Journal of Experimental Psychology: Learning, Memory, and Cognition

Qualitative Contrast Between Knowledge-Limited Mixed-State and Variable-Resources Models of Visual Change Detection

Robert M. Nosofsky and Chris Donkin

Online First Publication, March 7, 2016. http://dx.doi.org/10.1037/xlm0000268

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000268

Qualitative Contrast Between Knowledge-Limited Mixed-State and Variable-Resources Models of Visual Change Detection

Robert M. Nosofsky
Indiana University Bloomington

Chris Donkin
University of New South Wales

We report an experiment designed to provide a qualitative contrast between knowledge-limited versions of mixed-state and variable-resources (VR) models of visual change detection. The key data pattern is that observers often respond "same" on big-change trials, while simultaneously being able to discriminate between same and small-change trials. The mixed-state model provides a natural account of this data pattern: With some probability, the observer is in a zero-memory state and is forced to guess. Thus, even on big-change trials, there is a significant probability that the observer will respond "same." On other trials, the observer retains memory for the probed study item, and these memory-based responses allow the observer to show above-chance discrimination between same and small-change trials. By contrast, we show that important versions of the VR models that we refer to as *knowledge-limited* models are stymied by this simple pattern of results. In agreement with Keshvari, van den Berg, and Ma (2012, 2013), alternative *knowledge-rich* VR models that employ ideal-observer decision rules provide a significant improvement over the knowledge-limited VR models; however, extant versions of the knowledge-rich VR models still fall short quantitatively compared to the descriptive mixed-state model. We discuss implications of the knowledge-rich assumptions that are posited in current versions of the VR models that have been used to fit change-detection data.

Keywords: visual working memory, visual change detection, mixed-state models, variable-resources models

Visual working memory (VWM) is the short-term memory (STM) system that maintains visual representations of features and objects. In a typical VWM task, the observer is presented with a brief visual display of objects. Following a brief retention interval, one or more locations from the original display are probed and the observer's memory for the objects in the probed locations is tested. A fundamental result is that people's memory for the individual objects in the display declines dramatically as the number of to-be-remembered objects in the display increases (Luck & Vogel, 1997).

In this article, we contrast two major classes of theories that have been proposed to explain the severe capacity limit on VWM. In *mixed-state* models, the idea is that a studied object either resides in the VWM system or else all memory-based information regarding the object has been lost. Mixed-state models are most directly motivated by discrete-slots theories of VWM (e.g., Awh, Barton, & Vogel, 2007; Cowan, 2001; Cowan & Rouder, 2009; Donkin, Nosofsky, Gold, & Shiffrin, 2013; Luck & Vogel, 2013; Rouder et al., 2008; Vogel, Woodman, & Luck, 2001; Zhang & Luck, 2008). According to discrete-slots theories, VWM makes available some limited number of slots for storing objects in memory. If a studied object is retained in one of the slots, then the system can make use of this memory if the object's location is probed. By contrast, if the studied object has not been retained in one of the slots, then zero stimulus-based information regarding the object remains. In this case, if the location in which the object resided is probed, the observer is forced to guess regarding the object's identity. Importantly, however, in the present article, our focus is on the general class of mixed-state models rather than on specific processes that may give rise to the mixed states.

We will refer to the second main class of theories that we examine in the present work as the *pure continuous* class. This class is most directly motivated by shared-resources views of VWM (e.g., Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Ma, Husain, & Bays, 2014). According to such models, VWM makes available a limited pool of memory resources that is shared among the objects in the studied display. The greater the resources devoted to a given item, the more fine-grained is the memory for that item. According to modern versions of such models, the amount of resources devoted to individual items may be highly variable across the different items of the visual display (Fougnie, Suchow, & Alvarez, 2012; Keshvari, van den Berg, & Ma, 2012, 2013; Sims, Jacobs, & Knill, 2012; van den Berg, Shin, Chou, George, & Ma, 2012). However, for the pure continuous class that

Robert M. Nosofsky, Department of Psychological and Brain Sciences, Indiana University Bloomington; Christopher Donkin, School of Psychology, University of New South Wales.

This work was supported by AFOSR Grant FA9550-14-1-0307 to Robert M. Nosofsky; and Australian Research Council Discovery Early Career Award (DE130100124) to Christopher Donkin. Our thanks to Wei Ji Ma, Timothy Ricker, and a number of anonymous reviewers for their criticisms, comments, and discussions pertaining to earlier versions of this article.

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Department of Psychological and Brain Sciences, Indiana University, 1101 E. Tenth Street, Bloomington, IN 47405. E-mail: nosofsky@indiana.edu

we consider here, there is no true zero-stimulus-information state that requires the operation of a guessing process.[1]

In recent years, a major paradigm that has been used for testing the predictions of mixed state and continuous models is the continuous-recall paradigm (van den Berg, Awh, & Ma, 2014; Wilken & Ma, 2004; Zhang & Luck, 2008). In this paradigm, a particular location from the visual display is probed, and the observer is required to point to an appropriate location on a continuous response device, such as a continuous color wheel, that reproduces the value of the originally studied object.

According to mixed-state models, if the studied object has been retained in memory, then the continuous response is modeled as a random draw from a (circular) normal distribution centered on the value of the original stimulus. By contrast, if the studied object has not been retained, then the observer guesses with a value that is independent of the value of the originally presented stimulus. Because the studied objects are drawn uniformly from a continuous circular dimension (such as a color wheel), the aggregated distribution of guessing errors across trials of the experiment will be uniformly distributed. Thus, according to the mixed-state models, the aggregated distribution of response errors across trials of the experiment will be a mixture of a normal (centered on zero) and a uniform distribution, as illustrated schematically in the top panel of Figure 1 (e.g., Zhang & Luck, 2008).

By comparison, according to (pure) continuous variable-resources models of VWM, responses are always stimulus based and are modeled as random draws from circular normal distributions centered on the value of the original stimulus. On trials in which the observer devoted a large proportion of the pool of resources to the probed object, the memory will be highly precise and the particular remembered value will be a random draw from a normal distribution with a small variance. As the proportion of resources devoted to each study item decreases, the memories become less precise, and the particular remembered value on each trial will be a random draw from a normal distribution with increasing variance. Across trials, the aggregated distribution of response errors will be a mixture of random draws from (circular) normal distributions (centered on zero) with variable variances, as illustrated in the bottom panel of Figure 1 (e.g., van den Berg et al., 2012).

Comparing the top and bottom panels of Figure 1, it is apparent that the two classes of models make very similar qualitative predictions with respect to the distribution of response errors in the continuous-reproduction task. Recent work suggests some advantages in the quantitative predictions from the class of variable-resources models (van den Berg, Awh, & Ma, 2014). However, it is currently unknown the extent to which these quantitative advantages may depend on detailed psychophysical parametric assumptions. For example, applications of the competing models are based on the assumption that each adjacent color on the color wheel is evenly spaced. Evidence for variable resources may reflect, in part, violations of this assumption. In addition, there may be no very strong reason to assume that the component memory distributions are precisely normally distributed.

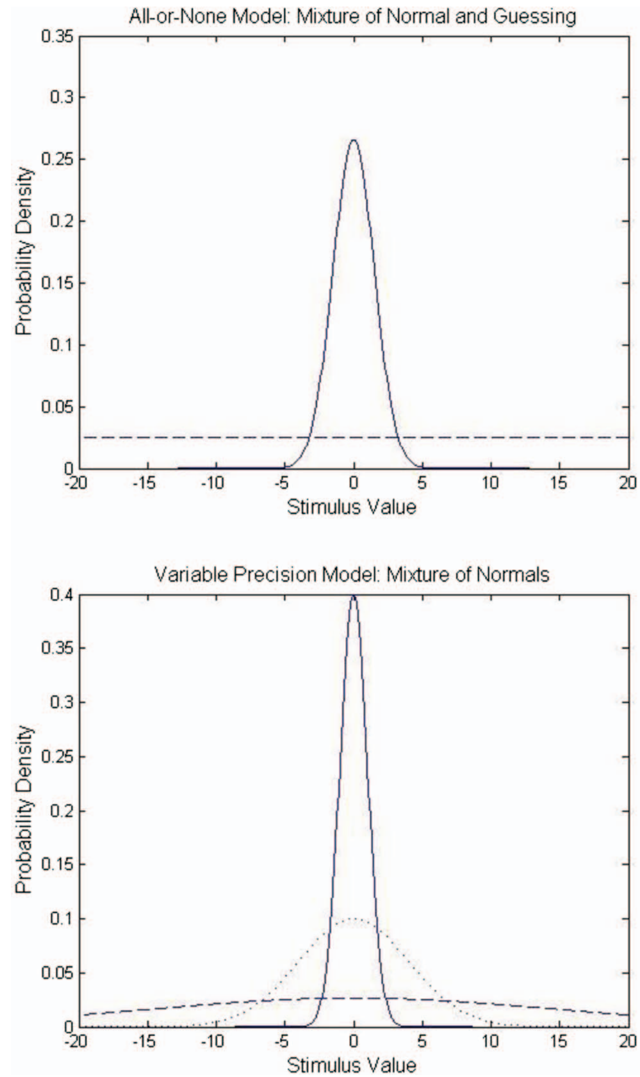

Figure 1. Schematic illustration of distribution of response errors produced by models in the continuous-reproduction task. Top panel: mixed-state memory-plus-guessing model. Bottom panel: variable-resources model. See the online article for the color version of this figure.

Goals of the Present Work

In the present work, our central goal was to develop a strong qualitative contrast between the predictions from important versions of the mixed-state and pure-continuous classes of models by making use of the alternative *change-detection* paradigm for the investigation of VWM (Luck & Vogel, 1997; Pashler, 1988). In the change-detection paradigm, the observer is presented with a specific value of a stimulus in a probed study location. The

[1] To reiterate, the intended contrast in this work is between mixed-state versus pure continuous models. Versions of shared-resources models can also yield mixed states, if, for example, there is presumed to be some lower threshold of resources below which a zero memory state arises. Using our present language, such a model would not be viewed as falling into the pure continuous class. Likewise, hybrid models assume combinations of mixed states arising from item limits and continuous variations in resolution arising from shared resources (e.g., Donkin et al., 2013; Nosofsky & Donkin, 2016; Sims et al., 2012; Swan & Wyble, 2014; Zhang & Luck, 2008). Obviously, evidence pointing to the role of mixed states supports hybrid models as well as all-or-none mixed-state ones.

observer's task is to judge whether the value at that location changed or stayed the same.

Specifically, the qualitative contrast that we pursue in this work is between versions of the mixed-state and continuous variable-resources (VR) models that we will refer to as *knowledge-limited* models of visual change detection. In these models, we presume that the observer has access to salient aspects of the experimental milieu, but not to highly detailed psychological and statistical processes that give rise to the memories themselves. For example, as is common in psychological theorizing, we will assume that the observer can adjust decision criteria in between-conditions fashion. For example, in conditions in which the objective probability of change trials is high, the observer may use lax criteria for making "change" judgments and may also tend to respond "change" if forced to guess. Such models also presume, of course, that the observer has access to the *outcome* of psychological processing involving individual items within a trial. For example, the observer will remember some specific value of the study item or will have knowledge that the item from the probed study location does not exist in memory at all. However, the models are knowledge-*limited* because they presume that observers do not have access to detailed hypothetical statistics associated with the underlying psychological and neurological processes that produced each individual-item memory representation in the first place. In particular, certain knowledge-*rich* versions of the models (see below) presume that observers have access—for each individual item in the memory set—to the standard deviation of the distribution of remembered values to which a mental process would give rise across trials of the experiment. For example, if the system allots some proportion p of its pooled resources to individual item i, then it has knowledge that, across trials of the experiment, the standard deviation of remembered values for such items is given by σ_p. The knowledge-limited models that we consider here will assume that observers do not have such access. In a nutshell, our experiment is designed to provide a strong qualitative contrast between the predictions from such knowledge-limited mixed-state and continuous VR models. To anticipate, our results will strongly favor the model from the mixed-state class.

Importantly, modern VR theorists will not be surprised by the shortcomings of the knowledge-limited versions of the continuous models. These researchers have in fact strongly endorsed knowledge-*rich* versions of the VR models and pointed to the shortcomings of knowledge-limited versions in their own work (e.g., Keshvari et al., 2012, 2013). Indeed, as will be explained later in our article, the favored model from this class is one that assumes that observers apply ideal-observer Bayesian decision rules to their knowledge-rich individual-item memory representations. Thus, the present work can be viewed as providing converging evidence to support the recent conclusions of the VR theorists.

However, our work will go beyond the recent demonstrations of the VR theorists in several respects. First, to date, the evidence in favor of the knowledge-rich VR models (compared with the knowledge-limited ones) has been based on the better quantitative fits yielded by the knowledge-rich versions. Although we will also evaluate the competing models in terms of their quantitative fit, the main contribution of the work involves the development of a version of the change-detection paradigm that yields a strong qualitative contrast between the predictions from the models. This approach will shine a stronger light on a core *reason* for the poorer fits of the knowledge-limited VR models, which, we believe, will advance the field's understanding of the properties of the models.

Second, in their recent work, Keshvari, van den Berg, and Ma (2013) compared versions of mixed-state and continuous VR models that all made the assumption that observers adopt ideal-observer decision rules. Whereas Keshvari, van den Berg, and Ma (2012) had shown that a knowledge-rich VR model outperformed knowledge-limited ones, Keshvari et al. (2013) showed in addition that the knowledge-rich VR model outperformed the representatives from the mixed-state class. Thus, Keshvari et al. (2013) concluded that "Our results suggest that working memory resource is continuous and variable and do not support the notion of an item limit" (p. 6). While acknowledging various hybrid possibilities, they went on to write "Our results, however, establish the VR model as the standard against which any new model of change detection should be compared" (p. 7). In the present work, we formulate a descriptive version of a model from the mixed-state class and compare its predictions to the ideal-observer VR model of Keshvari et al. (2013). To anticipate, although the knowledge-rich ideal-observer VR model does not suffer from the same qualitative shortcomings as the knowledge-limited version, we find that its quantitative predictions still fall short of those yielded by our descriptive mixed-state model.

Finally, in our General Discussion, we will argue that the knowledge-rich, ideal-observer assumptions that are embedded in the modern VR models can be questioned in terms of their psychological plausibility and internal consistency. To the extent that our concerns have merit, then the severe limitations of the knowledge-limited VR models compared to the mixed-state one provide powerful evidence in favor of the mixed-state class.

Qualitative Contrast

Having outlined the general issues addressed in this work, we now explain our approach to developing a sharp qualitative contrast between the predictions from the mixed-state and knowledge-limited VR models of visual change detection.

The first key to contrasting the predictions from the models is to include "big-change" trials within the design. For example, on a trial in which a particular studied color was "red," the test probe might be "green." On a psychological scale, the color green may be "miles away" from the color red in the context of the change-detection paradigm (see Model Analysis section for a formal statement).

From the perspective of mixed-state models, an event in which the observer responds "same" on a big-change trial is to be naturally expected. There will be some proportion of trials in which the observer has zero memory for the original study stimulus. On those trials, observers simply guess as to whether the stimulus changed or stayed the same. So long as the observer expects a reasonable number of trials on which study and test items are the same, then it is naturally expected that one will observe a significant proportion of "guess-same" responses, even on big-change trials.

From the perspective of VR models, however, the explanation of "same" responses on big-change trials is not so straightforward. It is true that a fundamental assumption of VR models is that there will be some proportion of trials in which minimal resources may be devoted to a particular item. Across such trials, the distribution of remembered values will have an extremely large variance, as depicted by the nearly flat normal distribution in the bottom panel of Figure 1. In the continuous-reproduction task, the responses produced on trials in which the flat normal operates will be very much like those produced by guessing in the mixed-state models.

Our key point, however, is that random draws from a nearly flat normal distribution seem very unlikely to produce a substantial number of "same" responses on big-change trials in the change-detection task. The situation is depicted schematically in the top panel of Figure 2. Let the original study stimulus have a reference value of zero. A big-change test probe (the solid circle) is depicted far from the original study stimulus. The distribution of between-trial remembered values in cases involving the nearly flat normal is also shown. Although the distribution has a very large variance, the chances that the *particular remembered value* that is selected from the normal distribution happens to be similar to the big-change stimulus are still minuscule. Thus, there would be little expectation that the observer would ever respond "same" on such big-change trials.

To allow the model to produce a reasonable proportion of "same" responses on such trials, one would need to assume that the observer adopts an extremely lax criterion for responding "same." The decision rule would be something like: "If the remembered value is anywhere close to the probed value, then respond 'same.'" This assumption is illustrated in the middle panel of Figure 2 by drawing criterion values that are far spread out from the big-change stimulus. Any remembered value that falls within the region defined by the criteria would be judged to be the same as the big-change test probe. A rationale for this assumption is that the observer may become aware during the course of the experiment that her memories for the studied items are sometimes highly variable, so a lax criterion needs to be used in order to perform the task.

The problem for the model is that *small-change* trials are also to be included in the task, and the inclusion of these trials is the second key to contrasting the predictions from the mixed-state and knowledge-limited VR models. The addition of the small-change trials is depicted in the bottom panel of Figure 2 (solid square). In line with the fundamental assumptions of the VR models, the figure also depicts a trial in which a high proportion of the resources is devoted to the study item, so the remembered value would now be a random draw from a normal distribution with a very small variance (the tall thin dashed distribution that is depicted in the figure). Indeed, the memory resolution that is assumed is sufficient to allow the observer to discriminate well above chance between *same* and *small-change* trials. Importantly, such discriminative capabilities are generally displayed by participants in the change-detection task. However, we have already assumed that the observer has adopted a lax criterion for making "same" responses (middle panel of Figure 2). The lax criterion is needed to allow the model to produce "same" responses on the big-change trials. The same lax criterion is also shown surrounding the small-change stimulus (bottom panel of Figure 2). It is apparent that even if memory resolution for the studied item were high, the use of the lax criterion would not allow the observer to successfully discriminate between *same* and *small-change* trials.

In a nutshell, it appears that the types of knowledge-limited VR models outlined above may have difficulty predicting a substantial proportion of *same* responses on big-change trials while simulta-

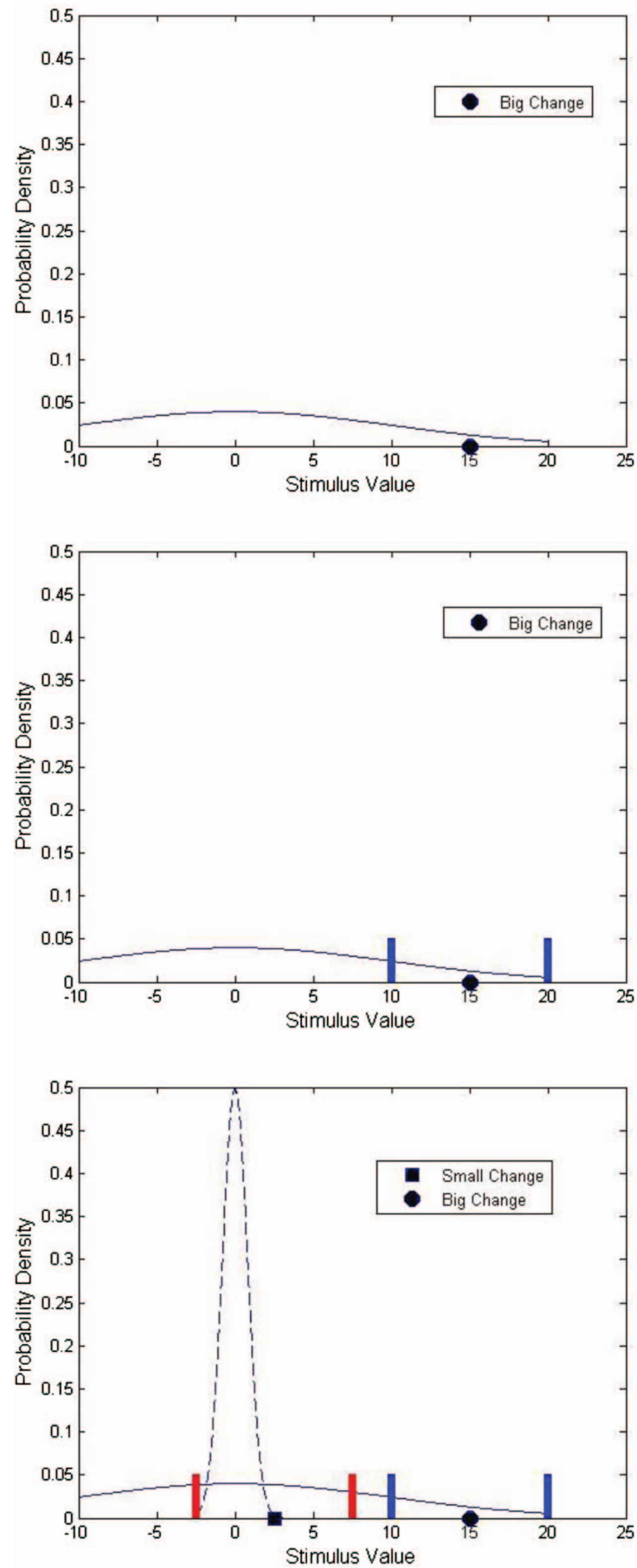

Figure 2. Schematic illustration of an application of a knowledge-limited variable-resources model to the change-detection task (see text for details). See the online article for the color version of this figure.

neously predicting successful ability to discriminate between *same* and *small-change* trials. The same dilemma is not faced by the mixed-state models. According to those models, *same* responses on big-change trials occur only when the subject is guessing. On many other trials, memory for the studied item will have been retained. Those memories will be sufficient to allow for above-chance discrimination of *same* and *small-change* trials.

Before proceeding to the experiment, we reiterate that the knowledge-rich VR models can account for the potentially challenging pattern of results as well. For example, if one assumes that: (a) the observer has knowledge of the amount of resources devoted to each individual item in the visual display; (b) knows the amount of between-trials variability in remembered values that is produced by the devoted resources; and (c) adjusts the criterion for responding "same" *for each individual item* in the memory set based on this known amount of individual-item variability, then such models can account for the qualitative pattern as well. It is an open question, however, how well extant versions of such models will fare in terms of their quantitative fit. In addition, in our view, the assumption that observers have detailed access to such knowledge for each individual item and apply ideal-observer decision rules with respect to such knowledge raises interesting questions about psychological plausibility that are worthy of open discussion and debate. We address such questions in our General Discussion.

Experiment

We conducted a visual change-detection experiment using as stimuli colors drawn from a color wheel. On each trial, subjects were briefly presented with a simultaneous visual display of two, five, or eight colored squares, with each colored square occupying a unique location of the display. Following a brief retention interval, a single location from the visual display was probed with a test color, and subjects judged whether the color in that location had changed or stayed the same. The test probe was either the same as the original color, a small change, or a big change. Across blocks of the experiment, we manipulated the objective probability with which change trials occurred (.3, .5, or .7). Subjects were informed of these objective change probabilities prior to each block. We expected the data from blocks with objective change probability (*cp*) equal to .3 to be the most diagnostic for discriminating between models. As developed in the Qualitative Contrast section of our introduction, on such blocks, subjects might respond *same* with high probability in cases in which they were guessing. Thus, even on big-change trials, there might be a high proportion of *same* judgments. If subjects simultaneously show above-chance discrimination between same and small-change trials under such conditions, we expect it will provide a severe challenge to the knowledge-limited VR models.

We should emphasize various novel components of our present experimental design relative to the related change-detection paradigms conducted by Keshvari et al. (2012, 2013). Each of our novel components was aimed at placing greater focus on the planned qualitative contrast between models. First, in Keshvari et al.'s (2012, 2013) paradigm, the magnitude of change that occurred on change trials was chosen randomly across the 360-degree circle of stimulus values. By contrast, as explained above, in our paradigm, change trials involved only small changes or big changes. This manipulation produced much larger sample sizes at the locations of the stimulus space that our paradigm-planning analyses indicated were most diagnostic for discriminating between the knowledge-limited mixed-state and continuous-VR models. Second, in Keshvari et al.'s (2012, 2013) paradigm, change trials always occurred with probability .5. By contrast, in our design, across blocks, change trials occurred with probability .3, .5, or 7. As explained earlier, the blocks with change probability equal to .3 were expected to be the most diagnostic for discriminating between the mixed-state and continuous-VR models, because they might promote a high probability of guessing "same" even on big-change trials. In addition, as will be seen, our modeling analyses will assume forms of selective influence of the change-probability manipulations on the values of the models' free parameters, thereby providing more diagnostic tests among the competing alternatives. Third, in Keshvari et al.'s (2012, 2013) paradigm, the test display involved probes at all locations from the study display; if the test trial was a change trial, then only a single location from the study display would change. Thus, the decision rules in Keshvari et al.'s (2012, 2013) models needed to integrate the likelihood of change across multiple probe locations, including all locations that had stayed the same. By contrast, in our design, only a single location was probed, so the decision rules assumed in the models pertained only to the likelihood of change at a single test location. Possibly, because in our design there is no role of memory noise across multiple locations in the models' decision rules, this aspect of our paradigm may also yield more focused tests of the key qualitative contrast of interest. Finally, we should note that each of our observers participated for 5,040 trials of testing (compared with 1,800 in the experiments of Keshvari et al., 2012, 2013). Beyond the larger sample sizes, we emphasize that our subjects were therefore highly experienced and extremely familiar with the structure of the experimental design. This factor will be an important consideration in evaluating the results of certain ideal-observer models applied to our data.

Method

Subjects

The subjects were six members of the Indiana University community who were paid for their participation. Each subject participated for 10 sessions, with each session lasting approximately 1 hr. Subjects were paid at the rate of $15 per session including a small bonus for good performance (average percent correct exceeding .70). The subjects all had normal or corrected-to-normal vision and all reported having normal color vision. None of the subjects was aware of the issues under investigation in the research.

Stimuli

The stimuli, similar to those described in Zhang and Luck (2008, 2009), were 180 colors that were evenly spaced around a circle in the $L^*a^*b^*$ color space (L = 50, a = 10, b = 10, with a radius of 40 units). The colors were presented as 30 × 30 pixel squares within a 200 × 200 pixel region centered on the computer screen. The background color of the screen was white.

All stimuli were generated using Matlab (version 7.1) on a single Apple iMac computer and displayed using the extensions provided by the psychophysics toolbox (Brainard, 1997) on a Sony Trinitron Multiscan 420GS CRT at a frame rate of 100 Hz (resolution: 1,024 × 768 pixels; size: 38.25 × 28.5 cm). The luminance and color calibration measurements were obtained using in-house software and a Photo Research PR-174 SecraScan radiometer. The maximum and minimum displayable luminances were 131.7 cd/m^2 and 0.02 cd/m^2, respectively. Viewing distance was approximately 57 cm and the visual angle of the individual squares was approximately .75° × .75°.

Procedure

On each trial, a set of colored squares was displayed simultaneously on the computer screen in random locations within the central rectangular region, subject to the constraint that the centers of all pairs of squares were at least 60 pixels away. The memory set size (number of squares in the display) was 2, 5, or 8, chosen randomly on each trial. The value of each individual study color from the color wheel (from 1 to 180) was chosen randomly on each trial.

On each trial, a single randomly chosen location from the study array was probed with a test square presented at that location. The test probe was either the same color as the original study square; a small-change color (24° away from the relevant study square); or a big-change color (180° away). On small-change trials, the direction of change was chosen at random.

Each of the 10 sessions of the experiment was divided into nine blocks of 56 trials each. We manipulated objective change probability across blocks: .3, .5, or .7. Each change-probability condition occurred once every three blocks in a random order. Within each block, if a trial was selected to be a change trial, then the degree of change (small = 24° or big = 180°) was chosen at random. (The number of small-change and big-change trials was not constrained to be equal in each individual block.)

Each trial started with the presentation of a fixation point (asterisk) at the center of the screen for 500 ms, followed by the display of the memory set for 500 ms. For Subjects 1–3, the screen then went blank for 500 ms, followed by the presentation of a location cue (open circle) for 200 ms, followed by another blank screen for 100 ms. For Subjects 4–6, the procedure was the same, except instead of presenting a location cue, colored pattern masks were presented at all study locations for 200 ms. The test probe was then presented and remained on the screen until the subject made a change or same judgment by pressing an appropriate button on the keyboard (J = change, F = same). Following a 500-ms blank interval, text feedback ("CORRECT!" or "INCORRECT") was provided for 1 s at the center of the screen. Following each block, subjects were informed of their overall percentage of correct responses.

Subjects were informed at the start of each block of the objective change probability operating during that block. The objective change probability was further emphasized by displaying a pie chart with the relative proportions of change and same trials. Subjects were instructed to adjust their response biases in accord with the objective change probability. Subjects were also instructed to rest their left and right index fingers on the F and J keys throughout each block and to press the appropriate key as soon as they made their same versus change judgment. Although we recorded subjects' response times, we did not analyze them here.

Results

As described in the Method section, three subjects performed the task under unmasked conditions whereas for the second three subjects we used masks. Even for the large memory set size, Subject 1 in the unmasked condition responded "change" with high probability on the big-change trials and also showed highly accurate responding in discriminating between *same* and *small-change* trials. (Subject 2 showed a similar pattern of results, although not as extreme as Subject 1.) We were concerned that these subjects were able to make use of forms of iconic memory to perform the task in the unmasked condition (Sperling, 1960). Therefore, we decided to test the three additional subjects (Subjects 4–6) in the masked condition. As it turned out, with the occasional exception of Subject 1, our model-based analyses yielded similar conclusions across the six subjects.

The complete set of change-same data matrices for the six subjects is reported in the Appendix. Although we test all candidate models at the individual-subject level, we start by reporting the results averaged across subjects in order to provide a sense of the main qualitative trends. Figure 3 displays the mean probability

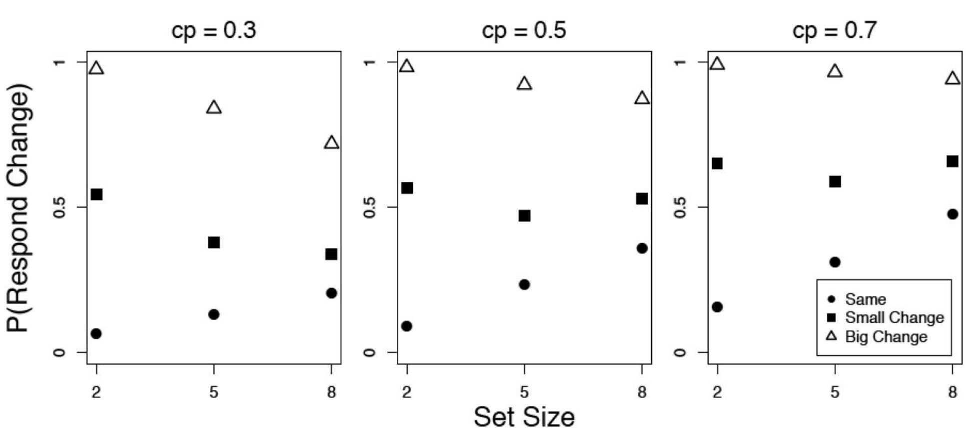

Figure 3. Mean respond-change probabilities as a joint function of stimulus type (same, small change, big change), memory set size, and objective change probability (*cp*).

with which observers responded "change" as a joint function of stimulus type (same, small change, big change), memory set size (2, 5, 8), and objective change probability (.3, .5, .7). Figure 4 averages across the different set sizes and shows how change-probability judgments varied with stimulus type and objective change probability. And Figure 5 averages across the different objective change-probability conditions and shows how change-probability judgments varied with stimulus type and memory-set size.

As is clear from inspection, the probability of a change response was greatest on big-change trials, intermediate on small-change trials, and least on same trials (Figures 3–5). Change responses grew systematically more frequent with increases in objective change probability (see Figure 4). "Hits" (responding change on change trials) tended to decrease with increases in memory set size, whereas "false alarms" (responding change on same trials) increased with increases in memory set size (see Figure 5). The most important result is that by the time one reaches set-size 8, there is a substantial probability of subjects failing to respond *change* to the big-change stimuli, particularly in the $cp = .3$ condition but also in the $cp = .5$ condition. These big-change "misses" are observed at the same time that subjects show above-chance ability to discriminate between the *same* and *small-change* trials (see below). As discussed extensively in our introduction, it is this pattern of results that we expect will provide a challenge to the class of knowledge-limited VR models.

We conducted statistical analyses to confirm our observation that participants discriminated at above-chance levels between the *same* and *small-change* trials in the set-size-8 condition. In one analysis, for each individual subject, we collapsed across the three change-probability conditions in the set-size-8 condition, and con-

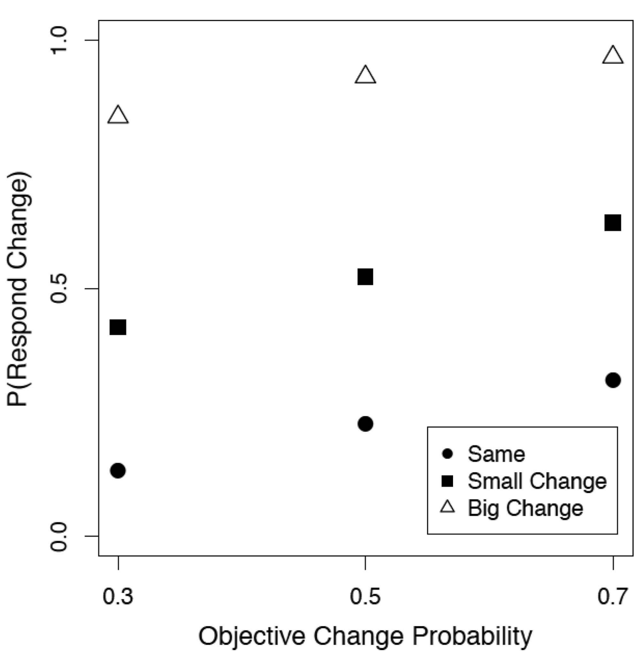

Figure 4. Mean respond-change probabilities as a joint function of stimulus type (same, small change, big change) and objective change probability. Data are collapsed across the different set-size conditions.

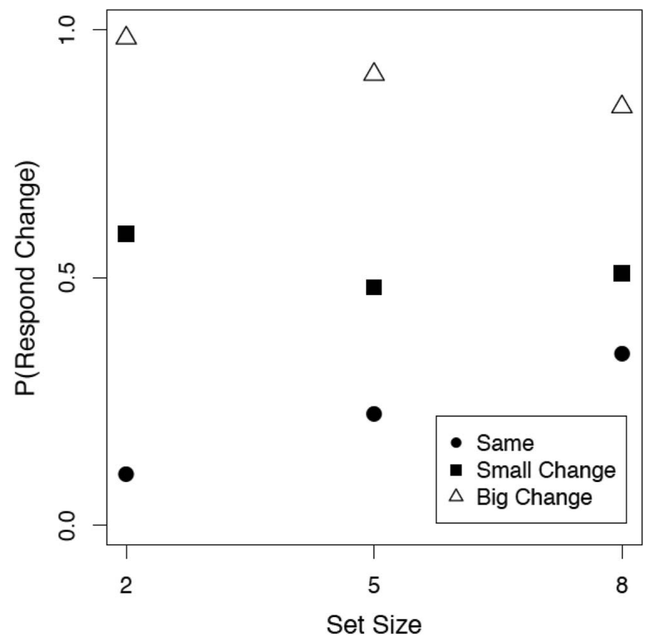

Figure 5. Mean respond-change probabilities as a joint function of stimulus type (same, small change, big change) and memory set size. Data are collapsed across the different objective-change probability conditions.

structed 2 × 2 contingency matrices in which the rows corresponded to trial type (small-change vs. same) and the columns corresponded to response type (change vs. no-change). For each individual subject, we conducted a chi-square test for the independence of the distribution of response frequencies across the two trial types. The results of these tests were significant at the $p = .001$ level for all six subjects; average $\chi^2(1, N = 1{,}202) = 72.2$. Even restricting the analysis to just the $cp = .3$ condition (in which the big-change *misses* were at their highest, but there is a greatly reduced total sample size), the results were still significant at the $p = .05$ level for all six subjects; average $\chi^2(1, N = 469) = 7.79$. Finally, again restricting consideration to only the highly diagnostic $cp = .3$ condition (set-size 8), we computed difference scores between hit rates on the *small-change* trials and false-alarm rates on the *same* trials for all six subjects. The mean difference score was significantly greater than zero, $t(5) = 8.19$, $p < .001$, further confirming that the subjects had discriminated with above-chance accuracy between the *small-change* and the *same* trials in the $cp = .3$, set-size-8 condition.

Formal Modeling Analyses

We divide the presentation of the formal modeling analyses into two sections. In the first section, we compare the mixed-state model to the knowledge-limited VR model. In the second section, we consider variants of knowledge-rich VR models with ideal-observer decision rules.

Knowledge-Limited Models

Mixed-state memory-plus-guessing model (MS). According to the baseline version of the mixed-state model, the probed study item resides in memory with probability p_{mem}, which is a function only of memory set size (ss = 2,5,8). If the study item is in memory, then the remembered value (v) is a random draw from a circular normal distribution with mean centered on the true value of the study item and standard deviation σ_M. Without loss of generality, the reference value for the mean of the memory distribution is set at zero. The distance (d) between the remembered value and the test probe is given by $d = |v - \mu|$, where $\mu = 0$ if the test probe is the same as the study item; $\mu = 1$ if the test probe is a small-change stimulus; and $\mu = 7.5$ if the test probe is a big-change stimulus. (Based on the $L^*a^*b^*$ color-space system, we assume that the big-change stimulus that is 180° from the study item is 7.5 times the distance as the small-change stimulus that is 24° from the study item.) In cases in which the remembered value of the study item "wraps around" the color wheel (i.e., is more than 180° from the test probe), the distance d is defined as the shorter distance to the test probe, thereby implementing the circular structure of the color space. The decision rule is to respond "change" if the memory distance d exceeds a criterion C.[2] The criterion is allowed to depend jointly on objective change probability (cp = .3, .5, .7) and memory set size (ss = 2, 5, 8):

$$C(cp, ss) = c_1(cp) + c_2(ss), \tag{1}$$

where $c_1(cp)$ and $c_2(ss)$ are freely estimated parameters. (Without loss of generality, $c_2(2) = 0$.) Presumably, as objective change probability increases, observers will set a more lax criterion (c_1) for making "change" judgments. Because previous evidence suggests that observers may also adjust response thresholds with changes in set size (e.g., Donkin et al., 2013), the parameter c_2 is allowed to vary as well.

If the study item from the probed location is no longer in memory, which occurs with probability $1\text{-}p_{mem}(ss)$, the subject is forced to guess. The guess-change probability, $g(cp)$, is presumed to depend solely on objective change probability. Thus, the model estimates free parameters $g(.3)$, $g(.5)$, and $g(.7)$.

In sum, the mixed-state model estimates 12 free parameters: three memory-state probabilities p_{mem} (one for each set size); a memory standard deviation parameter σ_M; three criterion-related parameters related to objective change probability (c_1); two free criterion-related parameters related to set size (c_2); and three guess-change parameters g.

Knowledge-limited variable-resources (KLVR) model. In the VR models, all study items are presumed to be in memory, so there are no memory-probability or guess-change parameters. The key idea is that because the observer splits the pool of resources among the items in the memory set, and the assignment of resources may differ across individual items, the remembered values are drawn from memory distributions with differing standard deviations. In these models, we followed van den Berg, Shin, Chou, George, and Ma (2012) and Keshvari et al. (2013) by assuming that the "precision" J associated with each item-memory distribution was a random draw from a gamma distribution. The standard deviation of each item distribution is then given by $\sigma_M = \sqrt{1/J}$. To provide the model with some flexibility, we allowed the scale and shape parameters (α and β) of the sampled gamma distribution to vary freely with changes in memory set size. Thus, allowance was made for the item distributions associated with larger memory set sizes to tend to have greater standard deviations than those associated with smaller set sizes; however, the individual item distributions associated with a particular set size would themselves have differing standard deviations, making the memory distributions doubly stochastic (cf. van den Berg et al., 2012). The other assumptions for the KLVR model were the same as already explained for the mixed-state model: In particular, if the sampled distance d exceeded $C(cp,ss) = c_1(cp) + c_2(ss)$, then the observer responded "change," and otherwise responded "same." This KLVR model estimates 11 free parameters: three shape (α) and scale (β) parameters for the sampled gamma distributions (a separate pair of shape and scale parameters for each memory set size); three criterion-related parameters related to objective change probability (c_1); and two free criterion-related parameters related to set size (c_2).

Knowledge-limited variable-resources model with mixed criteria (KLVR-MC). Although the knowledge-limited models presume that the observer does not have access to the variability of the between-trials distribution from which each individual item memory is drawn, perhaps the observer does have general knowledge that her memory representations have variable precision across items and trials. Thus, the observer may apply a mix of different criterion settings across items and trials, even within a given experimental condition (defined by objective change probability and set size). To test this possibility, we formulated an extended version of the KLVR model that allowed mixed criterion settings. We formalized the idea by randomly sampling on each trial a criterion-noise parameter κ drawn from a gamma distribution with free shape and scale parameters α_κ and β_κ. The criterion used on each trial was then given by

$$C(cp, ss) = [c_1(cp) + c_2(ss)] + \kappa, \tag{2}$$

where $c_1(cp)$ and $c_2(ss)$ are as defined previously. In all other respects, the KLVR-MC model was the same as the KLVR model. The KLVR-MC model adds two free parameters to the KLVR model, for a total of 13 free parameters.

Model-fitting method. We generated predictions from all models by using computer simulation.[3] For each set of candidate parameters, we used 50,000 simulated trials for each individual cell of the change-same data matrices. We used the Hooke and Jeeves (1961) algorithm as a computer-search algorithm for locating the best-fitting parameters. To guard against local minima, each model was fitted to each individual subject's data by conducting 25 separate computer searches with different random starting configurations of the parameters in each search.

We fitted all models to each individual subject's data by using a maximum-likelihood criterion. Specifically, we conducted computer searches for the free parameters that maximized the likelihood function

$$L = \prod_{c,ss,cp} N_C(c, ss, cp)P(C|c, ss, cp)^{F(C|c,ss,cp)} \times [1 - P(C|c, ss, cp)]^{F(S|c,ss,cp)}.$$

[2] The versions of the mixed-state and VR models that we report in this article used deterministic decision rules: The rule was to respond "change" if the memory distance d exceeded a criterion C. We also formulated parallel versions of the models that used a probabilistic choice rule. The models that used probabilistic choice rules tended to provide slightly worse fits than the versions we report here. Again, however, the mixed-state model yielded far better fits than did the knowledge-limited VR models.

[3] To avoid computer overflow and underflow problems, the lower limit on a sampled precision J was set at .001.

where the product is taken across all combinations of the factors *change* type (c = same, small-change, big-change), set size (ss), and objective change probability (cp); $P(C|c,ss,cp)$ denotes the model's predicted probability of change judgments at that combination of factors; $F(C|c,ss,cp)$ is the observed frequency of change judgments; $F(S|c,ss,cp)$ is the observed frequency of same judgments; and N_C denotes the binomial coefficient of C change judgments from $N = F(C)+F(S)$ total observations.

The fits of the models were evaluated by transforming the likelihood (L) values into Akaike Information Criterion (AIC) values:

$$\text{AIC} = -2 \cdot \ln L + 2n_p,$$

where n_p is the number of free parameters used by the model. The term $2n_p$ is a penalty term for use of free parameters. The model that achieves a smaller AIC is considered to provide a more parsimonious account of the data. A commonly used alternative fit criterion is the Bayesian Information Criterion (BIC), which uses a different penalty term. Although use of the AIC versus BIC had no bearing on our conclusions involving the knowledge-limited VR models, the issue became relevant in comparing the mixed-state model to the knowledge-rich VR models. Model-recovery analyses (reported in the second part of our model-fitting analyses) pointed clearly to the AIC as the better criterion under the present conditions.

Model-fitting results. The fits of the present models to each individual subject's data are reported in the first three columns of Table 1. As can be seen, overall, the mixed-state model yields far better AIC scores than either of the knowledge-limited VR models.

To gain some understanding of the basis for these model-fitting results, we consider the predicted and observed change probabilities for a representative subject (Subject 4) in some detail. Table 2 and Table 3 report the predictions from the mixed-state and knowledge-limited VR model, respectively. For ease of comparison, the observed change probabilities are presented in all tables.

Inspection of Table 2 reveals qualitative patterns of data for the single subject that mirror those seen in the averaged data that we displayed in Figure 3. First, note that when set size is equal to 2, the subject is essentially perfect in detecting the big-change trials. The subject also shows performance that is well-above chance in discriminating between the same and small-change trials. As set size grows, however, the subject begins to false-alarm (respond "change") on same trials and to miss (respond "same") on change trials. Indeed, in the $cp = .3$ condition, at set-size 8 the subject's hit rate is only .79 on the big-change trials. The substantial proportion of misses on these big-change trials is observed at the same time as the subject continues to show above-chance discrimination on the same versus small-change trials; $\chi^2(1, N = 463) = 4.39$, $p < .05$.

As can be seen in Table 2, the mixed-state model comfortably fits the data across the manipulations of change type, set size, and objective change probability. To understand the basis for these predictions, we report the best-fitting parameter estimates from the mixed-state model for the six subjects in Table 6. The pattern of Subject 4's parameter estimates is representative of those of the other subjects. First, note that as set size increased, the probability that the probed study item was retained in memory decreased (i.e., the p_{mem} parameter decreases systematically with increases in memory set size). Thus, the subject is forced to guess as set size grows larger. Furthermore, as can be seen in Table 6, the subject adjusted her guessing strategies (g) across the different objective-change-probability conditions. In conditions in which change-probability was low ($cp = .3$), the subject often guessed "same." This guessing strategy produced a high proportion of misses on big-change trials in the set-size-8 condition. At the same time, because the probed study item was also sometimes retained in memory in the set-size-8 condition, the subject continued to be able to discriminate with above-chance accuracy between same and small-change trials even in this large set-size condition.

By contrast, as can be seen in Table 3, the KLVR model has difficulty fitting the patterns of data. Recall that the challenge for the KLVR models is to simultaneously predict misses on the big-change trials while maintaining above-chance discrimination between the same and small-change trials. For Subject 4, the KLVR model apparently adopted parameter settings that allowed it to fit the big-change data by making use of lax criteria for responding "same." However, the use of lax criteria then prevents the model from being able to adequately discriminate between the same and small-change trials, so it greatly underpredicts the difference between hit and false-alarm rates for these trial types. In a follow-up analysis (see Table 4), we searched for parameter settings from the model that optimized its fits to just the same and small-trials data across the different experimental conditions. For those parameter settings, the model incorrectly predicted nearly 100% hit rates on the big-change trials across all memory set sizes. In still other analyses, we fitted the mixed-criterion version of the KLVR model to the data (third column of Table 1); with the

Table 1
−lnL and AIC Fits of the Models to the Individual Subject Data

Subject/model	Mixed state	KLVR	KLVR-MC	KRVR-BIO	KRVR-scp
1	76.2, 176.4	87.0, 196.0	69.8, **165.6**	349.5, 707.0	94.1, 200.2
2	69.9, **163.7**	110.8, 243.7	78.8, 183.7	322.5, 652.9	79.3, 170.5
3	73.9, **171.7**	111.5, 245.1	111.5, 249.1	454.8, 917.6	99.8, 211.6
4	70.4, **164.9**	117.9, 257.8	94.7, 215.5	740.9, 1489.8	100.9, 213.8
5	76.3, **176.6**	204.0, 429.8	200.2, 426.4	366.4, 740.9	80.9, **173.8**
6	67.8, **159.6**	157.6, 337.3	102.7, 231.5	341.9, 691.9	94.0, 199.9
Summed AIC	**1012.9**	1796.7	1471.8	5200.1	1169.8

Note. The first entry in each cell is the value of $-lnL$ (negative log-likelihood) and the second entry is the value of AIC (Akaike's Information Criterion). KLVR = knowledge-limited variable resources; KLVR-MC = knowledge-limited variable resources with mixed criteria; KRVR-BIO = knowledge-rich variable resources with biased ideal-observer decision rule; KRVR-scp = knowledge-rich variable resources with freely estimated subjective-change probabilities. The minimum AIC for each subject is highlighted in boldface font. (Multiple entries are highlighted in cases in which the AIC difference is less than five.)

Table 2
Mixed-State Model: Predicted Proportion of Change Judgments in Each Condition for Subject 4

Change type	Memory set size SS = 2	SS = 5	SS = 8
	Condition cp = .3		
Same	(.17)	(.21)	(.25)
	.17	.21	.23
Small change	(.50)	(.37)	(.35)
	.50	.41	.34
Big change	(.98)	(.80)	(.72)
	.98	.81	.79
	Condition cp = .5		
Same	(18)	(.29)	(.36)
	.18	.28	.37
Small change	(.52)	(.46)	(.47)
	.53	.47	.43
Big change	(.99)	(.88)	(.83)
	.99	.88	.84
	Condition cp = .7		
Same	(.27)	(.35)	(.43)
	.27	.35	.48
Small change	(.62)	(.56)	(.57)
	.61	.54	.59
Big change	(.99)	(.92)	(.89)
	.99	.91	.88

Note. SS = set size; cp = objective change probability. Top entry in each cell = predicted proportion, bottom entry in each cell = observed proportion.

exception of Subject 1, however, it continued to perform poorly relative to the mixed-state model. As explained in our introduction, to capture in detail the key qualitative pattern in the data, the mixing of criteria needs to be specific to whether the remembered study value was produced by a low-variability or high-variability process. In the present knowledge-limited models, the mixing of criteria is instead randomly decided across trials.

One approach to saving the knowledge-limited VR model might be to posit that "same" responses on the big-change trials occur because of attentional lapses on the part of the subjects (e.g., see Rouder et al., 2008). Alternatively, the big-change misses might arise due to positional uncertainty in which the observer compares the test probe to the incorrect study item (e.g., Bays et al., 2009), or to forms of obligatory averaging of item representations (for a review, see Dubé & Sekuler, 2015). Note, however, that subjects virtually never missed on the big-change trials in the set-size-2 condition (see Figures 3 and 5 and the individual-subject data in the Appendix). Thus, one would need to argue that attentional lapses, positional uncertainty, or perceptual averaging occurred only in the larger set size conditions, which places strain on such explanations. We should note in addition that the proponents of VR models of change detection (Keshvari et al., 2012, 2013) did not incorporate parameters related to attentional lapses, positional uncertainty, or perceptual averaging in their own modeling.

In sum, the focused qualitative contrast embedded in our design led to a decisive advantage in quantitative fit of the mixed-state model compared to the knowledge-limited VR model of visual change detection.

Knowledge-Rich Variable-Resources Models

As noted in our introduction, Keshvari et al. (2012, 2013) have already reported quantitative advantages in the fit of a knowledge-*rich* VR model that uses an ideal-observer decision rule. One purpose of the present work was to provide converging tests of Keshvari et al.'s (2012, 2013) findings in our modified paradigm. In addition, although we expected knowledge-rich versions of the VR model to be able to account for the main qualitative data pattern that was the focus of our design, it was an open question how such models would fare in quantitative comparisons with our mixed-state model. We pursue these questions in the present section.

Knowledge-rich variable-resources model with ideal-observer decision rule. In formulating the knowledge-rich versions of the models, we attempt to stay as close as possible to the versions of the models as actually proposed by the VR theorists (see Keshvari et al., 2012, 2013). First, the knowledge-rich component of the models involves the assumption that the observer has access to the standard deviation of the between-trials distribution from which each individual-item memory representation is drawn. Note that this standard deviation will vary randomly across different memory set sizes, trials, and individual items within each trial. Second, van den Berg et al. (2012) and Keshvari et al. (2012, 2013) introduced constraints on the form of the gamma distributions from which memory precisions (and the resulting individual-item standard deviations) are sampled: Let J_1 denote the mean of the gamma distribution (with scale parameter τ) associated with

Table 3
Knowledge-Limited Variable-Resources Model: Predicted Proportion of Change Judgments in Each Condition for Subject 4

Change type	Memory set size SS = 2	SS = 5	SS = 8
	Condition cp = .3		
Same	(.16)	(.28)	(.32)
	.17	.21	.23
Small change	(.28)	(.31)	(.34)
	.50	.41	.34
Big change	(.99)	(.85)	(.82)
	.98	.81	.79
	Condition cp = .5		
Same	(.23)	(.34)	(.38)
	.18	.28	.37
Small change	(.48)	(.37)	(.41)
	.53	.47	.43
Big change	(.99)	(.88)	(.86)
	.99	.88	.84
	Condition cp = .7		
Same	(.31)	(.39)	(.43)
	.27	.35	.48
Small change	(.67)	(.42)	(.45)
	.61	.54	.59
Big change	(.99)	(.90)	(.88)
	.99	.91	.88

Note. SS = set size; cp = objective change probability. Top entry in each cell = predicted proportion, bottom entry in each cell = observed proportion.

Table 4
Knowledge-Limited Variable-Resources Model That Optimizes Fit to the Same and Small-Change Trials: Subject 4

	Memory set size		
Change type	SS = 2	SS = 5	SS = 8
	Condition *cp* = .3		
Same	(15) .17	(.24) .21	(.27) .23
Small change	(.43) .50	(.38) .41	(.42) .34
Big change	(1.00) .98	(.98) .81	(1.00) .79
	Condition *cp* = .5		
Same	(.21) .18	(.27) .28	(.33) .37
Small change	(.53) .53	(.45) .47	(.49) .43
Big change	(1.00) .99	(.98) .88	(1.00) .84
	Condition *cp* = .7		
Same	(.30) .27	(.31) .35	(.41) .48
Small change	(.63) .61	(.57) .54	(.57) .59
Big change	(1.00) .99	(.98) .91	(1.00) .88

Note. SS = set size; *cp* = objective change probability. Top entry in each cell = predicted proportion, bottom entry in each cell = observed proportion.

memory set-size 1. Then the mean precision at memory set-size N is given by the power function $J_N = J_1 N^{-\alpha}$. Third, the observer is presumed to use an ideal-observer Bayesian decision rule to determine if the test probe is most likely to have arisen from a change trial or a same trial (given the remembered values of the probed study items and the known standard deviations of the item-memory distributions). The observer responds with the trial type that has higher likelihood. The precise form of the decision rule was different in Keshvari et al.'s (2012, 2013) application than the one we use here, because the structure of our paradigms differed. However, the conceptual underpinnings of the models are the same.

Specifically, in our application of the model, we compute the posterior probability that the test probe was most likely to have been generated from a same trial (p_s), a small-change trial (p_{sc}), or a big-change trial (p_{bc}). The observer makes a change judgment if the value $p_{sc} + p_{bc}$ exceeds the value p_s. The true ideal-observer version of the model makes use of only three free parameters: J_1, τ, and α. However, following Keshvari et al. (2013), we allow the assumed change probability operating in the experiment to be a free parameter as well. An interpretation is that subjects may have misestimated the objective change probability in applying the ideal-observer decision rule, so each individual subject was granted a subjective change-probability parameter. Recall that in our paradigm, there were separate objective change probabilities operating across blocks (.3, .5, and .7). There seemed to be two reasonable approaches to translating Keshvari et al.'s (2012, 2013) subjective change-probability approach to our paradigm. First, subjects may simply have an overall bias to respond "change" that systematically modulates the objective change probabilities. For this model, the subjective-change-probability (*scp*) in each block is assumed to be given by $scp = cp + b \cdot (1\text{-}cp)$, where *cp* is the objective change probability in the block and b ($0 \leq b \leq 1$) is a "respond-change" bias parameter. Second, the subjective change probabilities may be only monotonically related to the objective change probabilities, so the value of *scp* is allowed to be a free parameter for each of the three types of objective change-probability blocks. We refer to the first version of the model, which uses a total of four free parameters, as the knowledge-rich VR model with a biased ideal-observer decision rule (KRVR-BIO). The second version, which uses six free parameters, is termed the subjective change-probability version (KRVR-scp).

Model recovery analyses. Because the knowledge-rich models with ideal-observer decision rules have fewer free parameters than do the mixed-state and knowledge-limited VR models, the fit statistic for comparing the models takes on great importance. To investigate the issue, we conducted the following model-recovery analyses. First, using the best-fitting parameters for each individual subject from the mixed-state model, we generated 20 simulated data sets for each subject (120 total data sets). Analogously, using the best-fitting parameters from the KRVR-scp model, we generated 120 simulated data sets from that model. Each simulated data set had the same total number of observed trials in each condition as did the real subjects. We then fitted each of the models to each simulated data set. When using AIC as the criterion of fit, we found that the mixed-state model was correctly recovered for 91% of the data sets that it had generated, whereas the KRVR model was correctly recovered for 94% of the data sets that it had generated. By comparison, when using BIC as the criterion of fit, we found that the mixed-state model was correctly recovered for only 22% of the data sets that it had generated, whereas the KRVR model was correctly recovered for 100% of the data sets that it had generated. It appears, therefore, that the BIC places far too great a penalty on number of free parameters under the present conditions. By contrast, the AIC does an excellent job of recovering the true generating model. Therefore, we used AIC as our model-evaluation statistic.

Model-fitting results. The fits of the knowledge-rich models are reported in the final two columns of Table 1. The version of the model that assumes a single respond-change bias across blocks (KRVR-BIO) fits the data extremely poorly for all subjects. Inspection of the individual-subject predictions from that model revealed that it greatly overestimated the probability of change judgments in the *cp* = .7 condition, and generally underestimated the probability of change judgments in the *cp* = .3 condition. Applications of the version of the KRVR model with freely estimated subjective change probabilities (see below) confirm the shortcomings associated with the biased ideal-observer version. In particular, averaged across subjects, the estimated subjective change probabilities across the *cp* = .3, *cp* = .5, and *cp* = .7 conditions were .499, .513, and .562, respectively. Thus, if one interprets performance from the perspective of the KRVR models, then it appears that observers do not come close to employing the true objective change probabilities in their decision rules (despite having been told the objective probability with which changes occurred). Although the decision rule may have the form of an ideal observer, the operating parameters of the decision rule depart dramatically from those of an ideal observer.

As reported in the final column of Table 1, the version of the knowledge-rich VR model that allows freely estimated subjective change probabilities for each type of block does a far better job of fitting the data than the version with a single response-bias parameter. Indeed, for all subjects except Subject 1, the model yields far better fits than the knowledge-limited VR models that we considered in the previous section of our article (compare with columns 2 and 3 of Table 1). This result converges with the findings of Keshvari et al. (2012), who also found that knowledge-rich VR models performed better than knowledge-limited versions in their paradigm. The predicted change probabilities from the KRVR-scp model are shown along with the observed change probabilities for Subject 4 in Table 5. Unlike the knowledge-limited models, the KRVR-scp model is able to account for the challenging data pattern in the matrix, namely the subject's high probability of responding "same" on the big-change trials along with her ability to discriminate between the same and small-change trials.

Although the KRVR-scp model fits the data considerably better than do the knowledge-limited models, note that it is still the case that its quantitative fits are generally substantially worse than those of the mixed-state model—compare model fits in columns 1 and 5 of Table 1. We suspect that these quantitative shortcomings do not reflect limitations in the core properties of the KRVR-scp model. For example, by allowing more flexibility in the parameters of the gamma distributions from which precision values are sampled, it seems likely that the model could achieve improved quantitative fits.[4]

Table 5
Ideal-Observer Knowledge-Rich Variable-Resources Model With Freely Estimated Subjective Change Probabilities: Predicted Proportion of Change Judgments for Subject 4

Change type	Memory set size SS = 2	SS = 5	SS = 8
	Condition cp = .3		
Same	(14)	(.22)	(.26)
	.17	.21	.23
Small change	(.43)	(.40)	(.40)
	.50	.41	.34
Big change	(.91)	(.82)	(.77)
	.98	.81	.79
	Condition cp = .5		
Same	(.16)	(.28)	(.35)
	.18	.28	.37
Small change	(.45)	(.46)	(.50)
	.53	.47	.43
Big change	(.93)	(.88)	(.86)
	.99	.88	.84
	Condition cp = .7		
Same	(.20)	(.34)	(.43)
	.27	.35	.48
Small change	(.50)	(.53)	(.57)
	.61	.54	.59
Big change	(.96)	(.94)	(.94)
	.99	.91	.88

Note. SS = set size; cp = objective change probability. Top entry in each cell = predicted proportion, bottom entry in each cell = observed proportion.

Table 6
Best-Fitting Parameters From the Mixed-State Model

Parameter	Subject number S1	S2	S3	S4	S5	S6	Mean
$p_{mem}(2)$	.959	.982	.898	.962	.965	.988	.959
$p_{mem}(5)$	.839	.838	.544	.635	.538	.834	.705
$p_{mem}(8)$	.637	.680	.336	.490	.275	.624	.507
$g(.3)$	.807	.566	.282	.449	.263	.406	.462
$g(.5)$	.888	.782	.708	.672	.764	.477	.715
$g(.7)$	.942	.863	.907	.779	.950	.860	.884
$c_1(.3)$	.980	.807	1.542	.995	1.005	.824	1.026
$c_1(.5)$	.882	.696	1.498	.984	1.002	.836	.983
$c_1(.7)$	.816	.697	1.068	.815	.861	.726	.831
$c_2(2)$	[0]	[0]	[0]	[0]	[0]	[0]	[0]
$c_2(5)$	.171	.198	−.070	.307	.251	.376	.206
$c_2(8)$	.183	.315	.016	.483	.222	.416	.273
σ_m	.399	.421	.663	.705	.559	.358	.518

Note. S1 through S6 denotes Subjects 1 through 6. Values in brackets were held fixed at zero a priori.

Nevertheless, the version of the model applied here is directly analogous to the one advanced by Keshvari et al. (2013). In their article, these researchers pointed to the clear superiority of the KRVR model compared with certain versions of mixed-state models that were constrained to use ideal-observer Bayesian decision rules. By contrast, we obtain the opposite pattern of results using our descriptive version of the mixed-state model that is not constrained by such rules. At the least, these findings reopen the debate and point to the need for further research concerning the relative merits of mixed-state versus pure continuous models of visual change detection. In our view, however, a more fundamental question is whether the knowledge-rich assumption that is needed to allow the VR models to even be competitive is a psychologically plausible one. We pursue this question in some depth in our General Discussion.

General Discussion

Summary

Knowledge-limited mixed-state and variable-resources models. Much of the recent work that has compared mixed-state and variable-resources (VR) models of visual working memory has focused on performance in the continuous-reproduction task. In the context of that task, random draws from a nearly flat normal

[4] We do not explore such possibilities in the present article, however, because such changes to the assumed nature of the stimulus representation would have bearing on conclusions from numerous other studies beyond the one presented here. In particular, the VR theorists have incorporated the present assumptions involving how the parameters of the gamma distribution are related to stimulus set size in numerous previous articles (e.g., Keshvari et al., 2012, 2013; van den Berg et al., 2014; van den Berg, Shin, et al., 2012), and their conclusions regarding model selection and evaluation were based on those assumptions. Thus, it seems more appropriate for the VR theorists themselves to decide which routes of increased flexibility are most promising and would yield conclusions consistent with those reached in their numerous previous applications and tests of the model.

distribution (arising from low-precision memory) are very much like random draws from a uniform distribution (arising from an absence of memory and reliance on guessing), so the alternative models make similar qualitative predictions for that task (see Figure 1).

The primary purpose of the present work was to break this form of model mimicry in the context of the alternative change-detection task. In particular, we developed a version of the change-detection paradigm that yielded a strong qualitative contrast between the predictions from knowledge-limited forms of mixed-state and VR models. As described earlier, by "knowledge-limited," we mean that the observer is presumed to have access to the outcome of the psychological processes underlying visual working memory (e.g., the remembered value of a studied stimulus), but not to detailed hypothetical statistics associated with the underlying psychological and neurological processes that produced each individual-item outcome in the first place.

Our paradigm was highly diagnostic because it combined numerous presentations of both big-change and small-change trials, and also manipulated objective change probabilities across conditions. In the case in which there is a big-change trial, the mixed-state (memory-plus-guessing) model provides a natural explanation for the occurrence of "same" responses: The original study stimulus may simply be absent from memory, so it is a complete guessing game whether the test probe is the same or different from the original study item. And in conditions in which the observer expects a high proportion of "same" trials, she will often guess "same." By contrast, the explanation of same responses on big-change trials from the knowledge-limited VR models is not so straightforward. It is true that, according to the VR model, few resources may have been devoted to the probed study item, so the remembered value will be a random draw from a high-variability distribution. However, according to those models, *some* specific value is being remembered. The chances that the specific value is similar to the big-change test probe are still minuscule (Figure 2, top panel). To predict a significant proportion of same responses, the model would need to assume that observers are also adopting extremely lax criteria for judging "same" (Figure 2, middle panel); but this assumption then upsets the model's predictions of the ability of observers to successfully discriminate between same and small-change trials (Figure 2, bottom panel).

We verified these intuitive challenges by competitively testing the ability of formalized versions of the knowledge-limited mixed-state and VR models to quantitatively fit our big-change and small-change data. In line with the intuitive argument developed above, the mixed-state model provided a simple and natural account of the complete set of results, with best-fitting parameters varying in systematic and easy-to-interpret ways. By contrast, the knowledge-limited VR model failed dramatically to fit the data and was stymied by the qualitative challenge outlined above.

One might argue that our characterization of the mixed-state model as "knowledge-limited" is misleading because the model "knows" to guess when a probed item is in the absence-of-memory state. In our view, however, this property of the model involves an extremely weak form of "knowledge." The absence of memory at a given test location is an *outcome* of the processing that took place on a given trial. If a test probe is presented at a location for which memory of the study item is completely absent, then because there is nothing there, the observer is forced to guess.

Knowledge-rich VR models. Variable-resources theorists have previously argued in favor of knowledge-*rich* versions of the models compared to knowledge-limited ones (e.g., Keshvari et al., 2012). In the knowledge-rich models, the observer is presumed to have access to information such as the proportion of processing resources devoted to each individual item from the study set. Furthermore, given such information, the observer is presumed to know the standard deviation of the distribution of remembered values to which the process would give rise across trials of the experiment. The observer is then presumed to use an ideal-observer Bayesian decision rule in combination with these known statistics for making his or her change-detection judgments. In agreement with Keshvari et al. (2012), we found that the knowledge-rich version of the model yielded better quantitative fits to our change-detection data than did the knowledge-limited version. Thus, our results provide converging evidence from a modified paradigm in support of Keshvari et al.'s previous conclusions. In addition, our study complemented Keshvari et al.'s earlier one, by testing a paradigm that yielded a strong qualitative contrast between the predictions from the models. Thus, rather than relying solely on quantitative fit results, our approach highlighted a focused *reason* why the knowledge-limited model fell short compared to the knowledge-rich one.

Beyond pointing to the advantages of knowledge-rich compared with knowledge-limited VR models, Keshvari et al. (2013) further argued in favor of the VR model compared to models based on item limits (i.e., mixed-state models). However, the particular mixed-state models that they tested were all constrained by the assumption that observers use ideal-observer Bayesian decision rules. In the present work, we developed a more descriptive version of the mixed-state model that was not constrained by this assumption. Our finding was that the mixed-state model yielded much better quantitative fits than did the knowledge-rich VR model with an ideal-observer Bayesian decision rule. On the one hand, it seems likely that more flexible versions of the knowledge-rich ideal-observer VR models could be formulated that would yield improved quantitative fits to our data.[5] On the other hand, the versions of the models that we tested were directly analogous to those actually proposed in the literature. At the least, our results reopen the debate concerning the relative merits of mixed-state

[5] For example, the ideal-observer model that we tested assumed that the observer knows the exact size of the small and big changes when evaluating the likelihood of change. Analogously, Keshvari et al.'s (2012, 2013) analyses assumed that observers knew the exact uniform distribution of sizes of change in their alternative design. Note that in our procedure, the instructions informed subjects about the small-change/big-change structure of the paradigm at the outset. Moreover, one third of the trials were set-size-2 trials, which would lead to minimal noise in the observers' memory representations for the study items. Thus, subjects had a huge sample size for developing accurate estimates of the magnitude of change of the small-change and big-change trials. Nevertheless, a more flexible ideal-observer model might allow for uncertainty or bias in the observers' expectations about the size of the changes. The question then arises as to how one should formalize and parameterize such uncertainty and bias. Without constraint from theory, it is unclear how such choices should be made, so we abstain from developing and testing such models here. We emphasize that for such more flexible versions of knowledge-rich ideal-observer VR models to have predictive utility, a precise theory would be needed for understanding the manner in which the objective properties of each individual experimental design are translated into the assumptions that are made by the observer.

versus VR models as applied in the domain of change detection and point up the need for further research regarding these models.

Questions Regarding Psychological Plausibility

Even assuming that more flexible, principled versions of the knowledge-rich ideal-observer VR model can be formulated that yield improved quantitative fits, in our view the most critical issue is whether the "knowledge-rich" assumption is a psychologically plausible one in the first place. Admittedly, judgments of psychological plausibility are more difficult to formalize and use as model-evaluation criteria than are rigorous measures of comparative quantitative fit. However, such judgments have guided the course of theorizing in many other fields. In the present case, the view is that the observer applies an ideal-observer decision rule to individual-item memory representations each drawn from unique probability distributions with precisely known mathematical structure and parameter settings. For the present paradigm, this view amounts to the assumption that the observer applies a unique ideal-observer criterion to each and every individual item from the memory set. Similar types of assumptions have been strongly resisted in other domains of cognitive and perceptual psychology. We provide two brief examples.

First, consider the long history of research in response-time modeling in psychology. A wide variety of evidence-accumulation models have been proposed, including random-walk models (Link, 1975), diffusion models (Ratcliff, 1978), leaky-competing accumulator models (Usher & McClelland, 2001), and linear ballistic accumulators (Brown & Heathcote, 2008). In such models, responses are made once accumulating evidence reaches a response threshold or criterion. Although the models differ in their details, an essentially ubiquitous assumption is that the response thresholds vary only in a *between-conditions* fashion: The idea that a separate response threshold is set for each individual stimulus that might be presented on a given trial of a fixed experimental condition has not been viewed as psychologically plausible (e.g., Donkin, Brown, & Heathcote, 2011). The assumption embedded in the knowledge-rich VR models is even more extreme, namely that a separate criterion is adopted for each individual item of a briefly presented multiitem visual display.

To take a second example, consider the domain of long-term recognition memory. One of the challenging phenomena observed in that domain is the *mirror effect* (Glanzer & Adams, 1985). In brief, the effect refers to the finding that for a variety of different experimental variables, classes of stimuli that are accurately recognized as old when old are also accurately rejected as new when new. For example, hit rates for old low-frequency words exceed hit rates for old high-frequency words; while correct-rejection rates for new low frequency words also exceed correct-rejection rates for new high-frequency words. Although the generality of the mirror-effect has been questioned (for a review and analysis, see Greene, 2007), much theorizing has nevertheless been devoted in an attempt to explain it. In signal-detection terms, a good description would be that for a certain class of stimuli, the old-item and new-item distributions are pushed away from one another, but wherever they land on the evidence axis, the observer is able to set the old-new criterion for each class at a midway point between the two distributions. However, memory theorists have typically rejected this form of word-class-specific criterion placement. The idea that the observer would be able to adjust her criterion for a given trial in response to each such experimental variable was viewed as highly implausible. As an alternative, theorists have preferred to place the locus of the effects within the representations of the items themselves and to use Bayesian decision rules that make comparisons to fixed decision criteria. For example, in the retrieving effectively from memory (REM) theory of Shiffrin and Steyvers (1997), because of the assumed nature of the item representations, matches (or mismatches) to features of low-frequency words yield greater evidence that an item is old (or new) than do matches (or mismatches) to features of high-frequency words.

By contrast, in current versions of the knowledge-rich VR models, information regarding precision does not seem to be embedded in the item representation itself. Instead, the observer is presumed to remember some specific scalar value randomly drawn from a circular normal distribution. The manner in which that scalar value is evaluated for a change or same judgment then varies depending on detailed knowledge of the processes that led to that representation. The implications of the assumption that observers have access to these forms of detailed statistical processing knowledge remain to be explored in other paradigms for evaluating the nature of continuous versus discrete-state memory representations (e.g., Dubé, Starns, Rotello, & Ratcliff, 2012; Kellen & Klauer, 2014; Mickes, Wais, & Wixted, 2009; Province & Rouder, 2012).

Questions Regarding Internal Consistency

Finally, questions arise regarding the internal consistency of the types of assumptions embedded in the ideal-observer VR models. First, by assuming that observers have perfect access to the precision with which items were encoded, the model makes the counterintuitive claim that the memory for such information is without capacity. That is, the observer's memory for the precision with which each individual item is stored is as good for eight items as for two items. We find this claim unusual, given that the model simultaneously says that the precision of memory for the features of items must decrease as more items must be remembered. Second, the claim also seems unusual given our finding that, for the knowledge-rich VR model to be competitive, one needed to assume that subjects' subjective change-probability estimates departed in dramatic fashion from the true objective change probabilities that were operating (see section on Model-Fitting Results for the Knowledge-Rich Models). Yet, in our experimental methods, the objective change probabilities were given huge emphasis in the instructions prior to each block. In addition, our observers were highly practiced and extremely familiar with the structure of the task. The question arises why knowledge of the statistics of hypothetical between-trials distributions of individual-item memory representations would be so precise (and used in ideal-observer fashion), when knowledge of other more transparent aspects of the task structure are apparently so imprecise (or inappropriately used).

Our concerns regarding the assumption of perfect access to individual-item memory precision could likely be allayed with convincing studies that provide independent evidence of such access. There are indeed published studies that have demonstrated various forms of metamemory in tasks of visual working memory (e.g., Fougnie et al., 2012; Rademaker, Tredway, & Tong, 2012), and process models have been proposed from which such judgments of metamemory may emerge (e.g., Swan & Wyble, 2014). For example, Fougnie, Suchow, and Alvarez (2012) reported an experiment in which subjects engaged in the continuous-reproduction task. On half the trials, the subjects reported the color of the item from the study display that they judged they had best remembered. On the other half of the trials, subjects reported the color of an item that was probed at random. Measured precision was better for the trials in which subjects were given the opportunity to choose which item they wanted to reproduce. Such a demonstration falls far short, however, of providing evidence that subjects have access to the absolute precision associated with each individual item from the visual display. For example, various studies involving visual STM indicate that in designs involving rapid visual sequential presentations, performance is best for the most recently presented item (e.g., Kahana & Sekuler, 2002; Nosofsky, Cox, Cao, & Shiffrin, 2014; Nosofsky & Donkin, 2016; Nosofsky, Little, Donkin, & Fific, 2011). In simultaneous-presentation designs, there may be forms of covert sequential attention to individual items on the display. If an observer simply chooses to report the most recent item to which he or she attended, then performance is likely to better for that item than for one selected at random by the experimenter. (Note that a bias for choosing to report the most recently attended item may have nothing to do with an assessment of the memory precision associated with that item.)

In addition, various demonstrations of metamemory in VWM are compatible with the assumptions of mixed-state models as well. For example, observers presumably know (at least some of the time) when there is a complete absence of memory for some probed study item (which is what forces them to guess). If asked to provide confidence ratings, the distribution of ratings associated with the absence-of-memory state would be expected to differ from the distribution associated with the presence-of-memory state (e.g., Rademaker et al., 2012; see also Province & Rouder, 2012). Correlations between confidence and measured precision in continuous reproduction may also reflect, in part, the operation of an intermediate, categorical memory state, in which observers retain coarse-grained or verbal-labeling information regarding the value of a study stimulus in the absence of true perceptual memory (e.g., Donkin, Nosofsky, Gold, & Shiffrin, 2015; Logie, 2011; Ricker, Cowan, & Morey, 2010; Rouder, Thiele et al., 2015). In a nutshell, an important question for future research is whether evidence for metamemory in visual working memory tasks truly confirms the assumption of perfect access to individual-item continuous precision that is part of the extant ideal-observer VR models.

Conclusions

The purely continuous VR models put forward by Keshvari et al. (2012, 2013) and van den Berg et al. (2012) provide an impressive account of behavior in visual working memory change-detection tasks. In this article we highlighted, however, that to account for behavior in such tasks, the VR model assumes that observers have detailed access to the precision with which each and every item in a study display has been stored. Without that assumption, the model is unable to simultaneously predict performance on same, small-change, and big-change trials. One of our aims in this research was simply to shine a light on the importance of this assumption by developing a paradigm that yielded a sharp qualitative contrast between the predictions from knowledge-limited mixed-state and continuous VR models. This approach brings into sharp focus the crucial nature of the assumption; thus, we hope our results will lead the field to closely examine the psychological validity of the detailed-access assumption in future research. A second contribution involved the demonstration that, even allowing the detailed-access assumption, a straightforward descriptive version of a mixed-state model provided quantitative accounts of visual change-detection performance that were as good as or better than those of extant versions of the ideal-observer VR model. It is an open question whether this finding will generalize to other experimental paradigms. However, our findings reopen the debate concerning the relative merits of mixed-state and ideal-observer VR models of visual change detection.

References

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. http://dx.doi.org/10.1111/j.1467-9280.2007.01949.x

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. http://dx.doi.org/10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321,* 851–854. http://dx.doi.org/10.1126/science.1158023

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57,* 153–178. http://dx.doi.org/10.1016/j.cogpsych.2007.12.002

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114. http://dx.doi.org/10.1017/S0140525X01003922

Cowan, N., & Rouder, J. N. (2009). Comment on "Dynamic shifts of limited working memory resources in human vision." *Science, 323,* 877. http://dx.doi.org/10.1126/science.1166478

Donkin, C., Brown, S. D., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator model. *Journal of Mathematical Psychology, 55,* 140–151. http://dx.doi.org/10.1016/j.jmp.2010.10.001

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. http://dx.doi.org/10.1037/a0034247

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review, 22,* 170–178. http://dx.doi.org/10.3758/s13423-014-0675-5

Dubé, C., & Sekuler, R. (2015). Obligatory and adaptive averaging in visual short-term memory. *Journal of Vision, 15,* 13. http://dx.doi.org/10.1167/15.4.13

Dubé, 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 and Language, 67,* 389–406. http://dx.doi.org/10.1016/j.jml.2012.06.002

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications, 3,* 1229. http://dx.doi.org/10.1038/ncomms2237

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20. http://dx.doi.org/10.3758/BF03198438

Greene, R. L. (2007). Foxes, hedgehogs, and mirror effects: The role of general principles in memory research. In J. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 53–66). Hove, UK: Psychology Press.

Hooke, R., & Jeeves, T. A. (1961). "Direct search" solution of numerical and statistical problems. *Journal of the ACM, 8,* 212–229. http://dx.doi.org/10.1145/321062.321069

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research, 42,* 2177–2192. http://dx.doi.org/10.1016/S0042-6989(02)00118-9

Kellen, D., & Klauer, K. C. (2014). Discrete-state and continuous models of recognition memory: Testing core properties under minimal assumptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1795–1804. http://dx.doi.org/10.1037/xlm0000016

Keshvari, S., van den Berg, R., & Ma, W. J. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE, 7,* e40216. http://dx.doi.org/10.1371/journal.pone.0040216

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology, 9,* e1002927. http://dx.doi.org/10.1371/journal.pcbi.1002927

Link, S. W. (1975). The relative judgment theory of two choice reaction time. *Journal of Mathematical Psychology, 12,* 114–135. http://dx.doi.org/10.1016/0022-2496(75)90053-X

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science, 20,* 240–245. http://dx.doi.org/10.1177/0963721411415340

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. http://dx.doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17,* 391–400. http://dx.doi.org/10.1016/j.tics.2013.06.006

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17,* 347–356. http://dx.doi.org/10.1038/nn.3655

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science, 20,* 509–515. http://dx.doi.org/10.1111/j.1467-9280.2009.02324.x

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,* 1524–1539. http://dx.doi.org/10.1037/xlm0000015

Nosofsky, R. M., & Donkin, C. (2016). Response-time evidence for mixed memory states in a sequential-presentation change-detection task. *Cognitive Psychology, 84,* 31–62. http://dx.doi.org/10.1016/j.cogpsych.2015.11.001

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. http://dx.doi.org/10.1037/a0022494

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44,* 369–378. http://dx.doi.org/10.3758/BF03210419

Province, J. M., & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 14357–14362. http://dx.doi.org/10.1073/pnas.1103880109

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision, 12,* 21. http://dx.doi.org/10.1167/12.13.21

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108. http://dx.doi.org/10.1037/0033-295X.85.2.59

Ricker, T. J., Cowan, N., & Morey, C. C. (2010). Visual working memory is disrupted by covert verbal retrieval. *Psychonomic Bulletin & Review, 17,* 516–521. http://dx.doi.org/10.3758/PBR.17.4.516

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 of the United States of America, 105,* 5975–5979. http://dx.doi.org/10.1073/pnas.0711295105

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*. Advance online publication.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review, 119,* 807–830. http://dx.doi.org/10.1037/a0029856

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* 1–29. http://dx.doi.org/10.1037/h0093759

Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics, 76,* 2136–2157. http://dx.doi.org/10.3758/s13414-014-0633-3

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592. http://dx.doi.org/10.1037/0033-295X.108.3.550

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121,* 124–149. http://dx.doi.org/10.1037/a0035234

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 of the United States of America, 109,* 8780–8785. http://dx.doi.org/10.1073/pnas.1117465109

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114. http://dx.doi.org/10.1037/0096-1523.27.1.92

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4,* 1120–1135. http://dx.doi.org/10.1167/4.12.11

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453,* 233–235. http://dx.doi.org/10.1038/nature06860

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20,* 423–428. http://dx.doi.org/10.1111/j.1467-9280.2009.02322.x

(*Appendix follows*)

Appendix

Individual-Subject Data

Table A1
Proportion of Change Judgments in Each Condition for Subject 1

Change type	Memory set size SS = 2	SS = 5	SS = 8
	Condition *cp* = .3		
Same	.03	.13	.27
Small change	.56	.50	.43
Big change	1.00	1.00	.97
	Condition *cp* = .5		
Same	.07	.16	.34
Small change	.60	.52	.62
Big change	.98	.99	.95
	Condition *cp* = .7		
Same	.08	.19	.40
Small change	.71	.54	.68
Big change	.99	1.00	.97

Note. SS = set size; *cp* = objective change probability.

Table A2
Proportion of Change Judgments in Each Condition for Subject 2

Change type	Memory set size SS = 2	SS = 5	SS = 8
	Condition *cp* = .3		
Same	.06	.09	.19
Small change	.73	.55	.35
Big change	.99	.96	.91
	Condition *cp* = .5		
Same	.11	.15	.27
Small change	.71	.65	.63
Big change	.99	.96	.92
	Condition *cp* = .7		
Same	.11	.23	.28
Small change	.80	.60	.60
Big change	1.00	.98	.95

Note. SS = set size; *cp* = objective change probability.

(*Appendix continues*)

Table A3
Proportion of Change Judgments in Each Condition for Subject 3

Change type	Memory set size SS = 2	SS = 5	SS = 8
	Condition *cp* = .3		
Same	.03	.13	.20
Small change	.26	.26	.30
Big change	.92	.71	.49
	Condition *cp* = .5		
Same	.08	.35	.48
Small change	.30	.44	.52
Big change	.97	.86	.83
	Condition *cp* = .7		
Same	.22	.51	.65
Small change	.47	.69	.74
Big change	.97	.96	.95

Note. SS = set size; *cp* = objective change probability.

Table A4
Proportion of Change Judgments in Each Condition for Subject 4

Change type	Memory set size SS = 2	SS = 5	SS = 8
	Condition *cp* = .3		
Same	.17	.21	.23
Small change	.50	.41	.34
Big change	.98	.81	.79
	Condition *cp* = .5		
Same	.18	.28	.37
Small change	.53	.47	.43
Big change	.99	.88	.84
	Condition *cp* = .7		
Same	.27	.35	.48
Small change	.61	.54	.59
Big change	.99	.91	.88

Note. SS = set size; *cp* = objective change probability.

(*Appendix continues*)

Table A5
Proportion of Change Judgments in Each Condition for Subject 5

Change type	SS = 2	SS = 5	SS = 8
	Memory set size		
	Condition $cp = .3$		
Same	.08	.17	.17
Small change	.48	.25	.26
Big change	.98	.72	.43
	Condition $cp = .5$		
Same	.07	.39	.54
Small change	.59	.47	.66
Big change	.98	.91	.85
	Condition $cp = .7$		
Same	.19	.42	.72
Small change	.57	.68	.82
Big change	1.00	.98	.95

Note. SS = set size; *cp* = objective change probability.

Table A6
Proportion of Change Judgments in Each Condition for Subject 6

Change type	SS = 2	SS = 5	SS = 8
	Memory set size		
	Condition $cp = .3$		
Same	.01	.06	.17
Small change	.74	.31	.36
Big change	.99	.85	.73
	Condition $cp = .5$		
Same	.03	.07	.16
Small change	.68	.29	.32
Big change	.99	.94	.85
	Condition $cp = .7$		
Same	.08	.17	.34
Small change	.75	.48	.52
Big change	1.00	.97	.95

Note. SS = set size; *cp* = objective change probability.

Received December 6, 2015
Revision received January 23, 2016
Accepted January 30, 2016 ■