investigating the nature of VWM. The key idea was to develop and test formal computational models that account jointly for choice probabilities and complete distributions of *response times* (RTs) in VWM change-detection tasks (for related models and ideas, see Pearson, Raskevicius, Bays, Pertzov, & Husain, 2014; and Smith & Sewell, 2013). The change-detection tasks were of the form pioneered by Luck and Vogel (1997) and extended by Rouder et al. (2008). In particular, on each trial, subjects were presented with a display of to-be-remembered visual objects, with each object occupying a different location. Following the display, a single location was probed with a test object. The observer was required to respond as quickly as possible, while minimizing errors, whether the object in that location changed or stayed the same. Donkin et al. (2013) argued that the collection of RT data in these VWM change-detection tasks opened new windows into the underlying processes that are not revealed by accuracy and continuous-recall data alone. (For similar arguments and RT approaches that have been successfully adopted to investigate discrete and continuous processes in the domain of long-term recognition memory and cued recall, see, for example, Dube, Starns, Rotello, & Ratcliff, 2012; Nobel & Shiffrin, 2001; Rotello & Zeng, 2008; Starns, Ratcliff, & McKoon, 2012.)

Donkin et al. (2013) formalized and evaluated three main families of RT models of VWM change-detection. One family formalized the idea that responding in the tasks emerges from the all-or-none mixture of cognitive states posited by strong versions of the discrete-slots view. The second family was based on strong versions of the continuous, shared-resources view. A third, hybrid family, which combined assumptions from the two approaches, was also proposed and evaluated. We describe these models in some detail in the ensuing Section 5.1 of our article. The key intuition is that the models that propose a prominent role of mixed states predict that the observed RT distributions will reflect a mixture of processes with very different time courses: one type of process reflecting memory-based responding and the other process reflecting guessing. By contrast, the pure shared-resources models predict that a single memory-based process produces the RT distributions, with the parameters of that process varying in continuous fashion across different conditions of testing. The RT distributions predicted by the models are therefore different in character, allowing for diagnostic tests to tell the models apart. In brief, Donkin et al.'s model-fitting results provided strong evidence in favor of the role of mixed states in their VWM tasks, with pure guessing playing a prominent role. In addition, in cases in which the number of to-be-remembered objects was small, there was also clear evidence for continuous sharing of resources. Thus, the results pointed in the direction of hybrid models of VWM.

In the present research, our goal was to pursue this RT-based approach to contrasting models of VWM, with two major issues in mind. First, the paradigm tested by Donkin et al. (2013) used the typical procedure of VWM tasks in which the set of to-be-remembered objects was presented in simultaneous fashion for a brief period of time. Under such conditions, a reasonable concern is that the evidence for mixed states (i.e., the presence of a zero-information state) may reflect encoding processes rather than memory processes. For example, suppose that an observer believes that she would be unable to retain all elements of a visual display when memory set size is large. In such a case, the observer may decide to strategically focus attention on a subset of the items in a display and ignore remaining items. In such a situation, the ignored items would never have been fully encoded and would never have entered VWM in the first place.

As a potential approach to addressing this concern, in the present research we conducted change-detection experiments in which the to-be-remembered items were presented in a rapid visual sequence (rather than simultaneously). The procedure was similar to the classic *rapid serial visual presentation* (RSVP) methods used in past research (e.g., Chun & Potter, 1995; Shiffrin & Schneider, 1977). However, each memory-set item was presented in a separate location on the screen. A single location

was then probed and subjects were required to judge whether the object in that location changed or stayed the same. Memory-set size (2, 5 or 8 objects) was varied randomly across trials. Note that in this procedure, because the subjects would not know in advance the memory set size, it seems unlikely that they would adopt strategic encoding processes that involved purposely ignoring a subset of the display. (We acknowledge, however, that other types of involuntary encoding limitations may arise, and we address some possibilities later in our article.) Although other researchers have also sometimes tested sequential-presentation versions of visual working-memory tasks (e.g., [Endress & Potter, 2014](#); [Gorgoraptis, Catalao, Bays, & Husain, 2011](#); [Sewell, Lilburn, & Smith, 2014](#); [Shapiro & Miller, 2011](#); [Umamoto, Drew, Ester, & Awh, 2010](#); [Xu & Chun, 2006](#)), none had the present goal of testing modern models of RT in such tasks to distinguish between discrete-slots and shared-resources views.

The second main motivation for using the sequential-presentation change-detection task is that it allowed us to carefully examine the role of *lag* of presentation on VWM performance. Note that in the sequential-presentation design, for any given memory set size, some items in the display will have been presented recently, whereas others will have been presented in the more distant past. This potential influence of lag on performance could provide additional clues concerning mixed-state versus continuous models of VWM. For example, according to discrete-slots models, an item with long lag may have high probability of being “bumped” from a buffer of discrete slots, placing it into the zero-information state ([Atkinson & Shiffrin, 1968](#); [Donkin & Nosofsky, 2012a](#)). Note that similar forms of “covert lag” may also operate in simultaneous-display designs, if, for example, the observer attends in sequential fashion to individual items in the display. The present design has the advantage, however, of allowing one to “observe” the sequence of attended locations.

Indeed, the sequential-presentation change-detection task is closely related to classic memory-search paradigms ([Sternberg, 1966](#)), where lag of presentation is often observed to play a crucial role (e.g., [Donkin & Nosofsky, 2012b](#); [McElree & Doshier, 1989](#); [Monsell, 1978](#); [Nosofsky, Little, Donkin, & Fific, 2011](#)). In typical short-term memory-search tasks, an observer is presented with a sequential list of to-be-remembered items (i.e., the memory set). Following presentation of the list, a test probe is presented, and the observer judges whether the probe is “old” (a member of the memory set) or “new”. Note that the old/new judgment in the memory-search task is analogous to the same/change judgment that is made in a change-detection task. An important conceptual difference, however, is that in the change-detection task, the question is whether the test probe is the same as a single item in a given location; whereas in the memory-search task, the question is whether the test probe is the same as *any* item from the memory set. As is well known, in memory-search tasks, RTs get longer and accuracy declines as memory set-size increases. However, at least in conditions that discourage rehearsal and in which the objects are highly discriminable, the major variable that influences old-item RTs and accuracies appears to be lag, not memory set-size per se. In particular, holding lag fixed, there is generally little difference in RTs across different memory set sizes (e.g., [Donkin & Nosofsky, 2012b](#); [McElree & Doshier, 1989](#); [Monsell, 1978](#); [Nosofsky et al., 2011](#)). The reason why old-item RTs tend to be longer overall as set size increases is because larger memory sets contain items with greater lags.

Our sequential-presentation change-detection task allows even greater insights into the role of lag than does the traditional memory-search task. Note that in memory-search tasks, “lag” is clearly defined only for “old” test probes, not “new” ones. In particular, an old test probe occupies a particular serial position of the sequentially-presented memory set, so has an unambiguous lag. By contrast, by definition, a “new” test probe does not occupy any serial position of the list. In the case of our sequential-presentation change-detection task, however, lag is defined for both old (no-change) and new (change) test items. The reason is that the test probe is with respect to a single location of the visual display, and the object that was displayed in that location occurred in a specific order of the sequence. Thus, we can ask how both no-change (old) *and* change (new) trials are influenced by lag. As will be seen, the variable of lag has a profound influence on VWM change-detection performance and this influence provides additional constraints for contrasting discrete-slots and shared-resources models of change-detection RTs and accuracies.

2. Experiment

We tested six subjects in a sequential-presentation change-detection experiment. On each trial, subjects were presented with a rapid sequence of colored squares, with each square occupying a unique location within a rectangular region of the computer screen. Memory-set size was 2, 5 or 8, with the particular value chosen randomly across trials. Following the sequence of to-be-remembered items, a single location was probed with a test square. If the square in that location changed color, then the observer was supposed to respond “change,” and otherwise to respond “same”. We varied across blocks the prior probability (.3, .5, .7) with which change trials occurred and subjects were given careful instructions regarding the nature of this manipulation. In particular, subjects were asked to adjust their guessing strategies in accord with these manipulations of objective change probability. All individual colors were highly discriminable. Thus, if a location changed color, then it was a “big” change. Subjects were asked to make their responses as quickly as possible while minimizing errors.

3. Method

3.1. Subjects

Four of the six subjects were students from the University of New South Wales (UNSW) and two were undergraduate students from Indiana University Bloomington (IUB). All subjects completed 10 sessions of the task. Participants at UNSW were paid \$15 per session. The IUB undergraduates were paid \$12 per session (including a \$2 bonus per session for good performance). All subjects reported having normal color vision and none was aware of the issues under investigation in the study.

3.2. Stimuli

The stimuli were a set of 10 highly discriminable color squares (white, black, red, blue, green, yellow, orange, cyan, purple and dark-blue–green). Items were presented within a rectangular region whose visual angle was 9.8×7.3 deg. Each color square was 0.75×0.75 deg in size. Items were presented in randomly chosen locations within the rectangular region with the constraint that they had to be at least 2 deg away from any other item and from the center of the viewing area. Stimuli were presented on a grey background on 17-in. cathode ray tube monitors.

3.3. Procedure

There were slight differences in the procedures used for the UNSW subjects and the IUB subjects, so we describe them separately. (As noted in Section 4, these procedural differences did not lead to any notable changes in the patterns of empirical data and model-fitting results across the groups.) For the IUB subjects, on each trial, a fixation cross was presented for 500 ms in the center of the screen, followed by a 500 ms blank screen. A sequence of N unique study squares was then presented. Each square was displayed for 150 ms with a 100 ms inter-stimulus interval. Following the sequence, there was a 500-ms pause, and a circular cue was then presented for 200 ms to signal the location of the test probe. The test probe was then presented and remained on the screen until the subject responded. Following a 500 ms interval, feedback (“Correct” or “Incorrect”) was displayed on the screen for 1000 ms.

For the UNSW subjects, on each trial, a fixation cross was presented for 1000 ms in the center of the screen. A sequence of N colored squares were then presented, each for 150 ms and with a 100 ms blank screen between stimuli. Following the stimuli, there was a 500 ms blank screen, followed by a 500 ms presentation of a multi-colored mask in each of the N locations in which stimuli were presented on that trial. A circular cue was then presented at the same time as the test probe, which both remained on screen until a response was made. Feedback was provided to the participant for 1000 ms, followed by a blank screen for 100 ms to end the trial.

For both groups of subjects, memory set size ($N = 2, 5$ or 8) and the serial position of the probe were chosen randomly on each trial. The N unique study colors on each trial were chosen randomly from the complete set of 10. If the test probe was a new color, then it was chosen randomly from the remaining members of the set of 10 colors. (In other words, change trials always involved “external” changes, not “internal” ones in which a color from a different location of the array was used as a probe for the test location.) Subjects responded by pressing either the “F” key (same) or the “J” key (change).

Each session was organized into 9 blocks, with each block composed of either 54 trials at IUB or 60 trials at UNSW. Every three blocks, each of three change-probability conditions (.3, .5, .7) was tested once in a random order. On each trial, “change” versus “no-change” trials were determined randomly in accord with the objective change probability for that block. Following Rouder et al. (2008), we informed participants of the change-probability manipulation at the start of each block of trials. This information was provided by both telling the participant the percentage of upcoming trials on which the test item would change and by using a pie chart to illustrate the manipulation. Subjects were given explicit hints that they could increase their proportion of correct responses by tending to guess “change” on blocks with high objective change probability and by guessing “no change” on blocks with low objective change probability.

4. Results

We considered the results from the first three blocks of trials to be practice and we deleted those trials from the analysis. In addition, we deleted all trials in which the RT was less than 180 ms or greater than 2500 ms (less than 1% of the data). Visual inspection of the individual-subject data revealed generally similar patterns of responding for the six subjects.¹ Thus, in these initial analyses, we report the data averaged across subjects to provide information regarding the main trends in performance and to remove noise from the individual-subject results.

Following Rouder et al. (2008), in the following, we define a “hit” to occur when a subject responds “change” on a change trial. By contrast, a “false alarm” occurs when a subject responds “change” on a same trial. (Thus, a “miss” occurs when a subject fails to respond “change” on a change trial; whereas a “correct rejection” occurs when the subject responds “no change” on a same trial.)

Fig. 1 (top panel) plots the observed hit and false-alarm probabilities as a joint function of set size and lag. (These data are collapsed across the different objective change-probability conditions.) Perhaps the most dramatic result in our experiment is that subjects showed a precipitous increase in false alarms (responding change on same trials) with increases in lag. Indeed, by the time one reaches lag 4 in the set-size-8 condition, the average false-alarm rates exceed .50. There is a corresponding decrease in hit rates with increases in lag, although not nearly as dramatic as the increase in false alarms. All subjects showed this dramatic effect of lag and this strong asymmetry in false alarms versus hits.

For the very longest lag at set sizes 5 and 8, there is a drop in the false-alarm rate. This drop reflects a primacy effect: the longest lag in each set-size condition corresponds to the item in the first serial position of the sequence. Thus, the first item in each list seems to have a special status.

Once one takes account of the lag effects and the primacy effect, there appears to be little difference in results across the set-size 5 and 8 conditions: the hit and false-alarm curves across those two set-size conditions are nearly overlapping. However, the set-size-2 curve seems to have a special status: accuracy is greater in that condition compared to the other set-size conditions even after one conditions on lag. All six subjects showed this greater overall accuracy in the set-size-2 condition.

In our view, the precipitous increase in false alarms with increases in lag (coupled with the much smaller decrease in hits) seems most naturally interpreted in terms of mixed-state models of VWM. According to this interpretation, as new items enter the VWM system, previous items that occupied the discrete-slots buffer are ejected. If the observer is tested with a probe at an ejected location, she will be comparing the “presence” of some value with the absence of any stimulus information. Psychologically, this comparison corresponds to a form of change, so the natural response is to

¹ A notable exception is that S5 displayed markedly longer mean “miss” RTs as lag increased from 3 through 8, whereas for the other subjects the lag-RT functions were nearly flat.

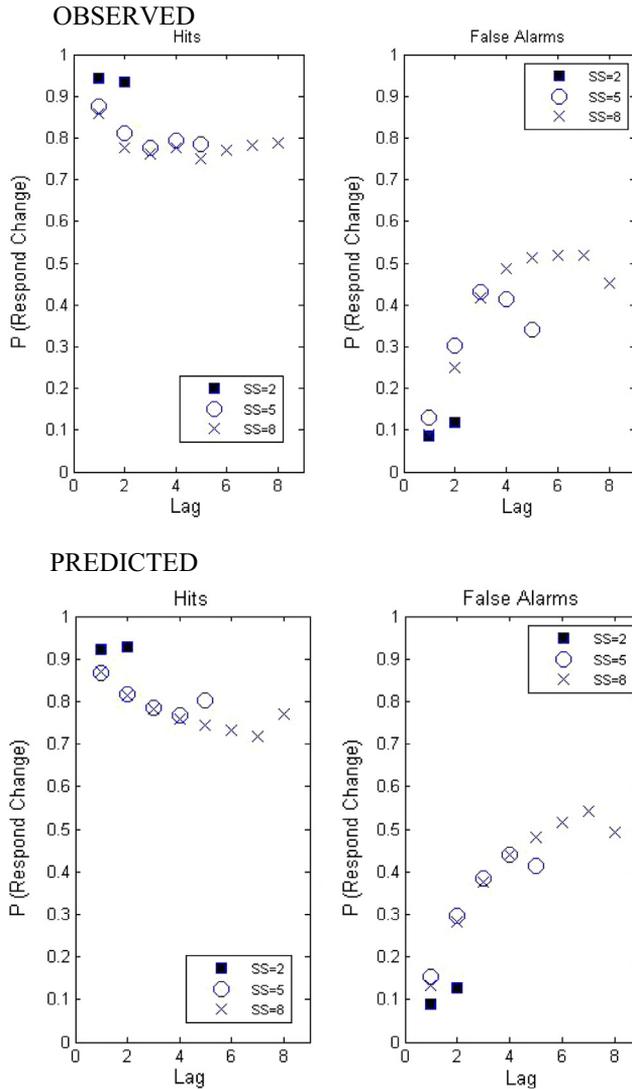


Fig. 1. Mean proportion of hits and false alarms plotted as a function of memory set size (SS) and study-test lag. Top panel: observed, bottom panel: predicted.

respond “change”. If the subject does indeed most often respond “change” on such trials, it would lead to a vast increase in false alarms (incorrectly responding “change” on trials in which the stimulus actually stayed the same). Hits would not show nearly as much decline with increasing lag: Although subjects would be guessing more with increasing lag, their guesses would tend to be “change” responses, leading them to make the correct response despite being in the zero-stimulus-information state. As will be seen, with appropriate choice of free parameters, the continuous models can also produce these general patterns of results. However, the strong asymmetry in hits and false alarms with increases in lag does not seem to be an *a priori* prediction of such models. Presumably, according to continuous models, as lag increases, memory gets “fuzzier” and performance will decline. The question, however, is why do the fuzzier memories lead to such a strong asymmetry towards responding

“change”? Regardless of these intuitive arguments, our key source of evidence in contrasting the mixed-state and continuous models will be provided in the ensuing Section 5.4.

Fig. 2 plots ROC curves that indicate how hits and false alarms varied as a joint function of set size and objective change probability. (There were insufficient data for a more detailed breakdown by lag.) As can be seen, subjects show lower sensitivity at larger set sizes. Across set sizes 5 and 8, the main reason for the lowered sensitivity is an increase in false alarms rather than a decrease in hits. Furthermore, both hits and false alarms increase with increases in objective change probability. That is, for both “change” and “no-change” trials, subjects respond “change” more often as objective change probability increases, suggesting a systematic shift in response criteria. This shift is in accord with the explicit instructions provided to subjects at the start of each change-probability block.

Evidence consistent with a criterion shift is also seen in Fig. 3, which plots how mean RTs for hits, misses, false alarms and correct rejections vary with objective change probability. As can be seen, as objective change probability increases, mean RTs for hits and false alarms (i.e., “change” responses) get shorter, whereas mean RTs for correct rejections and misses (i.e., “no-change” responses) get longer. We discuss many additional aspects of the Fig. 3 mean RT data in our Section 5.4.

The mean RTs for the different response types (hits, misses, false alarms and correct rejections) are plotted as a joint function of set size and lag in Fig. 4. (These data are collapsed across the different change-probability conditions.) Note that Donkin et al. (2013) had observed that in their simultaneous-presentation change-detection task, correct RTs (hits and correct rejections) got longer as set size increased. However, those data could not be conditioned on lag. In the present case, the main effect is that correct rejections get slower as lag increases (and all subjects showed this effect). Once one conditions on lag, there is little if any remaining effect of set size on the correct-rejection RTs. Furthermore, we were surprised to see that there was little effect of either set size or lag on the hit RTs. This latter result stands in contrast to the results from Donkin et al.’s simultaneous-presentation design, where hit RTs got significantly longer as memory set-size increased. The main explanation that Donkin et al. (2013) suggested for the longer hit and correct-rejection RTs in their study was that as set size increased, there was an increased probability that subjects would enter the zero-information state, so would be forced to guess. Furthermore, the guessing process operated more slowly than the memory-based processes for responding either “change” or “same”. In the present design, there may be an asymmetry, with the “guess-same” process operating more slowly than the memory-based respond-same process; but with the “guess-change” and memory-based respond-change processes operating at roughly the same speed. Thus, as the probability of guessing increases with increases in lag, there would be slowed correct-rejection responses, with little change in the hit RTs.

Regarding the error RTs (i.e., the RTs for false alarms and misses), the data are nearly flat across lags 3–8, and the RT functions for set sizes 5 and 8 are nearly overlapping. (The slight lengthening in miss RTs with lag is due almost completely to a single subject – see Footnote¹.) As discussed more fully in Section 5, the mixed-state models do indeed predict that the error-RT functions should be nearly invariant with set size and lag. The basic reasoning, as argued by Donkin et al. (2013), is that when stimulus changes are “big,” errors are almost always produced by the guessing process, not the memory-based process. Furthermore, the most natural assumption is that if the observer enters the zero-stimulus-information guessing state, then the speed of executing the guesses should be independent of stimulus-related factors such as set size and lag. The intuition is that the presentation of the test probe acts as a “gate” to one specific location of the stimulus display (see Section 5.1). If guessing is required at that specific location, then it seems reasonable to assume that the speed of the guessing process does not depend on factors such as the number of other locations that had been present in the display. Again, as will be seen, with appropriate choice of free parameters, the continuous models can fit the nearly invariant error RTs. However, this pattern does not seem to be an *a priori* prediction from that class of models. We will see that these multiple constraints appear to pose challenges for the purely continuous model, because its quantitative fits turn out to be substantially worse than those of the hybrid model for all six subjects.

Finally, the mean error RTs from lags 1 and 2 need to be interpreted with a good deal of caution, because subjects rarely made errors at these lags. Thus, the lag 1–2 error RTs are based on very small sample sizes. Furthermore, as discussed extensively by Donkin et al. (2013), a large proportion of the

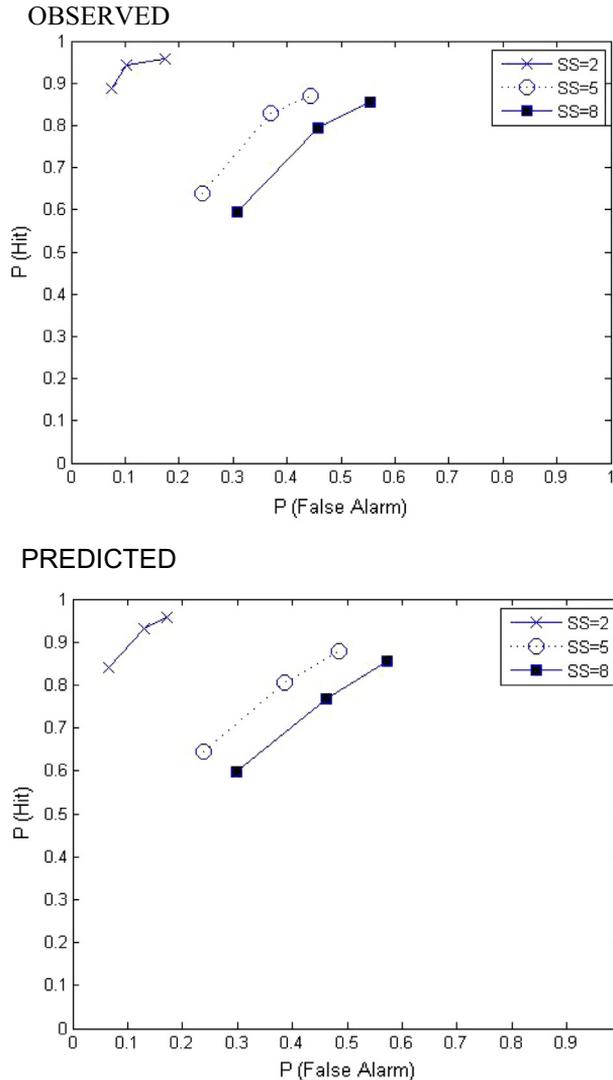


Fig. 2. Mean ROC curves for each memory set size (SS). Points within each ROC curve are for the different objective-change probability conditions. Top panel: observed, bottom panel: predicted.

observations in these cases may reflect processes that go outside the intended scope of the models (e.g., inadvertent responses based on inattention or motor-execution errors). Nevertheless, we include the data from all lag conditions in our ensuing modeling analyses.

5. Modeling analyses

5.1. The formal models

The models that we apply to the present data are extensions of the ones that Donkin et al. (2013) applied to their simultaneous-presentation design. We start by providing a conceptual outline of the models and then state their formal details.

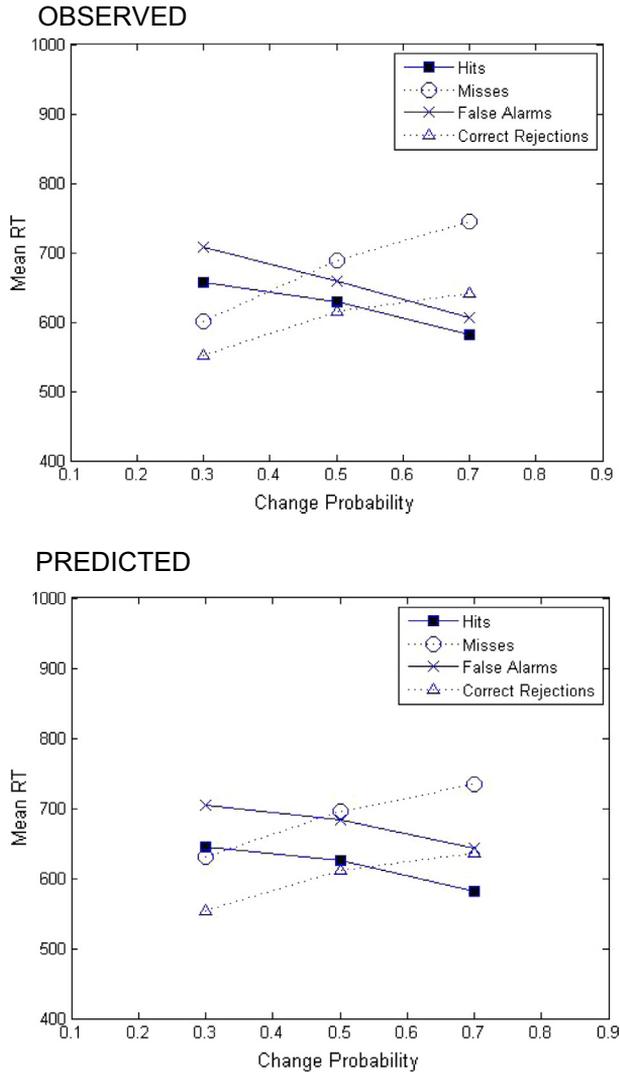


Fig. 3. Mean response times (RTs) for hits, misses, false alarms and correct rejections plotted as a function of objective change probability. Top panel: observed, bottom panel: predicted.

The basic idea underlying the class of discrete-slot (i.e., mixed-state) models is represented schematically in the top panel of Fig. 5. We presume that when a location from the study array is probed, there is an initial “gating” process that allows the system to determine whether memory-based information has been retained at that location (i.e., whether the study square is still in VWM). If so, a memory-based evidence-accumulation process operates that leads the observer to make a “change” or “no-change” decision. Furthermore, in the present big-change version of the paradigm, the memory-based accumulation process is presumed to operate with high accuracy. Alternatively, if the probed location is no longer in VWM, then a guessing-based accumulation process operates. In Donkin et al.’s (2013) previous applications to their simultaneous-presentation paradigm, the probability m that the subject entered the memory state (i.e., that the memory-based process operated) was presumed to depend only on memory set size. In particular, the greater the set size,

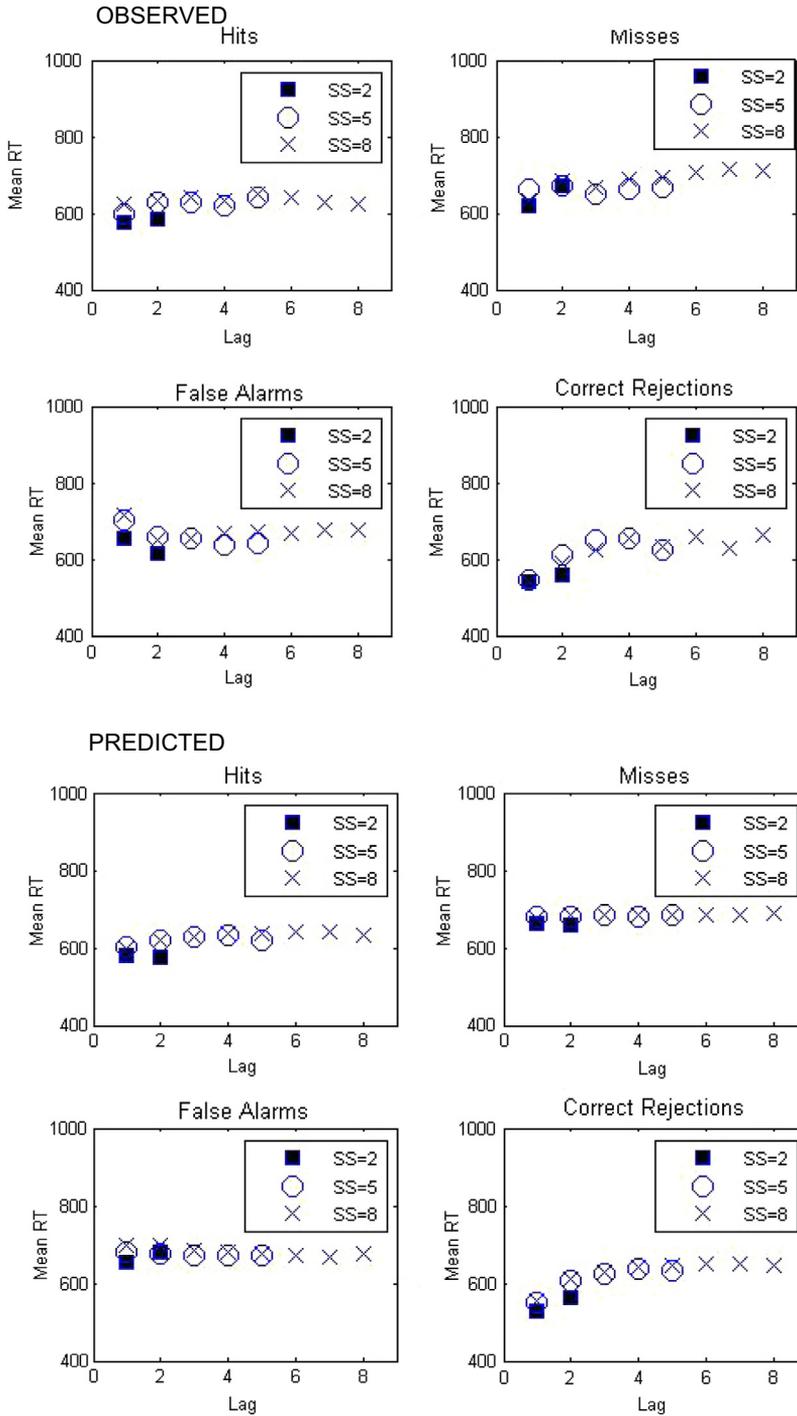


Fig. 4. Mean response times (RTs) for hits, misses, false alarms and correct rejections plotted as a function of memory set size (SS) and study-test lag. Top panel: observed, bottom panel: predicted.

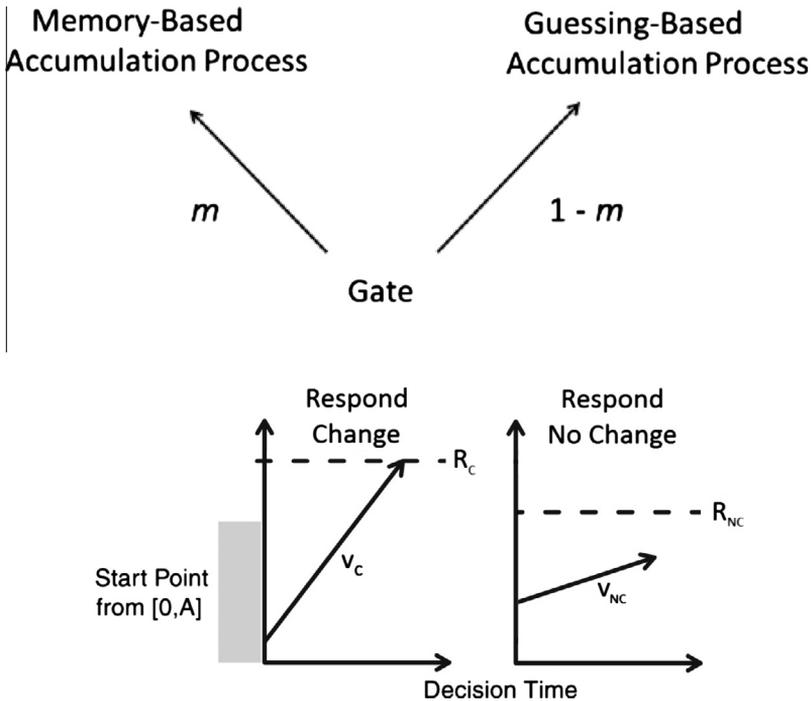


Fig. 5. Schematic illustration of components of the hybrid model of visual change detection. Top panel: gating process for memory-based versus guessing-based responding. Bottom panel: linear-ballistic-accumulator (LBA) model for the memory-based accumulation process.

the less likely that a probed item would be available in one of the discrete slots. In the present extension, this memory probability is presumed to depend jointly on set size and lag, with greater lags leading to a lower probability that the probed item remains in a discrete slot.

5.1.1. Linear-ballistic accumulator approach

The memory-based and guessing-based decision processes are both implemented by using a linear ballistic accumulator (LBA) approach (Brown & Heathcote, 2008). The LBA model is computationally simple and allows analytic computation of trial-by-trial likelihoods. Furthermore, research indicates that the LBA and the extremely influential diffusion model (e.g., Ratcliff, Van Zandt, & McKoon, 1999) mimic one another's predictions closely, and that parameters that share common psychological interpretation map closely onto each other (Donkin, Brown, Heathcote, & Wagenmakers, 2011).

To illustrate the approach, first consider the memory-based accumulation process. According to the LBA model, there are two accumulators, one for change responses and the second for no-change responses (Fig. 5, bottom panel). Evidence accumulates in linear and ballistic fashion (i.e., with no within-trial variability) until a response threshold is reached on either accumulator. Evidence accumulates with mean drift rate v_c on the change accumulator and with mean drift rate v_{NC} on the no-change accumulator. (If a stimulus change actually occurred, then v_c would tend to be large in magnitude and v_{NC} would tend to be small; and vice versa if no stimulus change occurred.) If the accumulating evidence first reaches the “change” response threshold (R_c), then the observer responds “change”; whereas if the evidence first reaches the “no change” response threshold (R_{NC}) then the observer responds “no change”. The decision time is determined by the time that it takes the accumulating evidence to reach the first threshold. Different versions of the model arise by varying assumptions about how drift rates and thresholds vary across conditions.

An analogous scheme operates for the guessing accumulation process (not pictured here), except in this case drift rates are not stimulus-dependent. Instead, one assumes some constant mean drift rate v_G in both a guess-change and guess-no-change accumulator. The observer is presumed to vary the response thresholds on these guessing accumulators in accord with experimental conditions. For example, in conditions in which change-probability is high, the observer would presumably set a low threshold on the “guess-change” accumulator, leading to more probable and faster “guess-change” responses. Although there are other approaches to modeling RTs and choices associated with guessing, it seemed most natural to us to develop a modeling approach that was analogous to what occurs on the memory-based accumulators.

The drift rate parameters v_C and v_{NC} on the memory-based accumulators reflect memory resolution: higher resolution leads to higher drift rate. Thus, in the pure mixed-state model, a fundamental assumption is that the mean drift rates on the memory-based accumulators are independent of memory set size and lag (because memory resolution is assumed to be independent of these factors). This assumption formalizes the “all-or-none” conception that is integral to the pure mixed-state view. Instead, the key parameters that vary across the different memory-set size and lag conditions are the memory-state probability parameters m (Fig. 5, top panel).

By contrast, in the pure shared-resources model, there are no guessing accumulators and *only* the memory-based accumulators operate. Instead, this model presumes that drift rates differ across stimulus conditions. These changing drift rates reflect the changes in memory resolution that are presumed to occur as memory set size and lag are varied. For example, the mean correct drift rates are large in magnitude when lag or memory set size is small.

Finally, the hybrid model combines these assumptions of the mixed-state and shared-resources models: Both the probability of entering the memory state (m) and the drift rates v_C and v_{NC} are presumed to vary across the different memory set size and lag conditions.

5.2. Detailed model specification

Our analytic strategy for contrasting the mixed-state, continuous shared-resources and hybrid models was to formulate descriptive versions that were reasonably general. If we instead imposed highly specific processing constraints on the formalized models, then the occurrence of poor fits might be attributed to inappropriate specific assumptions rather than to limitations of the general model class.

For example, although the mixed-state models are motivated most directly by a discrete-slots metaphor, there is a wide variety of different ways in which discrete-slots models can be formalized. One very strong approach is to assume that there is some fixed, invariant number of slots that are available, regardless of trial-by-trial fluctuations in attention, changes in memory set size, spatial layout of the study items, and so forth (e.g., Cowan, 2001; Luck & Vogel, 1997). We are neither endorsing nor arguing against such “fixed-slots” models in the present research; instead, our goal is to test more generally for the presence of the mixed memory states to which “slots” models are expected to give rise. Likewise, continuous models are most directly motivated by shared-resources metaphors, in which the precision of the memory representations gets gradually worse as set size increases. Some examples of strong versions of such models are those that impose constraints based on the sharing of some fixed sample size of units for coding the stimuli (e.g., Palmer, 1990; Smith & Sewell, 2013). We do not impose such constraints in our present formalizations, however, because our goal is to consider continuous models as a general class rather than to focus on some highly specific special case.

Although our modeling strategy was to develop general representatives of the mixed-state, continuous, and hybrid classes, the approach would become unwieldy without some reasonable simplifying assumptions. For example, as described earlier, within the framework of the continuous models, changes in memory resolution are modeled in terms of changes in drift rate on the LBA memory accumulators. Our goal is to generate predictions of performance as a joint function of the variables of memory set size, lag, objective change probability, and stimulus type (change versus same). If a separate free drift rate parameter were estimated for each and every unique combination of these factors, the number of free parameters would become prohibitive. To ensure our models are tractable,

therefore, we imposed what we view as reasonable functional constraints on the influence of these factors. For example, except for a residual primacy effect, it is almost certainly the case that memory storage probability and/or memory resolution will be a decreasing function of lag. Thus, rather than estimating separate memory-storage probabilities and drift rates for each and every combination of set size and lag, we use descriptive quantitative functions to model the joint effects of these factors in order to reduce the number of free parameters.

In the remainder of this section we state the specific manner in which the models are formalized and provide a listing of the models' free parameters (see Table 1 glossary). We start by describing the hybrid model. The mixed-state and shared-resources models are both special cases of the hybrid model that are produced by restricting some of its free parameters. Although all models use a large number of free parameters, we will see that the models are required to account for extremely challenging sets of RT-distribution and choice-probability data.

5.2.1. Hybrid model

In the hybrid model, the probability that the memory-accumulation process operates on a given trial is a joint function of memory set size (ss) and the lag of the probed location, given by

$$m(ss, lag) = \alpha(ss) \cdot lag^{-\beta} \cdot prim, \quad (1)$$

where $\alpha(ss)$ is a factor associated with set size; and β describes the rate at which the memory probability decreases with lag. The factor $prim$ is a “primacy boost” factor: its value is set at $prim\text{-boost-}m$ for items in the first serial position of the sequentially presented list, and is set at one for all other items. Because m is a probability, it is bounded between 0 and 1.

The mean correct drift rate on the memory-based change accumulator is also a joint function of set size and lag:

$$v_C(ss, lag) = \kappa_c \cdot v(ss) \cdot lag^{-\gamma(C)} \cdot prim, \quad (2)$$

where κ_c is a scaling constant associated with drift rates on the change accumulator; $v(ss)$ is a factor associated with set size; and $\gamma(C)$ is a drift-rate decay parameter for the change accumulator. Again, the factor $prim$ is a “primacy boost” factor. Its value is set at $prim\text{-boost-}v$ for items in the first serial position of the list and set at one otherwise. Analogously, the mean correct drift rate on the memory-based no-change accumulator is given by:

$$v_{NC}(ss, lag) = \kappa_{NC} \cdot v(ss) \cdot lag^{-\gamma(NC)} \cdot prim. \quad (3)$$

Note that Eqs. (1)–(3) assume a power function relation between memory strength and lag (or between memory-storage probability and lag). This assumption provides one example of our use of simple quantitative functions to develop tractable versions of the models.² In the present case, the power-function assumption is motivated by formal-modeling evidence obtained by Donkin and Nosofsky (2012a) in closely related probe-recognition tasks, and by past suggestions that forgetting can be well described as a power function of the retention interval (e.g., Anderson & Schooler, 1991; Wixted & Ebbesen, 1991). The crucial point, however, is that the power function is a reasonably flexible quantitative function for capturing the idea that memory storage probability and memory strength decrease with lag. We think it unlikely that any of our main conclusions would change if some of the major alternative descriptive functions of forgetting were assumed. Indeed, in Appendix A, we report an entirely parallel set of results and conclusions from alternative modeling analyses that assume that the forgetting functions are negative exponential in form.

The LBA model estimates a between-trials drift-rate variability parameter. (Drift rates are assumed to be normally distributed and the drift-rate variability parameter gives the standard deviation of the

² Another example involves the assumption embodied in Eq. (1) that the memory-storage probabilities can be expressed as multiplicative combinations of separable effects of set size and lag. Again, there is an infinite variety of ways in which discrete-slots processes (as well as other processes that give rise to mixed states) might be formalized. We make no claim that the storage probabilities resulting from each and every such process can be formally expressed in terms of Eq. (1). Instead, the equation is intended to provide a rough and general description of the patterns of storage probabilities without the requirement of estimating an extremely large number of free parameters.

Table 1

Glossary of parameters for the hybrid model.

<i>Memory-state probability parameters</i>	
$\alpha(ss)$	Memory probability factor associated with set size ss
β	Rate at which memory probability decreases with lag
<i>prim-boost-m</i>	Memory-probability primacy boost
<i>Memory-based drift-rate parameters</i>	
κ_C	Scaling factor for mean correct drift rate on memory-based change accumulator
κ_{NC}	Scaling factor for mean correct drift rate on memory-based no-change accumulator
$vss(i)$	Correct drift rate scaling factor associated with set size i
$\gamma(C)$	Rate at which correct “change” drift rate decreases with lag
$\gamma(NC)$	Rate at which correct “no-change” drift rate decreases with lag
<i>prim-boost-v</i>	Correct drift rate primacy boost
σ_M	Between-trial standard deviation of memory-based drift rate
<i>Memory-based start-point and response-threshold parameters</i>	
A_M	Start-point variability on memory-based accumulators
$ss_C(i) + cp_C(j)$	Additive response threshold parameters on memory-based change accumulator at set-size level i and change-probability level j
$ss_{NC}(i) + cp_{NC}(j)$	Additive response threshold parameters on memory-based no-change accumulator at set-size level i and change-probability level j
<i>Guessing-based drift-rate parameters</i>	
ν_G	Mean drift rate on guessing accumulators
σ_G	Between-trial standard deviation of guessing-based drift rate
<i>Guessing-based start-point and response-threshold parameters</i>	
A_G	Start-point variability on guessing-based accumulators
$G_C(j)$	Additive response-threshold parameter on guess-change accumulator at change-probability level j
$G_{NC}(j)$	Additive response-threshold parameter on guess-no-change accumulator at change-probability level j
<i>Base time</i>	
t_0	Base time

normal distribution.) On the memory-based accumulators, this drift-rate variability parameter is symbolized by σ_M . Finally, the mean error drift rate (i.e., the drift rate on the incorrect accumulator) is simply one minus the mean correct drift rate.

As illustrated in Fig. 5, in the LBA model, the start point of the evidence-accumulation process varies randomly across trials in accord with a uniform distribution. For the memory-based accumulators, the range of that uniform distribution is $[0, A_M]$. The response thresholds on the memory accumulators are allowed to vary as a function of both set size and change probability. For simplicity, we assume an additive combination rule, with the response-threshold setting on the change accumulator given by

$$R_C(i, j) = ss_C(i) + cp_C(j) + A_M, \quad (4)$$

where $ss_C(i)$ is the adjustment to the threshold at set-size level i , and $cp_C(j)$ is the adjustment to the threshold at change-probability level j . An analogous expression describes the response threshold settings on the no-change accumulator.

Eqs. (2)–(4) described the drift rates and response thresholds on the memory-based accumulators. Separate sets of parameters describe the drift rates and response thresholds on the guessing accumulators. For the guessing process, the mean drift rate on both the change and no-change accumulators is given by ν_G and the between-trials standard deviation of drift is given by σ_G . The start-point parameter of the guessing accumulation process is given by A_G . The response thresholds on the guessing accumulators are presumed to vary with change-probability only. $G_C(j)$ denotes the (additive) response-threshold parameter on the guess-change accumulator at change-probability level j , and analogously for $G_{NC}(j)$. Presumably, as objective change-probability increases, the values of $G_C(j)$ will tend to

decrease whereas the values of $G_{NC}(j)$ will tend to increase. That is, the observer's threshold for responding change will tend to become more lax, whereas the threshold for responding no-change will become stricter.

The hybrid model also estimates a base-time parameter t_0 to represent processes not associated with decision making, such as encoding and motor-execution times.

5.2.2. Mixed-state model

The pure mixed-state model is a special case of the hybrid model in which the mean-correct drift rates $v_C = \kappa_C$ and $v_{NC} = \kappa_{NC}$ are invariant with set size, lag and primacy. Thus, the $u(ss)$, γ and *prim-boost-v* parameters in Eqs. (2) and (3) are set at default values ($u(ss) = 1$ for all ss , $\gamma_C = \gamma_{NC} = 0$, and *prim-boost-v* = 1). All other aspects of the model are the same as in the hybrid model.

5.2.3. Shared-resources model

The pure shared-resources model is a special case of the hybrid model in which the memory-probability value (m) is equal to one across all set sizes and lags. Thus, the α , β and *prim-boost-m* parameters are set at default values ($\alpha(ss) = 1$ for all ss ; $\beta = 0$; and *prim-boost-m* = 1). Because $m = 1$, none of the parameters associated with the guessing accumulators are estimated for this model.

5.3. Model-fitting method

We fitted the models to the data of the individual subjects by conducting computer searches (Hooke & Jeeves, 1961) for the values of the free parameters that provided maximum-likelihood fits to the *individual-trials* choices and RTs. In this method, the likelihood of the choice and RT on each individual trial is computed, and the likelihood of the full data set is the product of these individual-trial likelihoods. The likelihood equations for the standard LBA model are presented in Brown and Heathcote (2008, Eqs. (1)–(3)) and it is straightforward to extend these equations for fitting the models that assume probabilistic mixtures of LBA processes.

The fit of each model was then assessed with the Bayesian Information Criterion (BIC; Schwarz, 1978):

$$\text{BIC} = -2 \ln(L) + p \ln(N),$$

where L is the maximum likelihood, p is the number of free parameters in the model, and N is the total number of trials on which the fit is based. The term $p \ln(N)$ penalizes a model for its number of free parameters. The model that yields the smallest BIC is considered to provide the most parsimonious fit to the data. To guard against local minima, each model was fitted to each individual subject's data by using 40 separate computer searches with different random starting parameters in each search.

5.4. Model-fitting results

The BIC fits of the models are reported in Table 2. The best-fitting model for each subject is indicated in boldface. For all subjects the hybrid model provides the best fit. The improvement in fit of the hybrid model compared to the other models is usually quite dramatic.

To gain insight into these model-fitting results, we start by considering the predictions from the hybrid model. Although the model was fitted to each individual subject's data, in this initial analysis we display the predictions averaged across the six subjects. The purpose is to evaluate the model's ability to capture the main trends in the data.

The averaged predictions are displayed in the bottom panels of Figs. 1–4. As can be seen in the bottom panels of Fig. 1, the model captures quite well how hit and false-alarm rates vary as a joint function of set size and lag. The model predicts the precipitous increase in false alarms that arise with lag and the smaller decline in hits. The main explanation is that items with long lags have a high probability of being in the zero-information state. Under such conditions, the observer has a bias to guess "change," accounting for the strong increase in false alarms and smaller decline in hits. The model also accounts for the finding that, once one conditions on lag, the set-size-5 and set-size-8 functions are nearly overlapping, whereas the set-size-2 function remains separated.

Table 2

BIC fits of the models to the individual-subject data.

Subject	Hybrid	Shared res.	Mixed state	Familiarity	Constrained hybrid
1	-3830.7	-3642.9	-3813.6	-3133.2	-3856.7
2	475.9	535.5	555.2	882.3	455.4
3	2631.1	3411.9	2861.8	3728.8	2603.6
4	-583.3	-416.5	-449.1	236.2	-594.7
5	1742.0	1836.7	1822.1	2776.6	1724.3
6	-2890.6	-2477.2	-2625.5	-1884.8	-2788.1

Note. Values in boldface are the minimum BIC scores among the hybrid, shared resources, mixed-state and familiarity-based models.

The averaged ROC curves predicted by the model are displayed in the bottom panel of Fig. 2. Naturally, the model predicts the manner in which overall sensitivity decreases with increases in set size. In moving from set-size 5 to set-size 8, the main reason for the decreased sensitivity is an increase in the averaged false-alarm rate rather than a decrease in the hit rate. The basis for this pattern is the same as previously described for Fig. 1 results. In addition, the model predicts correctly how both hit and false-alarm rates increase with increases in objective change probability. As will be seen, the main basis for this prediction is that the observers adjust their response thresholds in the manner that was encouraged by the instructions: as objective change probability increases, the observers' thresholds for responding change become more lax, and their thresholds for responding no-change become stricter. The shapes of the set-size-5 and set-size-8 ROCs are also captured reasonably well by the model, but there are some discrepancies for the set-size-2 ROC.

The bottom panel of Fig. 3 provides a plot of how mean RTs for hits, misses, false alarms and correct rejections are predicted to change with increases in objective change probability. The model captures very well the speed-up in hits and false alarms; the slow-down in misses and correct rejections; the cross-over in these functions across the objective change-probability conditions; that hits are always faster than false alarms; and that correct rejections are always faster than misses. The quantitative predictions are reasonably good as well, with the minor exception that the predictions for false alarms are slightly too slow in the .7 change-probability condition.

Finally, the bottom panel of Fig. 4 plots how mean RTs for hits, misses, false alarms and correct rejections are predicted to vary with changes in lag and set size. The model captures well the main result that correct-rejection RTs get longer with increases in lag: As lag increases, there is an increased guessing probability, and the guess-no-change process operates more slowly on average than does the memory-based respond-no-change process. The model also predicts correctly that, once one conditions on lag, there is little remaining effect of set size per se on the speed of the correct-rejection responses. Finally, the model predicts that the RTs for hits, misses and false alarms are nearly invariant across changes in set size and lag. As explained earlier, the predictions involving the flat error RTs arise as a natural consequence of the mixed-state structure of the model, in which error responses arise almost exclusively from the guessing process. However, the predictions of flat hit RTs are an accident of the particular parameter settings chosen by the model, in which the speed of the guess-change process turns out to be roughly the same as the speed of the memory-based respond-change process.

Although the hybrid model provides far better quantitative fits to most of the individual subjects' data than does the shared-resources model, it is difficult, unfortunately, to point to salient qualitative aspects of the results that are well captured by one model but not the other. Instead, the hybrid model yields an assortment of quantitative improvements. To illustrate, in Figs. 6–9 we display the individual-subject data from S3 along with the quantitative predictions from the hybrid model (top panels) and the shared-resources model (bottom panels). As can be seen in Fig. 6, the shared-resources model underestimates the hit rates for set sizes 2 and 5 to a greater extent than does the hybrid model. As shown in Fig. 7, although both models make similar qualitative predictions of the ROC data, the shared-resource model's quantitative mis-predictions are more pronounced, especially for the set-size-5 condition and also the set-size-2 condition when objective change-probability is at .3. As shown in Fig. 8, both models have some difficulty predicting how the speed of the different

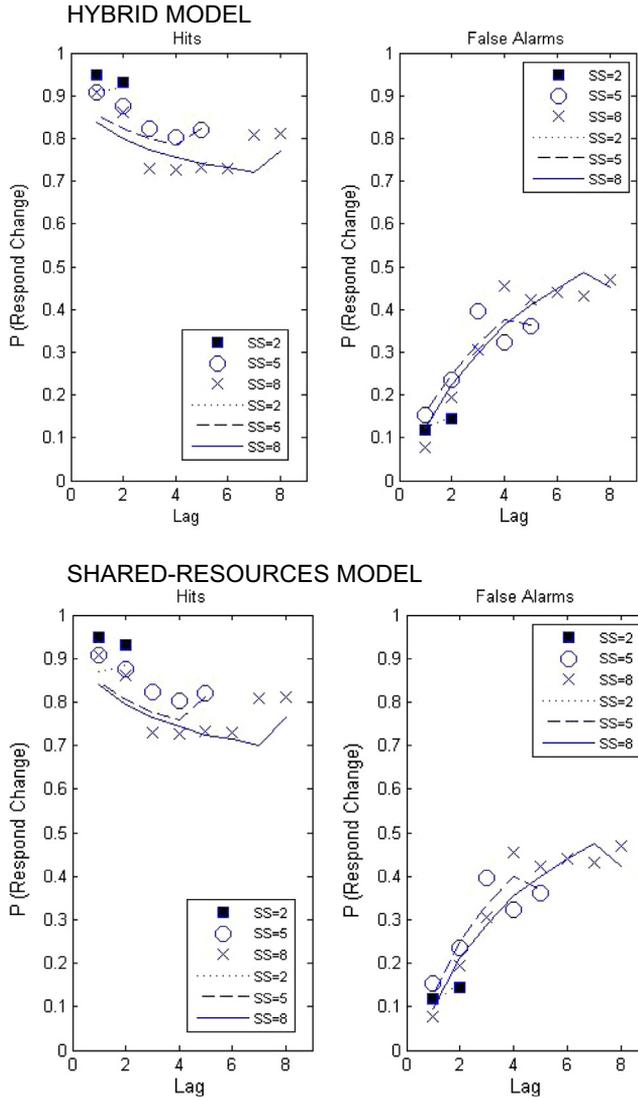


Fig. 6. Quantitative predictions of the hybrid model (top panels) and shared-resources model (bottom panels) of the proportion of hits and false alarms plotted as a function of memory set size (SS) and study-test lag for Subject 3.

response types varies with objective change probability. For example, both models predict hits that are too fast and misses that are too slow. Again, however, the mis-predictions are more severe for the shared-resources model. To take one example, whereas the hybrid model predicts well the speed of the correct rejections, the shared-resources model predicts correct rejections that are too slow. To take another example, when objective change probability is .5, hits and false alarms are slower on average than are correct rejections and misses. The hybrid model correctly predicts this ordering, but the shared-resources model predicts the opposite average ordering. Finally, as shown in Fig. 9, the hybrid model yields better quantitative predictions than does the shared-resources model of how the correct RTs (hits and correct rejections) vary with lag and set size. The shared-resources model predicts hit RTs that are too short across all lags at set-sizes 2 and 5, and predicts correct-

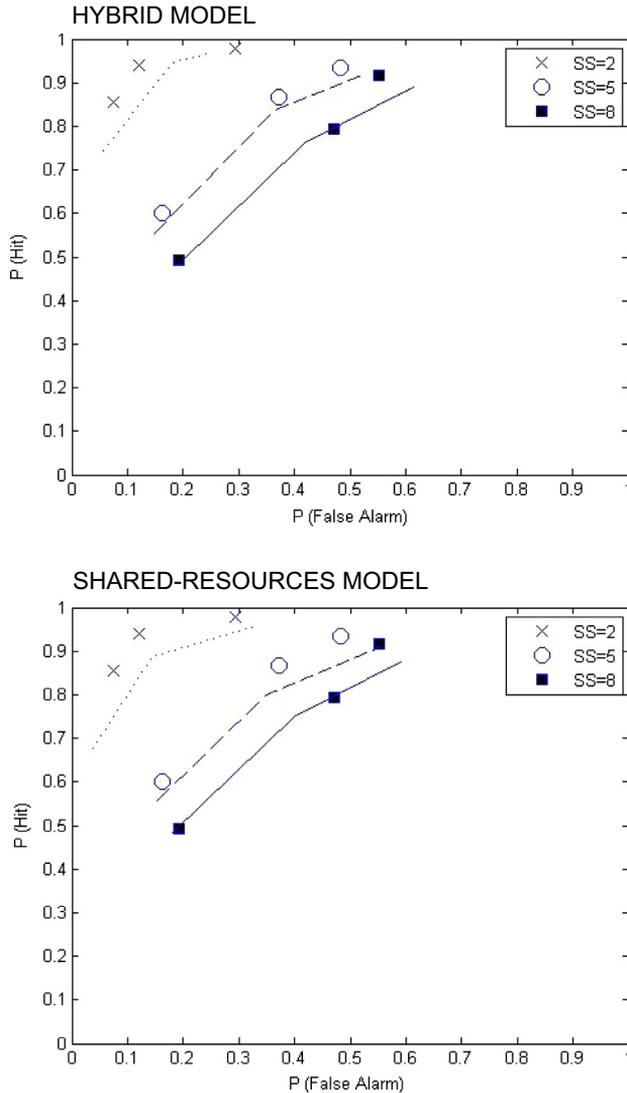


Fig. 7. Quantitative predictions of the hybrid model (top panel) and shared-resources model (bottom panel) of the ROC curves for each memory set size (SS) for Subject 3. Points within each ROC curve are for the different objective-change probability conditions.

rejection RTs that are too long across all lags at all set sizes. The error RTs are noisy, but both models clearly predict miss RTs that are too long. The magnitude of this mis-prediction is greater for the shared-resources model than for the hybrid model.

In Fig. 10 we display the detailed RT-distribution data for Subject 3 (open triangles), along with the predictions from the hybrid model (solid triangles). The RT distributions are plotted as a joint function of response type (hit, miss, false alarm, correct rejection), memory set size (2, 5, 8), and objective change probability (.3, .5, .7). (There were insufficient data to allow for fine-grained analyses of the RT distributions further broken down by lag.) The plotted points correspond to the RTs that form the .10, .30, .50, .70 and .90 quantiles in each distribution. Overall, the model appears to capture

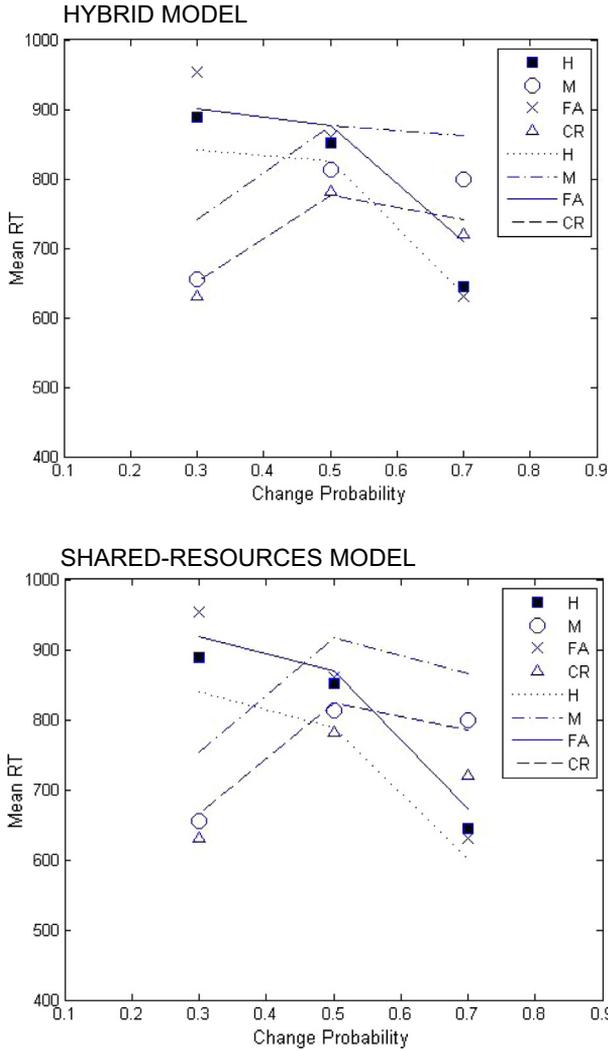


Fig. 8. Quantitative predictions of the hybrid model (top panel) and shared-resources model (bottom panel) of the mean response times (RTs) for hits, misses, false alarms and correct rejections plotted as a function of objective change probability for Subject 3.

the locations, spreads, and shapes of the RT distributions very well. We observed a similar pattern of results for all the other individual subjects. Perhaps the main systematic failing of the model for Subject 3 is that it predicts miss RTs at the .90 quantile that are too long. This failing is probably the key reason why the model predicted mean miss RTs that were too long (Fig. 9, top panel). Possibly, the subject may have incorporated some internal deadline for responding “same” and short-circuited on occasion the evidence-accumulation process if it ran on too long. The RT-distribution predictions from the pure continuous model (not shown here) were similar in form to those of the hybrid model. However, as was the case for the mean RTs, there was a varied assortment of cases in which its quantitative deviations (discrepancy between predicted and observed quantiles) were greater in magnitude than those of the hybrid model.

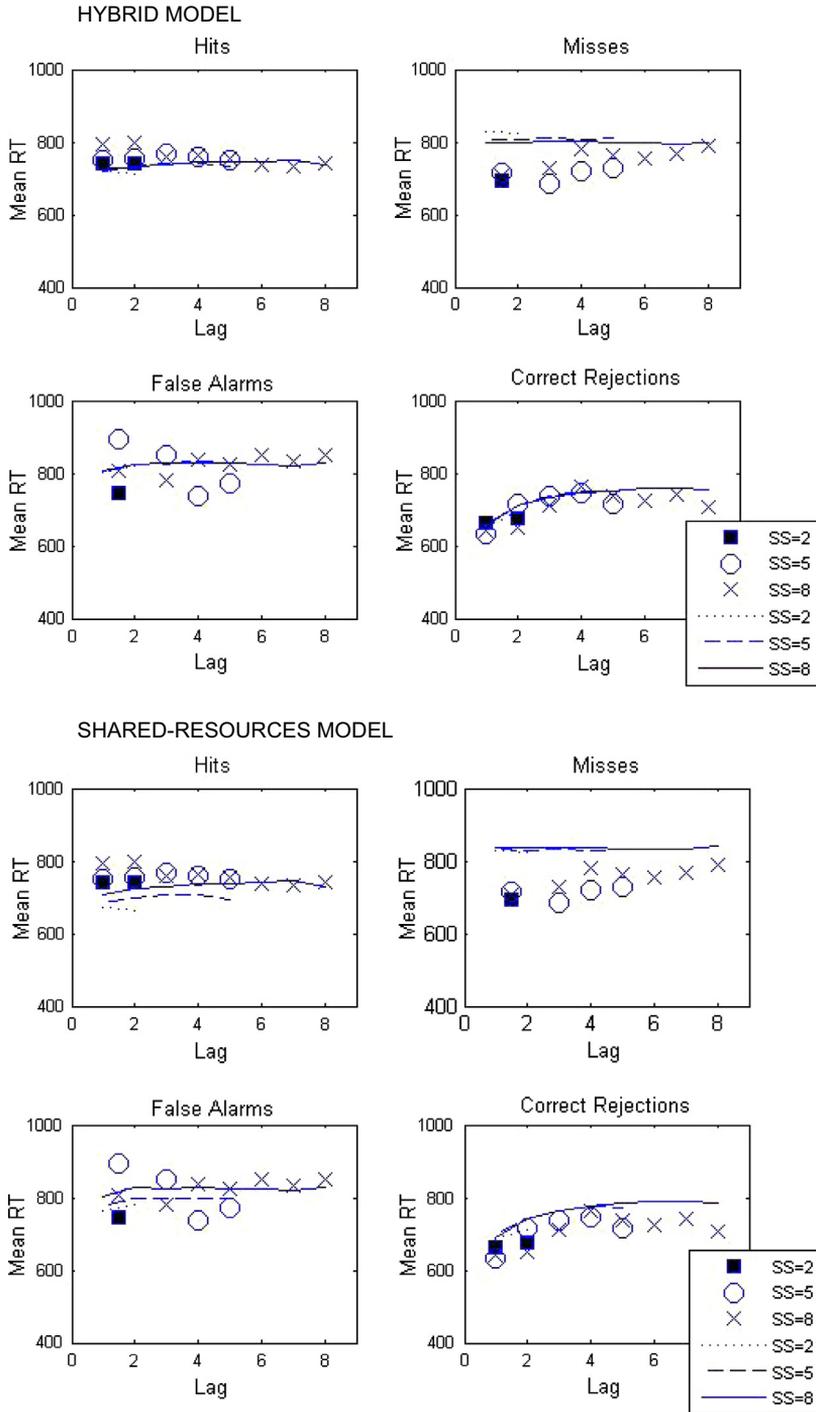


Fig. 9. Quantitative predictions of the hybrid model (top panels) and shared-resources model (bottom panels) of the mean response times (RTs) for hits, misses, false alarms and correct rejections plotted as a function of memory set size (SS) and study-test lag for Subject 3. Because of small sample sizes, error RTs (misses and false alarms) for the lag 1–2 results are averaged across these lags.

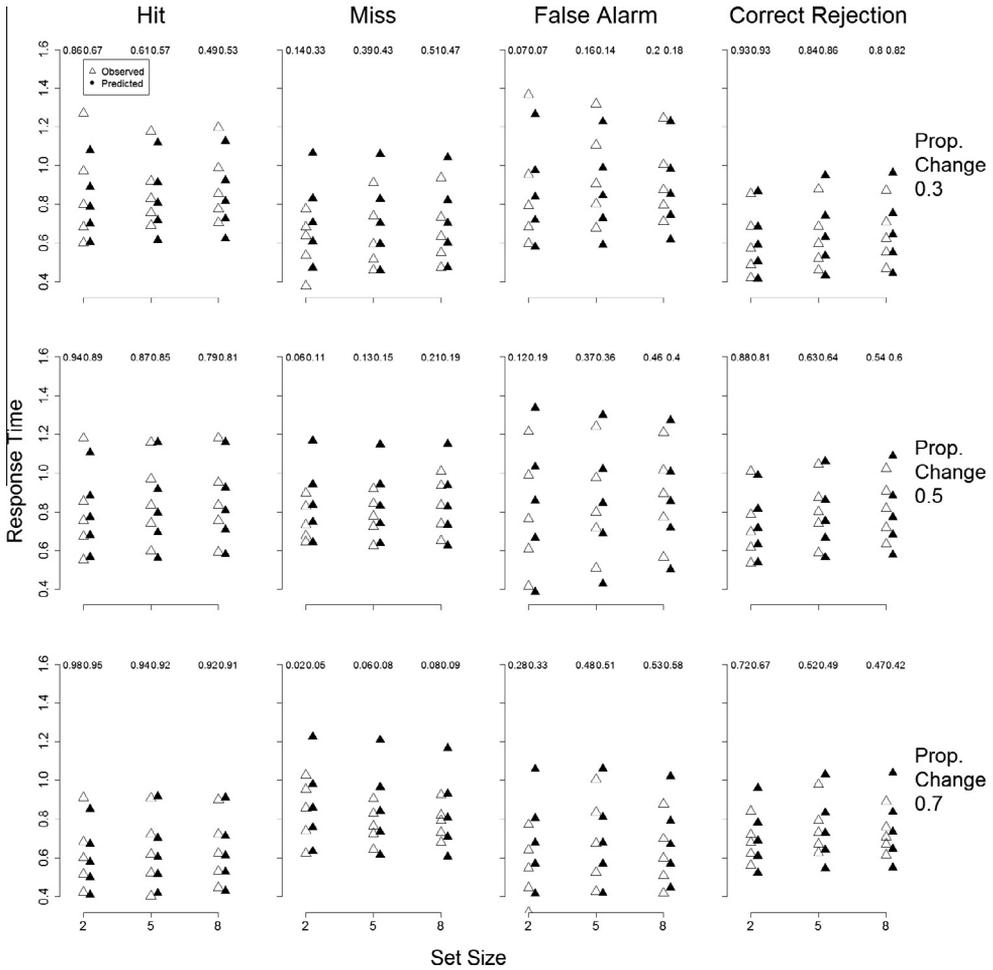


Fig. 10. Observed and hybrid-model predicted response-time (RT) distributions as a function of response type (hit, miss, false alarm correct rejection), memory set size (2, 5, 8), and objective change probability (.3, .5, .7) for Subject 3. Data and predictions are collapsed across the different lags. The solid triangles display the observed .10, .30, .50, .70, and .90 quantiles for each combination of conditions. The open triangles display the predicted quantiles.

5.5. Best-fitting parameters

In Table 3 we report the median values of the best-fitting parameters for each of the main models (hybrid, discrete-slots and shared-resources). We focus our discussion on the parameter estimates from the hybrid model, but the parameter estimates from the special-case models tell a similar story. First, set-size per se had little influence on the memory-storage probabilities across set sizes 5 and 8 (the α estimates are essentially identical). Instead, the main reason for the performance differences across set sizes 5 and 8 is that set-size-8 included items with greater lags, and memory-storage probability decreased with lag (the β memory-decay parameter is greater than zero). The implication is that items with greater lags have a greater probability of being “bumped” from the discrete-slots buffer. Set-size-2 was special, with an enhanced memory-storage parameter. Because essentially all versions of discrete-slots models presume that there are more than two slots available for storing items in VWM, it seems reasonable that the set-size-2 condition would have a higher memory-storage-

Table 3

Median values of best-fitting parameters from the hybrid, discrete-slots, and shared-resources models.

Model	Hybrid	Discrete-slots	Shared-resources
$\alpha(2), \alpha(5), \alpha(8)$.91, .85, .86	.92, .84, .82	–
β	.17	.73	–
<i>prim-boost-m</i>	1.08	1.77	–
k_G, k_{NC}	.81, .72	.96, .76	.72, .71
$vss(2), vss(5), vss(8)$	[1], .93, .91	–	[1], .91, .91
$\gamma(C), \gamma(NC)$.07, .25	–	.06, .21
<i>prim-boost-v</i>	1.07	–	1.07
σ_M	.19	.22	.20
A_M	.17	.21	.16
$ss_C(2), ss_C(5), ss_C(8)$	[0], –.01, –.03	[0], .05, .06	[0], .00, .01
$ss_{NC}(2), ss_{NC}(5), ss_{NC}(8)$	[0], –.00, –.00	[0], .02, .02	[0], .01, .01
$cp_C(.3), cp_C(.5), cp_C(.7)$.33, .32, .27	.37, .34, .31	.27, .24, .19
$cp_{NC}(.3), cp_{NC}(.5), cp_{NC}(.7)$.19, .23, .27	.18, .24, .27	.17, .23, .26
ν_G	.48	.31	–
A_G	.20	.19	–
$G_C(.3), G_C(.5), G_C(.7)$.22, .12, .10	.14, .11, .08	–
$G_{NC}(.3), G_{NC}(.5), G_{NC}(.7)$.12, .21, .26	.09, .16, .18	–
t_0	.120	.100	.100

Note. *ss*, and *cp* parameters denote the extent to which the memory-based response thresholds are incremented beyond the value of the A_M start-point parameter. The G parameters denote the extent to which the guessing-based response thresholds are incremented beyond the value of the A_G start-point parameter. Parameters in brackets are held fixed at default values for scaling convenience. In all cases, $\sigma_G = .1$.

probability factor. In particular, in this condition, the stored items would not be “bumped” from the VWM buffer by subsequent items in the sequence.

These same patterns are observed for the estimated memory drift-rate parameters. Set sizes 5 and 8 gave rise to similar drift rates (the *vss* estimates are nearly identical). Instead, mean correct drift rate decreased with lag (the γ memory-decay parameters are greater than zero), and set-size-8 includes items with greater lags. Again, however, set-size-2 is special, and gives rise to greater mean correct drift rates even after one conditions on lag. Hybrid versions of discrete-slots models, such as those proposed by Zhang and Luck (2008) and Cowan and Rouder (2009), are consistent with this general pattern of results. According to such models, there is sharing of resources if the number of retained items does not exceed the number of discrete slots. Once the number of discrete slots is exceeded, no further sharing takes place. Thus, if there is a set-size-related factor that influences drift rates, the factor should be the same for set sizes 5 and 8, but the factor should be greater for set-size 2. Note that these drift-rate parameter estimates help explain the empirical patterns that we described in Figs. 1 and 4: once one conditions on lag, the set-size 5 and 8 curves are nearly overlapping for both choice probabilities and mean RTs, but the set-size-2 curves are separated apart from the others.

The manipulations of objective change probability across conditions had systematic effects on the magnitudes of the response thresholds on both the memory-based accumulators and the guessing accumulators. As objective change probability increased, the response thresholds on the change accumulators grew more lax, whereas the response thresholds on the no-change accumulators became stricter. These parameter estimates help explain the empirical patterns that we described in Figs. 2 and 3 in which choice probabilities and mean RTs were systematically affected by the manipulation of objective change probability.

Under the present conditions of testing, the manipulations of set size had little effect on the magnitudes of the response thresholds (the ss_C and ss_{NC} parameter estimates are all near zero). This result contrasts with the one reported by Donkin et al. (2013) in their simultaneous-presentation change-detection tasks, in which response thresholds on the memory-based accumulators increased with set size. (Increases in the magnitude of response thresholds with set size are also often observed in tasks of memory search, e.g., Donkin & Nosofsky, 2012b; McElree & Doshier, 1989; Nosofsky et al.,

2011; Ratcliff, 1978.) It is an open question why set size had little effect on the response thresholds under the present conditions of testing.

Finally, the asymmetry with which false-alarm and hit rates varied with lag (Fig. 1) can also be understood from inspection of the best-fitting parameters. First, recall that with increasing lag, there is increased guessing behavior. Furthermore, as can be seen in Table 3, the magnitudes of the response thresholds on the guess-change accumulator tended to be lower than the magnitudes on the guess-no-change accumulator. Thus, increased guessing behavior would lead to more “change” responses, explaining the steep increase in false alarms but the smaller decrease in hits with increasing lag.³

Because some of the free parameters appeared to be nearly invariant across different conditions of testing, and others did not depart very much from default values of zero, we decided to evaluate a somewhat more parsimonious version of the hybrid model that imposed further parameter constraints. In particular, we fitted a version that imposed the constraints $\alpha(8) = \alpha(5)$; $vss(8) = vss(5)$; and $ss_C(i) = ss_{NC}(i) = 0$ for all set sizes i . The BIC fits of this constrained hybrid model, reported in Table 2, are better than those of the full hybrid model for Subjects 1–5 (but worse for Subject 6). The generally good fits of this constrained hybrid model lend further support to the hypotheses from Zhang and Luck (2008) and Cowan and Roudier (2009) that sharing of resources takes place for small set sizes, but further sharing does not take place once some item-limit (not greater than 5) is exceeded. There are a number of other routes one might pursue to achieve still greater parsimony. For example, in our present approach, we used descriptive functions for modeling the memory-storage probabilities in the hybrid model. An important goal for future work is to test low-parameter process models defined by explicit mechanisms that govern the storage and ejection of study items from the discrete-slots buffer. As a second example, recall that the magnitude of the response thresholds on both the memory-based and guessing accumulators varied systematically with our manipulations of objective change probability. It seems likely that these variations could be well captured by imposing some quantitative function for relating the magnitude of the response-threshold settings to the objective change-probability values. Because the motivating theme of our research plan did not involve the search for parsimonious special-case models, we leave such investigations to future research.

5.6. Alternative model classes

As explained at the outset, the models we fitted to our data were intended to be reasonably general representatives of the hybrid, mixed-state, and continuous classes. In this section, we briefly consider some alternative possibilities from other general classes. First, all three models that we fitted to the data presumed that the change/no-change decision process was with respect to a single object at the probed location. A much different possibility is that the observer assesses the overall “familiarity” of the test probe with respect to all objects in the entire display. Nosofsky and colleagues demonstrated that such familiarity-based models provide excellent accounts of RT and accuracy data in a wide variety of visual short-term probe-recognition paradigms (e.g., Nosofsky, Cao, Cox, & Shiffrin, 2014; Nosofsky, Cox, Cao, & Shiffrin, 2014; Nosofsky et al., 2011; see also Kahana & Sekuler, 2002; Sekuler & Kahana, 2007). Although the observer’s task in those “memory-search” paradigms was to assess whether the test probe matched *any* item from a presented list (rather than a single object from a specific location), it seemed reasonable to test such a global-familiarity model in the present domain. Indeed, recall that “change” trials in the present design always involved “external” changes, i.e., colors that were big changes from all items in the memory set. Thus, there was no logical requirement that observers assess changes with respect to only a specific probed location. A strategy involving a familiarity assessment with respect to the entire memory set might have worked just as well.

The detailed implementation of the familiarity model is described in Appendix B. The bottom line is that the model provided dramatically worse fits to our data than did any of our main candidates (see Table 2), suggesting that subjects’ assessments of change were indeed specific to the probed location.

³ Because we interpret the “guess-change” asymmetry as arising from a comparison of “presence” versus “absence”, a more appropriate modeling approach might be to assume asymmetric drift rates on the guessing accumulators rather than asymmetric threshold settings. We re-fitted all our data using this alternative drift-rate approach and found that it yielded similar results as the currently reported asymmetric-threshold version.

A limitation of the implemented model, however, is that familiarity was assessed only with respect to how similar a probe was to each of the memory-set items in terms of their color (see [Appendix B](#)). A more nuanced familiarity-based model might assess this similarity as a joint function of both color match *and* location match ([Donkin, Tran, & Le Pelley, 2014](#)). Unfortunately, because we did not record individual-item location coordinates in our data files, we could not evaluate this more sophisticated possibility for the current experiment.

A second alternative concerns the possibility that a “variable” shared-resources model (e.g., [van den Berg et al., 2012](#)) might yield better accounts of our data than did the present models. According to such models, the pooled resources of VWM may be shared in highly uneven fashion among the set of to-be-remembered items. To extend such models to the present domain, the idea would be to allow the drift rates on the memory-based accumulators to be “doubly stochastic” ([van den Berg et al., 2012](#)): not only would there be random variability around the mean correct drift rates, but the mean correct drift rates would themselves vary among the items of a memory set. It is important to note that [Donkin et al. \(2013\)](#) gave explicit consideration to variable shared-resources models in their simultaneous-presentation design. In particular, they fitted a probability-mixture version of the shared-resources model in which the mean correct drift rates for individual items varied in accord with a power function. Thus, some items in a set might give rise to high mean drift rates (high memory resolution), whereas others might give rise to very low ones (low memory resolution). Even this highly flexible variable-resources model yielded far worse fits to Donkin et al.’s RT and accuracy data than did the mixed-state alternatives (for discussion, see [Donkin et al., 2013, pp. 897, 901–902](#)).

[Donkin et al.’s \(2013\)](#) conceptual justification for the doubly-stochastic drift rates in their simultaneous design was that observers may have attended in sequential fashion to different items in the display, yielding “covert lag” effects. In the present design, the idea was to make these lag effects explicit by presenting the items in a rapid visual sequence. Furthermore, we explicitly incorporated the lag effects in the formal models. Thus, it is important to note that, for any given memory set size, the drift rates associated with the items are *already* doubly stochastic in the present models. We acknowledge that there may be yet other reasons for high variability in mean drift rates across items, so one might formulate models in which the drift rates are stochastic at three levels. We leave such more complicated possibilities, however, to future research.

6. General discussion

6.1. Summary and conclusions

In this study we tested subjects in a rapid sequential-presentation visual change-detection task. We measured how accuracies and RTs varied as a function of memory set size, lag of the probed location, and objective change probabilities. The main purpose was to test mixed-state, continuous, and hybrid models of VWM on their ability to account for the individual-subject change-detection RT and accuracy data.

In a related paradigm, [Donkin et al. \(2013\)](#) had reported evidence in favor of hybrid models of change-detection performance. Models that made allowance for mixed memory states – including a zero-stimulus-information state that required guessing – yielded far better quantitative accounts of the RT and accuracy data than did models that assumed only continuous variations in memory resolution across items. However, that previous evidence was obtained in a simultaneous-presentation change-detection task. Conceivably, under those conditions, subjects may have adopted strategies in which only a subset of the items from the simultaneous display were encoded. The evidence for a prominent role of a zero-stimulus-information state may simply have reflected that some items never entered VWM in the first place.

One motivation for our use of the present sequential-presentation design was to potentially address this encoding explanation. Under the present conditions, memory set-size (2, 5, 8) varied randomly across trials, so subjects could not anticipate the number of to-be-remembered items on any given trial. If they chose to ignore the initially presented items of any given sequence, then performance would have been poor on the set-size-2 trials; however, performance was by far the best on

those trials, so it appears that the initial items in the sequences were encoded. If subjects chose to ignore subsequent items in the sequences – or if the initial items interfered with subjects' ability to encode the subsequent ones – then items near the end of each sequence should have yielded poor performance. However, exactly the opposite pattern was observed: regardless of memory set size, items near the end of each sequence (i.e., those with short testing lags) yielded the best performance. Instead, it was items with long testing lags that seemed to drop out of memory. This pattern of results suggests limits on visual working *memory* capacity rather than encoding-based explanations.

The general qualitative result that performance was worst for items with long lags is compatible with both discrete-slots (i.e., mixed-state) and continuous shared-resource models of VWM. According to the discrete-slots models, the items with long lags have a higher probability of being ejected from the discrete-slots buffer (e.g., Atkinson & Shiffrin, 1968). According to the shared-resources models, the items with long lags develop increasingly diffuse memorial representations (e.g., Kinchla & Smyzer, 1967). By requiring the models to simultaneously fit both accuracy and RT data, we in fact obtained evidence in favor of both mechanisms. The hybrid model yielded far better quantitative fits to the individual-subject RT and accuracy data than did the pure versions of either the mixed-state or shared-resources models. Furthermore, parameter estimates from the hybrid model revealed that as lag increased, both memory-storage probability and memorial resolution decreased.

Among the most salient empirical results in our study was that as lag increased, there was a precipitous increase in false alarms, i.e., responding “change” on trials in which the probed stimulus stayed the same. This dramatic increase in false alarms was accompanied by a much more modest reduction in hits, i.e., responding “change” on trials in which the probed stimulus did in fact change. In our view, this strong asymmetry (bias to respond “change”) is most naturally interpreted in terms of the mixed-state component of the hybrid model. Items with long test lags have a higher probability of being ejected from the discrete-slots buffer, resulting in a zero-stimulus-information state. If the observer is probed at a location with zero stimulus information, the natural interpretation is that the stimulus is “new”, so it is natural to guess “change”. As acknowledged earlier, with appropriate choice of free parameters, the continuous shared-resources model was able to fit the strong bias to respond “change” as lag increased; however, if memory simply becomes “fuzzier” as lag increases, then this strong asymmetry in making “change” versus “same” judgments does not seem to be an *a priori* prediction from this class of models.⁴ At the least, the presence of this dramatic asymmetry should be viewed as a major target phenomenon to be explained by all competing classes of models.

Another important result is that, for most of our observers, error RTs (false alarms and misses) were nearly invariant with changes in set size and lag. This result too seems most naturally interpreted in terms of the mixed-state component of the hybrid model. In the present big-change experiment, in which all colors were highly discriminable from one another, almost all errors would arise from guessing. The gating process (Fig. 5) “tells” the observer that there is zero stimulus information available at the probed location, so the system must guess. Because the time course of the guessing process is not stimulus-dependent, the prediction is that the error RTs should therefore be invariant with set size and lag. Again, although pure shared-resource models can fit such data with appropriate adjustment of their free parameters, the error-RT invariance does not seem to be a natural prior prediction of such models.

Interestingly, under the present conditions of testing, hit RTs were also nearly invariant with lag. This invariance does not reflect a general lack of sensitivity of RTs to the experimental manipulations. In particular, correct-rejection RTs got systematically longer with increases in lag; and RTs for all response types (hits, misses, correct rejections and false alarms) also varied systematically with our

⁴ An alternative metaphor from the class of shared resources models is that memory does not become “fuzzier”; instead, some specific value is always remembered, but the between-trials distribution from which the specific remembered value is drawn becomes more variable as memory set size or lag increases. This type of model might indeed predict a bias towards responding “change”, because as between-trials variability increases, the distance between the remembered value of a study item and a *same* test probe would tend to increase (e.g., see Nosofsky & Gold, 2015). A detailed discussion of this alternative modeling approach goes beyond the scope of the present article. In brief, certain sophisticated versions of such models presume that the observer has access to the intrinsic variability associated with each individual item representation in the memory set and adjusts criterion settings in ideal-observer fashion in accord with the known variabilities (e.g., Keshvari et al., 2013). A wide variety of such models are possible and the predictions from such models would vary depending on the knowledge to which the observer is presumed to have access and the precise decision rules that are used.

manipulations of objective change probability. Instead, our interpretation is that under the present conditions, the guess-change process apparently operated at roughly the same speed as the memory-based respond-change process. Thus, the increased guessing that took place with increases in lag did not result in slowed hit responses.

Although the pattern of nearly-invariant hit RTs was unexpected, it is interesting to note that it seems to challenge alternative accounts of the change-detection RT data based on memory “search”. According to these alternative accounts, the observer needs to search through the memory representations of the individual objects to determine if a change has occurred. Presumably, search times would increase as the number of stored items increases or if it takes longer to complete the search for items that have longer lags. The present change-detection RT models are not “search” models. Instead, the models assume that when a location on the visual display is probed, the probe initiates a gating process that allows the system to directly access any information that may be remembered at that location. Changes in RT with set size or lag would arise because of how these variables might influence that direct-access process, not because the observer is searching through other memory representations from the display.

6.2. *Limits to generality*

The present research was conducted under conditions associated with a highly limited-capacity visual working memory. In particular, a small set of items was used that repeated on each and every trial. Furthermore, the task was conducted under “varied-mapping” conditions (Shiffrin & Schneider, 1977): Across trials, the same stimulus would sometimes constitute a “change” and other times constitute “no change”. Under alternative conditions, very different patterns of performance would be expected to arise. For example, in some tests of visual memory search, completely new sets of items are presented on each and every trial, which leads to vast reductions in proactive interference (e.g., Endress & Potter, 2014; Nosofsky, Cox, et al., 2014; see also Banks & Atkinson, 1974). Under such conditions, memory-search performance improves dramatically. Current research has been aimed at understanding the mechanistic basis for these improvements. One possibility, for example, is that under such conditions, observers can rely on information preserved in longer-term memory stores to make their judgments (e.g., Atkinson & Shiffrin, 1968; Endress & Potter, 2014). Likewise, even under conditions in which there is a small set of repeating items, visual and memory search performance improves dramatically if the items are consistently mapped to alternative response categories (Nosofsky, Cao, et al., 2014; Nosofsky, Cox, et al., 2014; Shiffrin & Schneider, 1977). Again, rather than relying on a limited-capacity visual working memory, performance in such cases may reflect the use of long-term memory and categorization processes (e.g., Nosofsky, Cao, et al., 2014). The modeling accounts developed in the present work were aimed at diagnosing performance under conditions revolving solely around the use of highly limited-capacity VWM and would need to be extended to account for performance under more general conditions.

6.3. *Relation to probe-recognition memory search*

Finally, the present results also leave us with something of a mystery that remains to be resolved. As discussed at various points in our article, the sequential-presentation change-detection is closely related to tasks of probe-recognition memory search.⁵ In both cases, a sequence of

⁵ Likewise, Hyun, Woodman, Vogel, Hollingworth, and Luck (2009) noted and investigated relations between change-detection and visual search, including the measurement of RTs in certain types of change-detection tasks. However, there were important differences between the issues and change-detection tasks examined by Hyun et al. and those that were the focus of the present work. First, because Hyun et al. considered change-detection tasks in which set size was always four or less, the issues had nothing to do with testing for mixed states in visual working memory. Second, in Hyun et al.’s tasks, rather than presenting a single test probe in a single location, the test probes were multi-item displays and the subjects’ task was to judge whether any of the individual items had changed. Whereas the RT models developed in the present work were aimed at characterizing the nature of information processing with respect to a single probed location, the RT models that would be required to account for performance in Hyun et al.’s paradigm would be considerably more complex. In particular, the models would need to characterize the manner in which observers processed an entire display of individual probed locations, rather than just a single one.

to-be-remembered items is presented, and the observer judges whether a test probe is new (i.e., a change) or old (i.e., the same). Among the main differences is that in the memory-search task, the observer needs to decide if the test probe is the same as *any* item from the memory set; whereas in the change-detection task, the question is whether the test probe is the same as one specific item from a given location. Clearly, the decision processes that operate are likely to differ across these tasks in order for the cognitive system to answer the different questions that are posed. However, one might expect to see evidence of similar types of memory representations across these closely related paradigms.

To our knowledge, however, despite the enormous variety of models of probe-recognition memory search that have been proposed throughout the history of cognitive psychology (for an early review, see [Townsend & Ashby, 1983](#)), none have posited a role of mixed states in that closely related paradigm. Although memory-search tasks typically involve small set sizes, there are various examples in which the set sizes exceed the limits posited in standard discrete-slots models (e.g., [Burrows & Okada, 1975](#); [Donkin & Nosofsky, 2012a, 2012b](#); [Endress & Potter, 2014](#); [Nosofsky, Cao, et al., 2014](#); [Nosofsky, Cox, et al., 2014](#); [Wolfe, 2012](#); [Öztekin, Davachi, & McElree, 2010](#)). Likewise, although studies from the memory-search literature typically use alphanumeric or verbal stimuli, there are many examples that have used visual stimuli (e.g., [Endress & Potter, 2014](#); [Nosofsky, Cao, et al., 2014](#); [Nosofsky, Cox, et al., 2014](#); [Nosofsky et al., 2011](#); [Wolfe, 2012](#)). Thus, the lack of theorizing involving mixed states in memory search does not seem to arise from procedural differences involving memory set size or types of stimuli.

Instead, perhaps the absence of hypotheses involving mixed states in memory search has arisen because that paradigm may not be as diagnostic as is the change-detection task for identifying such states. Suppose, in particular, that there has indeed been a complete loss of information for a particular study item. In the change detection task, suppose the location corresponding to that study item is probed. We have argued that it is more natural for the observer to guess “change” in such cases. However, responding “change” based on the absence of information is clearly different than responding “change” in cases in which the study item has been remembered and a true change has occurred. Furthermore, because the two processes are based on different psychological mechanisms, it is natural to expect that they will often follow different time courses. Finally, according to the present theorizing, the processing of *other* items from the study display does not influence the behavioral data: the time course of the response is based solely on the observer’s judgment with respect to the single probed location. Under those conditions, it becomes feasible to identify the influence of zero-stimulus-information memory states.

But in a corresponding memory-search task, the situation is not as clear-cut. The question now is whether the test probe is the same as *any* study item from the memory set. First, under such circumstances, the difference between *absence* of information as opposed to *changed* stimulus information may no longer be psychologically prominent: In both cases, there is simply a failure of the test probe to match any information stored in memory. Second, even if the processing mechanisms for those two situations continue to differ, the differences are embedded in the system’s processing of all other study items from the memory set. Under such circumstances, it may be exceedingly difficult to identify the operation of the zero-stimulus-information memory state that is posited in the mixed-state and hybrid models of visual working memory.

6.4. Conclusion

We have provided evidence that the time course of decisions in a sequential-presentation change-detection task reflects the operation of multiple response modes – one based on use of information from working memory; and a second, guessing-based process that reflects a zero-stimulus-information state. Our formal analyses were based on use of general descriptive models, and do not speak directly to more specific assumptions such as whether a fixed number of slots led to the mixture of memory- and guessing-based responses. Rather, it is the role of future research to develop a process model that can reproduce the general pattern of results that is summarized by our hybrid, mixed-state model.

Acknowledgments

This work was supported by grant FA9550-14-1-0307 from the Air Force Office of Scientific Research to Robert Nosofsky and by Australian Research Council Discovery Early Career Award (DE130100129) and Discovery Project (DP130100124) to Chris Donkin.

The authors would like to thank Edward Awh, Philip Smith, and three anonymous reviewers for their criticisms of earlier versions of this article.

Appendix A. Exponential lag function

We repeated the modeling analyses involving the hybrid, continuous and mixed-state models, except Eqs. (1) and (2) were modified such that

$$m(ss, lag) = \alpha(ss) \cdot \beta^{-lag} \cdot prim, \quad (A1)$$

and

$$v_C(ss, lag) = \kappa_c \cdot v(ss) \cdot \gamma(C)^{-lag} \cdot prim. \quad (A2)$$

The free parameters are as defined in the main text. The BIC fits of the modified models are reported in Table A1. In all cases, the BIC fits are somewhat worse for the present alternative versions that use an exponential lag function compared to the original versions that used a power function (compare fits in Table A1 to those in Table 2). However, as can be seen in Table A1, the hybrid model still yields a much better BIC fit than does the continuous shared-resources model for all six subjects. The pure mixed-state model continues to provide an intermediate fit for all six subjects.

Appendix B. Familiarity-based model

As is the case in the shared-resources model, in the familiarity model only the memory-based accumulators operate. We start by describing how the drift rates on those accumulators are computed (see Nosofsky, Cao, et al., 2014, for more detailed motivation). First, the “memory strength” for an item presented at lag j in a sequence of set-size ss is given by

$$m_j = \alpha(ss) \cdot j^{-\beta} \cdot prim, \quad (B1)$$

where $\alpha(ss)$ is a set-size factor; β is a memory-strength decay parameter; and $prim$ is the primacy multiplier.

The degree to which exemplar j (e_j) from the study list is “activated” when test-item i (t_i) is presented is a joint function of exemplar j ’s memory strength and its similarity to test item i :

$$a_{ij} = m_j, \quad \text{if } t_i = e_j \quad (B2a)$$

$$a_{ij} = m_j s, \quad \text{if } t_i \neq e_j \quad (B2b)$$

where s is a similarity parameter. Note that in the present implementation, similarity is computed only with respect to color: If the test item matches the memory-set item’s color, then the similarity is set equal to one; whereas if the test item mismatches the memory-set item’s color, then the similarity is set equal to s .

The mean drift rate on the “old” (no-change) accumulator is given by

$$p_i = A_i / (A_i + k), \quad (B3)$$

where A_i gives the summed activation (“familiarity”) of the test probe to all memory-set items:

$$A_i = \sum a_{ij}, \quad (B4)$$

and k is a criterion parameter that is assumed to be linearly related to memory set size:

$$k(M) = u + v \cdot M. \quad (B5)$$

Table A1

BIC fits of the negative-exponential version of the models to the individual-subject data.

Subject	Hybrid	Continuous	Mixed state
1	– 3793.3	–3620.3	–3778.9
2	537.0	584.4	561.8
3	2669.7	3453.0	2865.8
4	– 508.5	–282.7	–429.9
5	1779.1	1924.7	1847.7
6	– 2818.0	–2409.8	–2642.0

Note. Values in boldface are the minimum BIC scores among the hybrid, continuous, and mixed-state models.

The mean drift rate on the “new” (change) accumulator is equal to one minus the mean drift rate on the “old” (no-change) accumulator.

The other free parameters are the same as in the shared-resources model, including the between-trials standard deviation of drift; start-point variability parameter; additive response-threshold parameters (which are allowed to depend on both set size and objective change probability); and the base-time parameter.

References

- 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*, 106–111.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89–195.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628.
- Banks, W. P., & Atkinson, R. C. (1974). Accuracy and speed strategies in scanning active memory. *Memory & Cognition*, *2*(4), 629–636.
- Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1359–1367.
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 1–11.
- Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*, 1–15.
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.
- Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science*, *188*(4192), 1031–1033.
- Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*, 87–114.
- Cowan, N., & Rouder, J. N. (2009). Comment on “Dynamic shifts of limited working memory resources in human vision”. *Science*, *323*, 877c.
- Donkin, C., Brown, S. D., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, *18*, 61–69.
- Donkin, C., & Nosofsky, R. M. (2012a). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science*, *23*, 625–634.
- Donkin, C., & Nosofsky, R. M. (2012b). The structure of short-term memory scanning: An investigation using response-time distribution models. *Psychonomic Bulletin & Review*, *19*, 363–394.
- Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working memory response times. *Psychological Review*, *120*, 873–902.
- Donkin, C., Tran, S. C., & Le Pelley, M. (2014). Location-based errors in change detection: A challenge for the slots model of visual working memory. *Memory & Cognition*, *1*–11.
- Dube, C., Starns, J. J., Rotello, C. M., & Ratcliff, R. (2012). Beyond ROC curvature: Strength effects and response time data support continuous-evidence models of recognition memory. *Journal of Memory & Language*, *67*, 389–406.
- Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General*, *143*(2), 548.
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229.
- Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience*, *31*, 8502–8511.

- Hooke, R., & Jeeves, T. A. (1961). "Direct search" solution of numerical and statistical problems. *Journal of the ACM (JACM)*, 8(2), 212–229.
- Hyun, J. S., Woodman, G. F., Vogel, E. K., Hollingworth, A., & Luck, S. J. (2009). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1140.
- Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, 42, 2177–2192.
- Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLOS Computational Biology*, 9, e1002927.
- Kinchla, R. A., & Smyzer, F. (1967). A diffusion model of perceptual memory. *Perception & Psychophysics*, 2(6), 219–229.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Science*, 17, 391–400.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17, 347–356.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short term memory: The time course of recognition. *Journal of Experimental Psychology: General*, 118, 346–373.
- Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, 10(4), 465–501.
- Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 384–413.
- Nosofsky, R. M., Cao, R., Cox, G. E., & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology*, 75, 97–129.
- Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1524.
- Nosofsky, R. M., & Gold, J. (2015). Memory strength versus memory variability in visual change detection. *Attention, Perception, & Psychophysics*, 1–16.
- Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, 118, 280–315.
- Öztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. *Psychological Science*.
- Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2), 332.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44, 369–378.
- Pearson, B., Raskevicius, J., Bays, P. M., Pertzov, Y., & Husain, M. (2014). Working memory retrieval as a decision process. *Journal of Vision*, 14, 1–15.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106, 261–300.
- Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review*, 15, 825–832.
- Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, 105, 5975–5979.
- Rouder, J. N., Thiele, J. E., Province, J. M., Cusumano, M., & Cowan, N. (2015). The evidence for pure guessing in working-memory judgments. *Cognitive Psychology*.
- Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, 6, 461–464.
- Sekuler, R., & Kahana, M. J. (2007). A stimulus-oriented approach to memory. *Current Directions in Psychological Science*, 16(6), 305–310.
- Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*.
- Shapiro, K. L., & Miller, C. E. (2011). The role of biased competition in visual short-term memory. *Neuropsychologia*, 49, 1506–1517.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2), 127.
- Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review*, 119, 807–830.
- Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological Review*, 120, 589–627.
- Souza, A. S., Kerko, L., Lin, H.-Y., & Oberauer, K. (2014). Focused attention improves working memory capacity: Implications for flexible-resource and discrete-capacity models. *Attention, Perception, & Psychophysics*, 76, 2080–2102.
- Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variability and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, 64, 1–34.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153, 652–654.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.
- Umemoto, A., Drew, T., Ester, E. F., & Awh, E. (2010). A bilateral advantage for storage in visual working memory. *Cognition*, 117, 69–79.
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, 71, 1–26.
- van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparisons of working memory models. *Psychological Review*, 121, 124–149.
- van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, 109, 8780–8785.

- 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.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4, 1120–1135.
- Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, 2, 409–415.
- Wolfe, J. M. (2012). Saved by a log how do humans perform hybrid visual and memory search? *Psychological Science*. 0956797612443968.
- 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.