Informed Guessing in Change Detection

Stephen Rhodes[1,2], Nelson Cowan[1], Kyle O. Hardman[1], and Robert H. Logie[2]

[1]Department of Psychological Sciences, University of Missouri

[2]Department of Psychology, University of Edinburgh

Accepted for publication in *Journal of Experimental Psychology: Learning, Memory, & Cognition*

Author Note

This work was supported by grant R01 HD-21338 (NICHD) to NC. SR, NC, and RHL are currently supported by the United Kingdom ESRC, grant ES/N010728/1. We would like to thank Levi Doyle-Barker for collecting the data for Experiment 4 and members of the University of Missouri working memory lab for helpful discussion.

The materials, data, and analysis code for the experiments reported here can be found at: `https://github.com/stephenrho/Guessing`

Email: rhodessp@missouri.edu.

Abstract

Provided stimuli are highly distinct, the detection of changes between two briefly separated arrays appears to be achieved by an all-or-none process where either the relevant information is in working memory or observers guess. This observation suggests that it is possible to estimate the average number of items an observer was able to retain across a series of trials, a potentially highly informative cognitive characteristic. For each version of the change detection paradigm, for this estimate to be accurate, it is important to specify how observers use the information available to them. For some instantiations of this task it is possible that observers use knowledge of the contents of working memory even when they are in a guessing state, rather than selecting between the response alternatives at random. Here we test the suggestion that observers may be able to use their knowledge of the number of items in memory to guide guessing in two versions of the change detection task. The four experiments reported here suggest that participants are, in fact, able to use the parameters of the task to update their base expectation of a change occurring to arrive at more informed guessing. (194 words)

Keywords: Change Detection, Working Memory, Guessing

Informed Guessing in Change Detection

The change detection paradigm is a simple task in which participants study an array of items (e.g. colored squares) and, following a brief delay, indicate whether a change has occurred or not in a probe array. It has become a very popular task in cognitive psychology due to this simplicity and, in no small part, the availability of simple processing models that allow researchers to estimate the number of items an individual could hold in mind over a series of trials (Cowan, 2001; Pashler, 1988; Rouder, Morey, Morey, & Cowan, 2011). These easily obtained estimates are found to correlate highly with important outcomes, such as scholastic aptitude (Cowan et al., 2005) or fluid reasoning ability (Johnson et al., 2013), suggesting that this is an important cognitive characteristic.

The assumption underlying these models is that the observer has k of the N items in memory with sufficient precision to detect a change. On a certain proportion of trials the relevant information, needed to make the discrimination, is in memory and the observer chooses the correct response option (a 'high-threshold' assumption). For example, for the task depicted in Figure 1A, the probability of this occurring is given by, $d = \min(k/N, 1)$. On the remaining proportion of trials $(1 - d)$, however, the participant must guess between 'change' and 'same'. This all-or-none assumption appears to be justified for highly distinct simultaneously presented stimuli (e.g. categorically distinct colors) as indicated by linear receiver operating characteristics (Rouder et al., 2008; Donkin, Tran, & Nosofsky, 2014) and the analysis of response time distributions (Donkin, Nosofsky, Gold, & Shiffrin, 2013). Thus, while the way in which visual working memory capacity is conceived is still very much evolving (Donkin, Kary, Tahir, & Taylor, 2016; Luck & Vogel, 2013; Ma, Husain, & Bays, 2014), this observation appears to justify talking about *number of item* metrics for the canonical change detection task.

However, the way in which this metric is derived depends on the exact change detection task given and the logical constraints it places on how information can be used to perform the discrimination (Rouder et al., 2011). Take the standard single

probe task, for example (see Figure 1A). In this case a single test item occupies a location previously filled in the study array. According to the high-threshold assumption, where the relevant information is either in mind or not, on $1 - d$ proportion of trials the observer does not have the requisite information in memory and must guess. In this case it is assumed that participants go with their base expectation of a change occurring, which we will call u, for *uninformed* guessing rate (cf. Rouder et al., 2011).

For the whole display paradigm, the possible constraints on information use are different (see Figure 1B). Here the observer is assumed to match the k items in memory to the corresponding items in the probe array and if a mismatch is found (on d proportion of trials) the response *change* is given (Pashler, 1988). If a mismatch is not found, however, and the array size exceeds the number of items they could retain ($k < N$) the observer is in a state of uncertainty; it may be that an item outside of memory changed or it could be that the array is an exact repetition of the study set. In this case the observer may respond *change* in line with their uninformed guessing rate. According to this basic model, first outlined by Pashler (1988), the probability of making a hit (h) for the whole display task is, $h = d + (1 - d)u$, whereas the probability of making a false-alarm (f) is, $f = u$, provided that the observer was unable to encode and retain all memory array items ($k < N$).

Alternatively, it is possible for the observer to use knowledge of the number of items they have in memory to guide their response. Rouder et al. (2011) outline one plausible way in which guessing might be *informed* in the whole display change detection task. They note that the observer's base expectation of a change can be updated by their knowledge of the probability that, had an item changed, they would have missed it given the number of items they had in memory (which, in terms of the parameters above, is $1 - d$). In an application of Bayes' theorem the prediction for this informed guessing is,

$$g^{(\mathrm{WD})} = \frac{(1 - d)u}{(1 - d)u + (1 - u)}$$

where WD stands for whole display. This equation reflects the assumption that informed guessing updates the prior expectation of a change (given by the uninformed

guessing rate, u) via knowledge of k and, consequently, the probability of missing a change (given by $1 - d$). The equation can be read more simply as the probability that all items in memory match items in the test array *given that a change occurred* divided by the probability that all items match regardless of whether this is a same or change trial.[1] Note that, unlike the uninformed guessing rate, this informed guessing depends on d, which in turn depends on set size (N) and the number of items in working memory (k). A consequence of this is that, for a given uninformed guessing rate and set size, observers with more items in memory should be *less* likely to guess *change*. Or alternatively, for an observer with fixed k, as the number of to-be-remembered items increases, the probability that they guess *change* should also increase. This is because increasing N relative to k increases the chances of a true change being missed.

It is also possible for guessing to be informed in single probe versions of the change detection task. When the single probe is presented in a neutral location that was not occupied by items in the memory array (usually the center of the screen, see Figure 1C) it is not possible for observers to evaluate the status of only a single item, as can be done in the standard single probe task (Gilchrist & Cowan, 2014). Rather if the observer does not identify a match between the probe and an item in memory, and capacity was exceeded by the number of to-be-remembered items, they must guess whether or not a change has actually occurred. Note that in this case the uncertain state begins with failure to match, whereas for a whole display uncertainty arises when all items in memory match those in the probe (a failure to detect change). Cowan, Blume, and Saults (2013) outline the basic uninformed model for this task, where the probability of making a false-alarm is, $f = (1 - d)u$, and the probability of making a hit for arrays exceeding capacity is, $h = u$.[2]

Note that in this case guessing can be informed by knowledge of k too. Here if a

[1] Note that the prior expectation of a *same* trial is given by $1 - u$ and the probability that all items in memory match those in the test array on a same trial is necessarily 1 (as no items change), which gives the right side of the denominator

[2] We have slightly changed the way in which this model is presented here relative the the original presentation in Cowan et al. (2013) but the model remains the same.

match is not detected the informed guesser can deduce that this could have happened on a *same* trial if the relevant item was not retained in memory. This occurs with the probability, $1 - d$. In this case the same logic used by Rouder et al. (2011) leads to the following prediction for informed guessing in this single probe task;

$$g^{(\mathrm{SP})} = \frac{u}{u + (1 - d)u},$$

where SP stands for single probe. This may be interpreted more simply as the probability that a match between the probe and an item in memory was not detected given that a change occurred (which for this task is given by the uninformed guessing rate) divided by the probability of no match being detected for both *same* and *change* trials.[3]

The prediction arising from this is that for a given uninformed guessing rate and set size, observers with more items in working memory are predicted to be *more* likely to guess *change*. Alternatively, for a fixed k, increasing the number of to-be-remembered items should decrease the probability of guessing *change*. This is because increasing N relative to k increases the probability that the probe was not retained in memory and consequently that a true match would be missed. Note that this is the opposite to the prediction informed guessing makes for the whole display task, due to the different constraints imposed on information use (i.e., the possibility to detect a match as opposed to the possibility to detect a change).

The clear assumption in these informed guessing accounts is that participants know the number of items they have in working memory and use this information to guide responding, even when they are in an uncertain state. The extant literature provides some reason to believe that observers have insight into the contents of working memory and that they can use this to guide responses. Cowan et al. (2016) used a modified change detection task, in which participants had to indicate how many changes had occurred between study and test, rather than whether or not a change had

[3]The numerator and left side of the denominator are u in this case because the probability that no match occurs on a *change* trial is necessarily 1. Multiplied by the prior (or base) expectation of a change gives u.

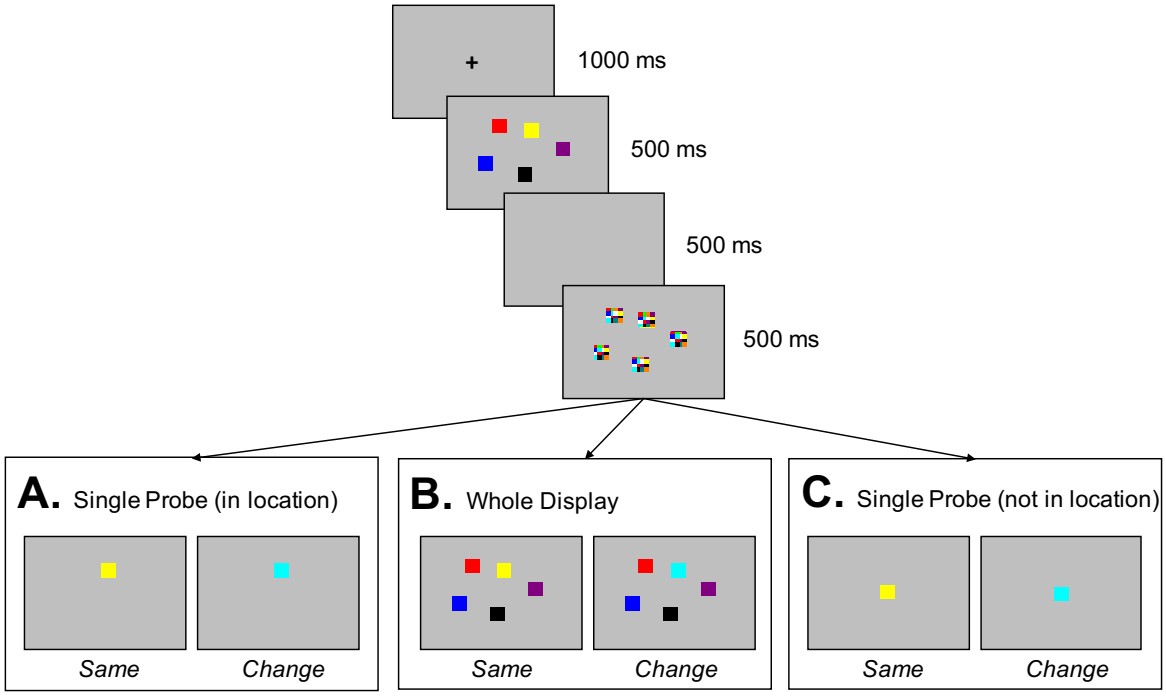

Figure 1. The general change detection procedure with a study array followed by a pattern mask then a test array. Examples of arrays requiring *same* and *change* responses are given for the standard single probe task in location (A), the whole display task (B), and the single probe task where the probe appears in a previously unoccupied location (C). Note participants see only one probe array per trial.

occurred. Occasionally, in the retention interval of their trial procedure, Cowan et al. (2016) would probe their participants to report the number of items they felt they had in memory. This introspective judgement corresponded quite well to model based estimates of capacity, with a slight tendency for participants to overestimate the number of items in memory. This meta-cognitive insight is also present in recall tasks requiring the precise reconstruction of a studied item (e.g. selecting a color from a continuous set). Rademaker, Tredway, and Tong (2012) found that ratings of the quality of memory for a given probed item tracked the precision with which that item was recalled. Further, using a mixture model to separate guesses from memory responses, they found that these meta-cognitive judgements also related to the probability of the probed item being in memory (although see Van den Berg, Yoo, & Ma, 2017, for an alternative analysis).

In recent years the simple models outlined above have been extended to better account for extant data. For instance, it is common practice to include an attention lapse parameter to account for occasional errors at small set sizes that are presumably smaller than capacity (e.g. Donkin et al., 2014; Rouder et al., 2008; Hardman & Cowan, 2016). With this additional parameter, closed form solutions for k are not available and model fitting techniques must be used instead. This requires precise specification of the model, as guessing is no longer subtracted out like it is for closed form estimators (Pashler, 1988; Cowan et al., 2013). Failure to accurately account for how observers use information will lead to biased estimates of parameters and possibly dubious experimental effects (see Rouder et al., 2011). Thus, the difference between informed and uninformed guessing is not a trivial one. To accurately measure the quantity of interest to many researchers, namely the number of items a participant could accurately maintain in mind across a series of trials, a more precise specification of guessing behavior is required. Four experiments aimed to help move towards this more precise specification with two versions of the change detection task.

The present experiments

The informed guessing described above depends on two factors; the participants' prior expectation of a change, u, and the probability that the relevant information is in memory when the probe appears, d. The precise way in which this information is used depends on the parameters of the task (i.e. whether a single probe or whole display is used). Therefore, one way to disambiguate this from uninformed guessing is to systematically manipulate each of these factors. Figure 2 depicts hypothetical ROC curves for the whole display and single probe tasks according to uninformed and informed guessing. Each line depicts a different base expectation of a change, or u value, whereas each of the three points along the lines represents a probability, d, of the relevant information being in memory when probed. The former may be manipulated by informing participants that change trials will be more or less frequent and the latter may be manipulated by increasing the number of to-be-remembered items.

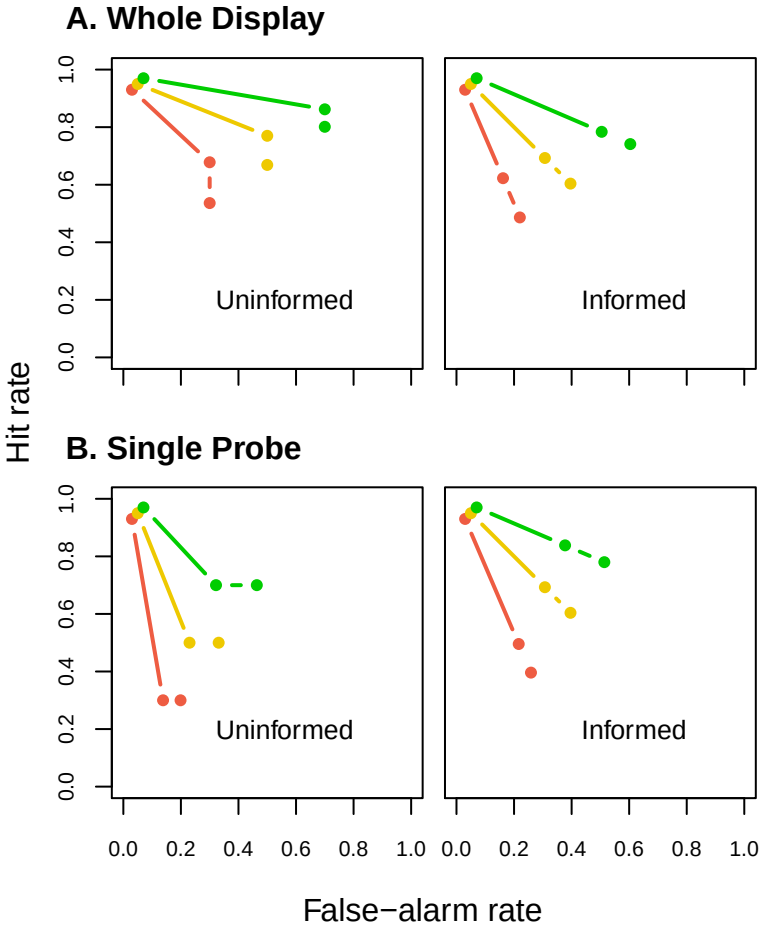

Figure 2. Hypothetical ROC curves for the uninformed and informed guessing models for the whole display (A) and single probe (B) change detection tasks. These ROCs depict the situation where uninformed guessing rate is manipulated along three levels (the three lines in different colors) and the probability of having relevant information in memory is manipulated by increasing set size (the points along each line). Note that the underlying model parameters are the same and depict a hypothetical participant with $k = 3$, a (attention lapse parameter) $= 0.9$, and one uninformed guessing rate for each bias condition (see Method for more detail), $u = [0.3, 0.5, 0.7]$.

In Experiment 1 we use these two manipulations with the whole display change detection task. To preview the results of this experiment, the resulting ROC curves appear to be better reconciled with an informed guessing account. In Experiment 2 we seek converging evidence with the whole display task by this time varying the number of possible changes on a given trial, rather than the probability of a change occurring.

Finally, in Experiments 3 and 4 we use the bias and set size manipulations from Experiment 1, this time with the single probe version of the change detection task (see Figure 1C). In all four experiments the data support an informed guessing account over a uninformed account.

Experiment 1

As noted above, according to the informed guessing model the propensity of an observer to guess *change* depends on the number of to-be-remembered items (set size) and their base expectation of a change. Thus in Experiment 1 we followed the general procedure of Rouder et al. (2008) to trace out ROC curves for the whole display task. Participants studied arrays of 2, 5, or 8 colored squares and were told, prior to each block of trials, the probability that a change would occur between study and test. It was emphasized to participants that a change would only occur to a single color and that this would introduce a color that was not in the studied set. The base rates used were 0.3, 0.5, and 0.7.

Uninformed guessing predicts that, for array sizes exceeding capacity, the false-alarm rate should only depend on the uninformed guessing rate. As an example, for an individual with $k = 3$ their propensity to say a change occurred when it did not should remain constant between set sizes 5 and 8 (see Figure 2A). Informed guessing, on the other hand, predicts that the false-alarm rate depends both on set size (N) and base expectation of a change (u). This is because informed guessing predicts that observers are less likely to guess *change* at smaller set sizes (as there is less chance that a change was missed).[4]

Method

Participants. Thirty-one students (25 female) from the University of Edinburgh, aged between 18 and 33 (mean = 22.3), took part in this experiment in

[4]The two models also predict different hit rates but in the description above we focus on false-alarm rate for illustrative purposes as these differences are most pronounced.

return for £7 for the 45 minute session. One participant was excluded for failing to achieve greater than 60% accuracy averaged across all experimental blocks.

Stimuli, Design, and Procedure. Stimuli were colored squares presented on a grey background. Colors on each trial were selected without replacement from a set of 10 taken from Rouder et al. (2008); black, white, red, blue, green, yellow, orange, cyan, purple, and teal. Each square measured $0.75° \times 0.75°$ of visual angle at an approximate viewing distance of 50 cm. Stimuli were randomly positioned on each trial within a $9.8° \times 7.3°$ area with the constraint that each item could not be within 2° of the center of the screen or the center of another item. Pattern masks presented following the memoranda also subtended $0.75° \times 0.75°$ and were made up of a 4×4 grid with each cell filled with a color from the study set.

The general trial procedure is depicted in Figure 1B. Participants saw a fixation cross for 1000 ms which was then replaced with the study array for 500 ms. Following this there was a blank interval of 500 ms before a 500 ms pattern mask to ensure any sensory trace was eliminated. Immediately following the mask the probe array was presented (Figure 1B) and remained visible until participants made their response by pressing either the 'z' or 'm' key for 'same' or 'change' (the mapping of keys was varied between participants).

The experimental session was split into 9 blocks. Prior to each block the participant was informed of the probability that a given trial in that block would involve a change to a single square between study and test. Pie charts were used to depict the relative probabilities (in percentages) of same and change trials prior to each block and the probability of a change was reiterated on screen prior to each trial. Participants progressed through the 3 base rate conditions (30%, 50%, and 70%) in a random order 3 times. Each block contained 60 trials (for a total of 540) with the number of change and same trials determined by the base rate manipulation (as the base rates were genuinely informative). The different set sizes (2, 5, and 8) were randomly distributed throughout each block with equal probability. All experiments were programmed using PsychoPy (Peirce, 2007).

Modeling approach. Hierarchical Bayesian versions of the informed and uninformed guessing models (similar to those outlined by Morey, 2011) were fit to the data. Each of these models has five parameters estimated for each participant; the number of items in working memory, k, the probability that attention is paid on a given trial, a, to account for the presence of errors at low set sizes (Rouder et al., 2008), and an uninformed guessing rate, u, for each of the base rate conditions. Individual participant parameters were assumed to be drawn from population-level distributions with weakly-informative priors placed on the mean and standard deviation of these distributions (see Appendix for more information). To quantify model fit, whilst penalizing for the effective number of parameters, we use the widely applicable information criterion (WAIC; Watanabe, 2010), calculated from the posterior chains as described by Vehtari, Gelman, and Gabry (2017). WAIC is a fully Bayesian measure (that is, it uses the full posterior distribution) of out of sample predictive accuracy (see Gelman et al., 2014, for further detail). We also report the more frequently used deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002; as calculated by Plummer, 2008) alongside parameter estimates in Table 1.

In addition to the guessing models, we also estimated hit and false-alarm rates from the raw data with a hierarchical model (see Donkin et al., 2016, for a similar approach). Rates for each individual in each condition were assumed to be drawn from normal distributions on the log odds scale with separately estimated means and a common estimated standard deviation. This model, used to estimate the rates with no theoretical assumptions, provides a basis for comparison with the theoretical models, producing shrinkage of estimated rates by using group means to constrain individual estimates. It also gives appropriate interval estimates (confidence intervals) that are bounded between 0 and 1. More detail on the modelling approach is given in the Appendix.

Results and Discussion

Estimated hit and false-alarm rates are plotted in Figure 3 with their 95% highest density intervals (HDI; see Kruschke, 2015, for detailed discussion) along with posterior predictive samples from the population mean parameters of the uninformed (left) and informed (right) models. In these figures points progressing along the negative diagonal reflect the different set sizes (2, 5, 8) and points following the positive diagonal reflect the different base-rates (0.3, 0.5, 0.7). Visually the informed model appears to provide a better account of the data relative to the uninformed model, which predicts that increasing the number of to-be-remembered items from 5 to 8 has no effect on false-alarm rate. These predictions of the uninformed model were not borne out, as the false-alarm rate does indeed increase with set size. In the 0.3 base rate condition the estimated false-alarm rate was 0.058 [0.011, 0.109] (median and 95% HDI) larger at set size 8 than 5, with similar results in the 0.5 (0.067 [0.004, 0.132]) and 0.7 (0.094 [0.017, 0.174]) conditions (0.073 [0.036, 0.111] collapsed across conditions). This increase in false-alarm rate is as predicted by the informed guessing model (parameter estimates for each model are given in Table 1). This impression was backed up by the WAIC which was smaller for the informed model (ΔWAIC = 206.18).

Some deviation can be seen between the aggregate data and informed model fit (Figure 3 right panel). For example, the model appears to predict shallower isobias functions at set sizes 5 and 8 than are present in the estimated rates. Nevertheless a comparison to the saturated model, used to estimate hit and false-alarm rates, via WAIC favored the informed guessing model, suggesting that the misfit is not too extreme (ΔWAIC = 113.56).

In summary, the data from Experiment 1 are better reconciled with the assumption that observers use knowledge of the number of items in memory to inform guessing behavior. We sought to obtain converging evidence for informed guessing in whole display change detection in a second experiment.

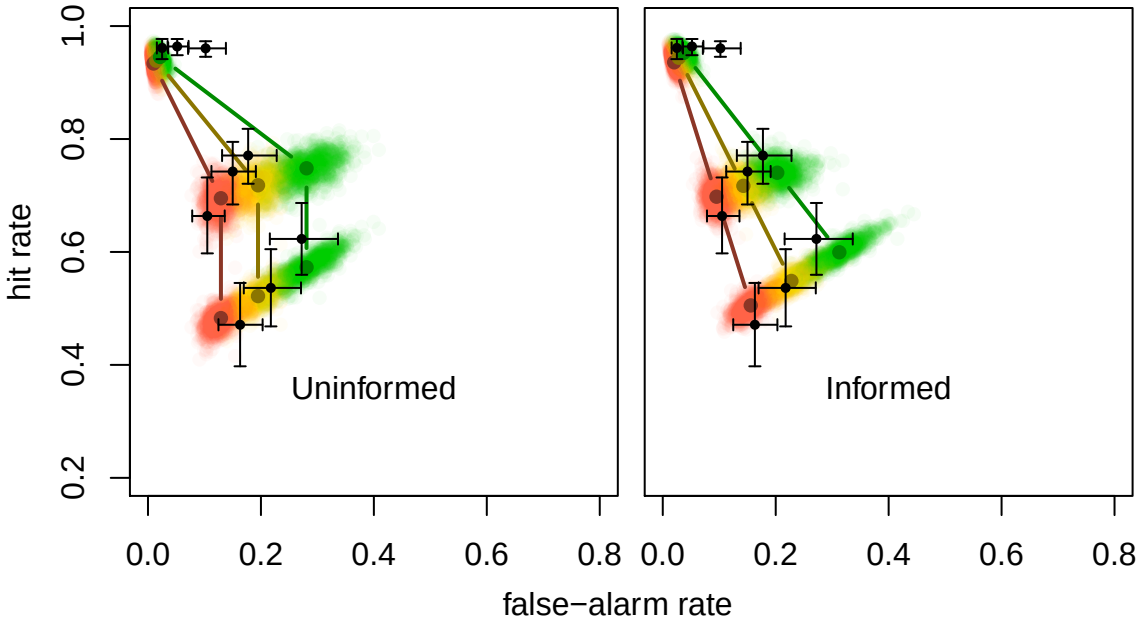

Figure 3. Estimated false-alarm and hit rates (posterior medians and 95% highest density intervals) from Experiment 1 shown in black over posterior predictive samples from the uninformed (left panel) and informed (right panel) guessing models (red = samples from the 0.3 base-rate condition, yellow = 0.5, green = 0.7). Points progressing along the negative diagonal denote the set size 2, 5, and 8 conditions, respectively. Posterior median predictions are also given joined by lines.

Experiment 2

In Experiment 1, for the most part, participants' whole display change detection performance was in line with the expectations of informed guessing. In Experiment 2 we sought to obtain converging evidence from a different experimental manipulation. In this case, rather than vary participants' base expectation of a change, we varied the number of changes (C) that would occur on a given change trial. Before starting a block participants were informed that when a change occurred (which would happen on 50% of trials) it would occur to 1, 2, 3, or 4 of the items. Wilken and Ma (2004) report a similar manipulation, however, in their case the different numbers of changes were mixed with non-changed probe arrays within the same trial blocks, thus there was no separate false-alarm rate for the different possible number of changes (see also, Gibson,

Wasserman, & Luck, 2011). Separating these trials out results in clear differential predictions for uninformed and informed guessers. Uninformed guessing predicts that, as long as $k < N - C$, there will be no dependence of false-alarm rate on the number of possible changes. Therefore, for example, at set size 5 we may predict a discontinuity where, for an observer with $k = 3$, false-alarm rate is close to zero for 3 and 4 possible changes but raises to u when only 1 or 2 changes can occur. This is because the observer cannot be sure that an item outside of memory has not changed, and therefore must guess. Informed guessing, on the other hand, predicts a more graceful dependence of false-alarm rate on C, as knowledge of the number of items is used to update the base expectation of a change. Further, given that only a single base-rate (0.5) was used in this experiment, the number of parameters needed to fit the informed and uninformed models is reduced from 5 to 3.

Method

Participants. A new sample of 28 students from the University of Edinburgh, aged between 18 and 41 (mean = 24.8), were recruited for Experiment 2. Each participant received £ 7 for the 45 minute session. All participants achieved greater than 60% accuracy averaged across blocks.

Stimuli, Design, and Procedure. As Experiment 2 included trials on which a large number of changes could occur, the following colors were added to the set used in Experiment 1; fuchsia, lime, and maroon, so we could sample without replacement. These colors were also included in the pattern mask for this experiment. Other aspects of the stimuli were identical to those used in Experiment 1.

The general trial procedure was identical to Experiment 1 (see Figure 1B) with the exception that change trials could consist of 1–4 squares changing to brand new colors, not present in the memory array. The experiment session was split into 8 blocks of 68 trials (for a total of 544). The two set sizes (5 and 8) were randomly intermixed within each block. Prior to each block participants were informed that a change trial would involve 1, 2, 3, or 4 squares changing to new colors, whereas the probability of a

change trial occurring was 50%. The number of possible changes remained constant within a block and was reiterated on screen prior to each trial.

Results and Discussion

Once again hierarchical versions of uninformed and informed models were estimated. In this case each model had three parameters estimated for each participant drawn from population distributions with weakly-informative priors. These free parameters were k, a, and a single uninformed guessing rate, u, as the probability of change was fixed at 50%. More details are given in the Appendix.

Posterior samples from the mean population-level parameters of the two models are presented in Figure 4 with estimated hit and false-alarm rates superimposed. Once again the informed model provides a visibly better fit to the data than the uninformed model. Uninformed guessing predicts a discontinuous relationship between the number of changes and false-alarm rate at set size 5. When the number of possible changes is small, and thus the observer cannot be certain that an item outside of working memory did not change, false-alarm rate should match the guessing rate, whereas the false-alarm rate should drop dramatically when the number of changes becomes large, as at least one item in memory will have changed between study and test (as the mean capacity estimate was approximately 3; see Table 1). This pattern is not present in the estimated false-alarm rates, although this is aggregated across individual participants with different capacities which may lead to a smoother change of false-alarm rate with C. The clearer discrepancy between the data and the uninformed model is at set size 8. Given the estimated number of items in memory the uninformed model predicts no change of false-alarm rate with number of possible changes. This is clearly not the case, estimated false-alarm rates do change significantly between 1 and 4 possible changes (-0.143 [-0.207, -0.08]). For the informed model the mean prediction lines match up nicely with the estimated hit and false-alarm rates, generally falling withing the 95% HDIs (see bottom of Figure 4).

Once again this visual inspection was supported by the WAIC comparison

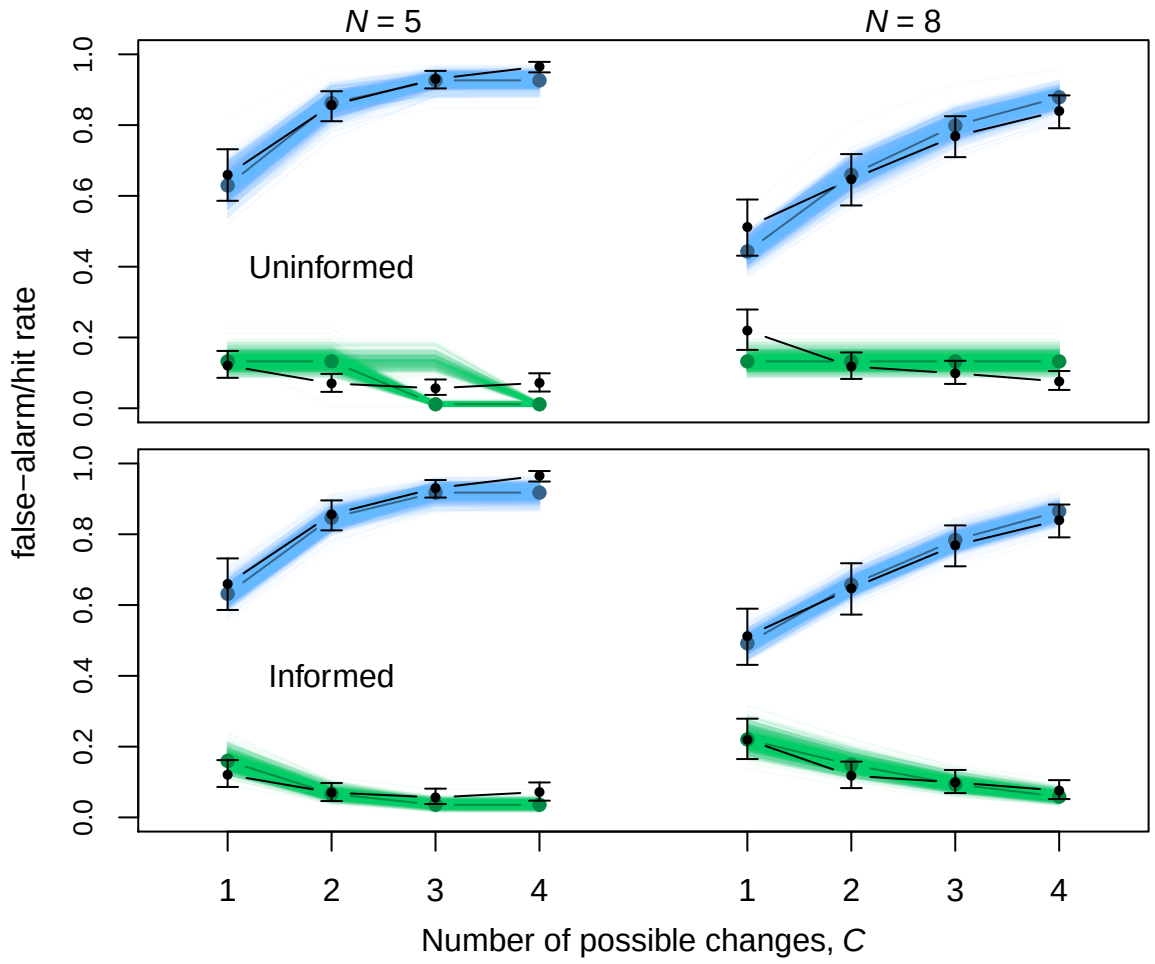

Figure 4. Estimated false-alarm (lower points) and hit (upper points) rates (posterior medians and 95% highest density intervals) from Experiment 2 shown in black over posterior predictive samples from the uninformed (top panel) and informed (bottom panel) guessing models (green = false-alarms, blue = hits). Posterior median predictions are also given joined by lines.

(ΔWAIC = 264.15). Whilst the informed model outperformed the uninformed model, there appears to be room for improvement as the WAIC for the saturated model, which more freely estimated hits and false-alarms, was slightly lower (ΔWAIC = -32.54). We discuss potential extensions to the informed guessing model prior to the General Discussion.

Experiment 3

Experiments 1 and 2 provide converging evidence that observers make more use of information afforded to them in whole display change detection than uninformed guessing would predict. Experiments 3 and 4 asked if this is also the case for single probe change detection. The whole display change detection task includes much more contextual information at test which appears to benefit performance relative to a single probe (Jiang, Olson, & Chun, 2000). Indeed, informed guessing assumes that participants have some notion of the number of items that were presented or the proportion of the array that they encoded. A whole display makes the number of presented items explicit, whereas a single probe does not provide this information at test. Further, with a whole display, it is possible that participants make a gist based comparison of the study and probe arrays on the basis of their higher order properties (e.g. average hue or luminence Brady & Alvarez, 2011). If this is the case, the evidence for informed guessing, particularly in Experiment 2, may reflect participant's adjusting their responding in accordance with the expected magnitude of change between study and test (i.e. in the 4 possible change condition, a minimal change in ensemble properties between study and test is unlikely to have happened given a change). This version of the single probe task has been increasingly used recently, particularly in studies assessing the binding of features in working memory, where it is desirable to detach the probe features from their initially studied locations (e.g. Wheeler & Treisman, 2002). Finding evidence of similar information use with a single probe would strengthen the argument that guessing is informed in change detection.

Method

Participants. Thirty students (21 female) from the University of Edinburgh, who had not taken part in Experiments 1 or 2, were recruited for Experiment 3. They were aged between 19 and 31 (mean = 22.8) and received £ 7 for the 45 minute session. All participants obtained greater than 60% accuracy averaged across blocks.

Stimuli, Design, and Procedure. For this experiment we reverted back to the stimulus set used in Experiment 1, consisting of 10 distinct colors. The general trial procedure was also the same as Experiments 1 and 2 with the exception that the test array would only contain a single item presented at the center of the screen, which was never occupied by memory array items (see Figure 1C). On *same* trials this probe item was taken from the presented set, whereas on *change* trials this was a color taken from the remaining items not presented during study.

Like Experiment 1, in this experiment the participants' expectation of a change occurring on a given trial was manipulated across 9 blocks. At the beginning of each block participants were informed that either 30%, 50%, or 70% of trials in that block would be change trials. Participants cycled through these base rates in a random order 3 times with the different set sizes (2, 5, 8) mixed together.

Results and Discussion

Hit and false-alarm rates for Experiment 3 are plotted in Figure 5 along with posterior predictive samples from the uninformed and informed guessing models. As in Experiments 1 and 2 comparison of the penalized fit statistic, WAIC, preferred the informed model over the uninformed one (ΔWAIC = 226.03). The informed model was also preferred over the hierarchical model used to estimate hit and false-alarm rates (ΔWAIC = 79.69).

Despite the good overall fit of the informed model, it is clear from Figure 5 that there is considerable overlap between the three base-rate conditions, which may limit our ability to provide a fair comparison of the two accounts. Further the estimates of uninformed guessing parameters (u) in this experiment (Table 1) suggest a strong overall bias towards responding change (all estimates > 0.6) and no great difference between the three base-rate conditions. It is possible that our base-rate manipulation was ineffective in modifying the observer's expectation of change. However, it is not clear why this manipulation would be less effective with a single probe than a whole display. Alternatively, the bias towards responding *change* may have, unintentionally,

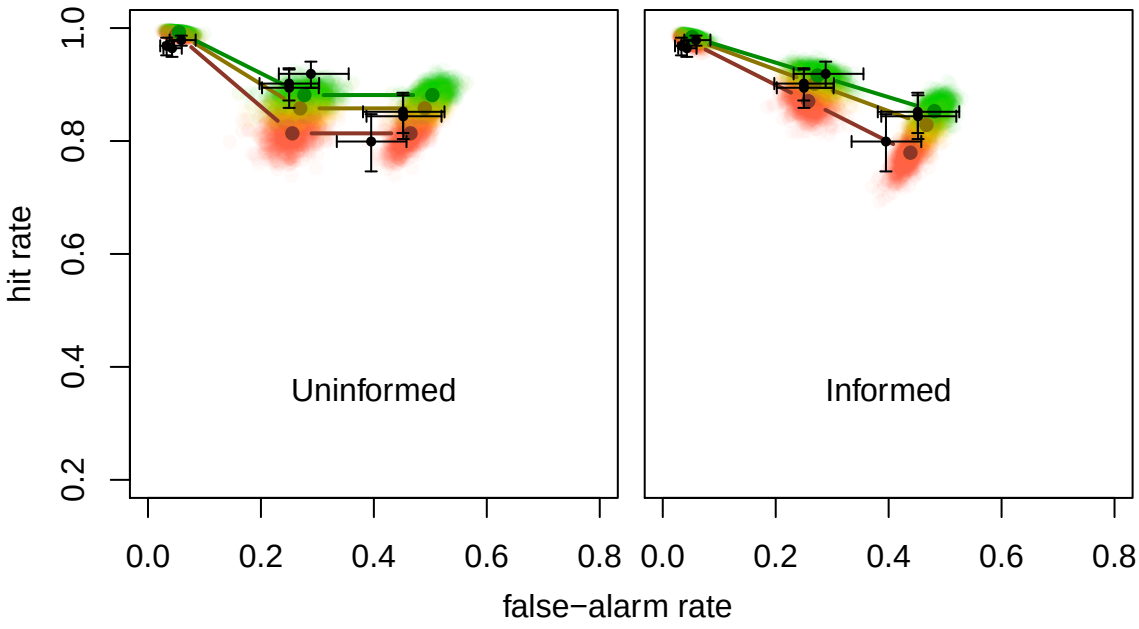

Figure 5. Estimated false-alarm and hit rates (posterior medians and 95% highest density intervals) from Experiment 3 shown in black over posterior predictive samples from the uninformed (left panel) and informed (right panel) guessing models (red = samples from the 0.3 base-rate condition, yellow = 0.5, green = 0.7). Points progressing along the negative diagonal denote the set size 2, 5, and 8 conditions, respectively. Posterior median predictions are also given joined by lines.

been caused by the nature of probe. The placement of the probe item in the center of the screen immediately following the pattern mask had the effect of producing apparent motion where the mask items all collapsed to a single point. This effect may have contributed to the impression that a change had occurred between study and test. Alternatively, in spite of being instructed that the probe item would always appear at the center where no memory items would appear, participants may have been biased towards saying *change* due to the mismatch of color and location. Experiment 4 attempted to eliminate these potential biasing factors.

Experiment 4

In this experiment, rather than presenting a single square as the probe item, we used an annulus that surrounded the area in which memory items were presented. This

annulus was present throughout the trial and changed to the color that served as the probe feature at the offset of the pattern mask, thereby avoiding the apparent movement effect in the previous experiment and emphasizing the irrelevance of location in the judgement. Figure 6 depicts an example trial.

Method

Participants. For Experiment 4, 32 students (12 female) from the University of Missouri, aged between 18 and 21 (mean = 18.7), took part in return for course credit for the 45 minute session. One participant did not complete the full session and two participants were excluded for failing to achieve greater than 60% accuracy averaged across all experimental blocks, leaving 29 for analysis.

Stimuli, Design, and Procedure. The stimuli, design, and general procedure was almost identical to those used in Experiments 1 and 3. The major difference was that an annulus with radius of 7° and thickness of approximately 0.5° appeared throughout the trial in a darker grey than the background (see Figure 6). This annulus surrounded the memory array and immediately following the offset of the pattern mask was filled with a single color, which participants had to judge as either being selected from the initial study array or being a color from outside this set.

Results and Discussion

Estimated hit and false-alarm rates for this fourth experiment are presented in Figure 7 superimposed on posterior predictive samples for the two guessing models. Relative to the ROC points in Experiment 3, which used a central probe (Figure 5), the present points exhibit a greater spread, with larger differences between the different conditions (see also estimates of u in Table 1). What is also clear from Figure 7 is that the fit of the informed model to the aggregate data, while far from perfect, is much better than that of the uninformed model. As discussed above, the uninformed model predicts that hit rate should not change between set sizes 5 and 8, which is clearly not the case in the estimated rates. Hit rate dropped by -0.098 [-0.148, -0.05] between set sizes 5 and 8 collapsed across the three base rate conditions (0.3 = -0.135 [-0.241,

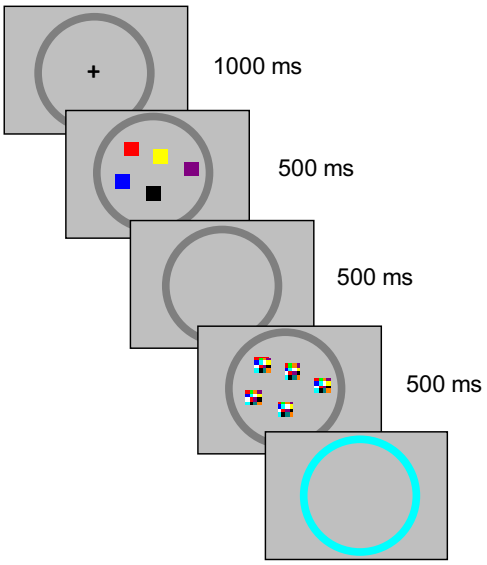

Figure 6. Illustration of the task used in Experiment 4. The figure depicts a trial on which the correct response is 'change'.

-0.032]; 0.5 = -0.097 [-0.177, -0.019]; 0.7 = -0.062 [-0.133, 0.003]). Once again the sampled WAIC values favored the account provided by informed guessing (ΔWAIC = 166.08). The 5 parameter informed model is also preferred over the model which freely estimated the hit and false-alarm rates from the raw data (ΔWAIC = 76.28).

Before proving an overview of our findings it is worth discussing some unexpected results and alternative models that allow for flexibility in the extent to which information is utilized when guessing.

Exploring Alternative Models

In the informed guessing model that we have applied here, participants are assumed to update their base expectation of a change with knowledge of the proportion of the array they retained (d) and then *probability match*; that is, choose between the options in accordance with their probabilities. Approximate probability matching has been observed in visual working memory tasks before (Cowan et al., 2016; Rouder et al., 2008). However, an optimal approach would be to select the response option with maximum likelihood all of the time (Friedman & Massaro, 1998; Shanks, Tunney, & McCarthy, 2002) and it is possible that people vary in the extent to which they use

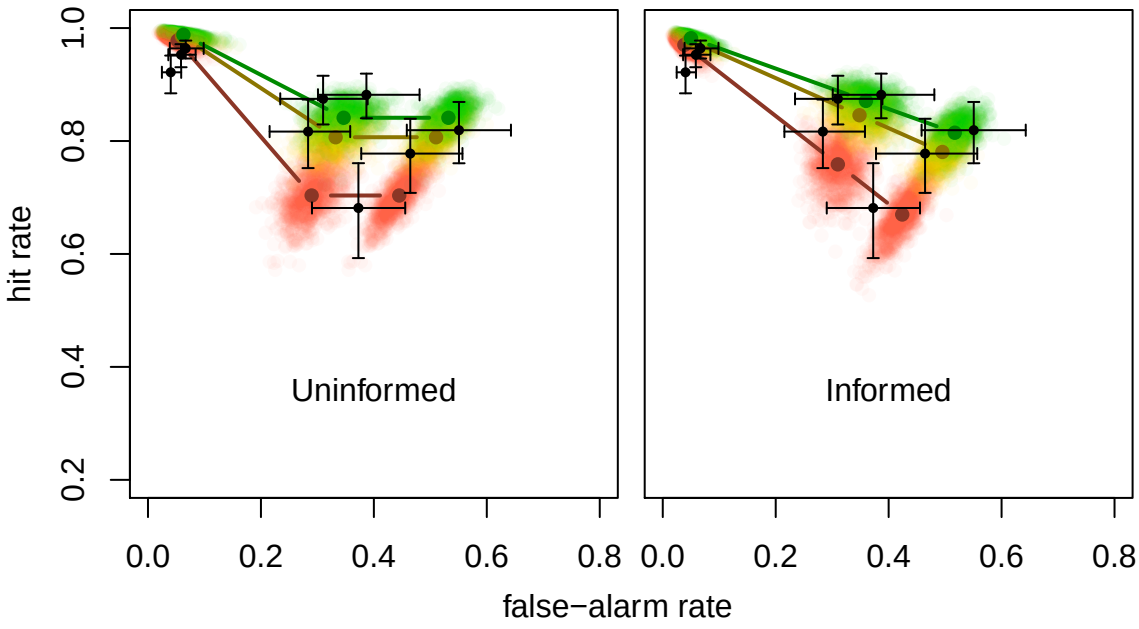

Figure 7. Estimated false-alarm and hit rates (posterior medians and 95% highest density intervals) from Experiment 4 shown in black over posterior predictive samples from the uninformed (left panel) and informed (right panel) guessing models (red = samples from the 0.3 base-rate condition, yellow = 0.5, green = 0.7). Points progressing along the negative diagonal denote the set size 2, 5, and 8 conditions, respectively. Posterior median predictions are also given joined by lines.

information efficiently in tasks like these. Hardman and Cowan (2016) recently addressed this very issue. They presented participants with an array of colored triangles at different orientations and then tested memory with either a single change detection probe, which was either a conjunction of features from the array or a new combination, or with a feature matching task, in which participants were presented with one feature of a studied item (e.g. orientation) and had to select the accompanying feature (e.g. color). They then fit a mixture model in which participants could use information minimally, ideally, or in an intermediate fashion, in-between these two extremes. Minimal information use was analogous to our uninformed guessing model where participants select from possibilities with no consideration for other information (see Chen & Cowan, 2013). Their ideal guessing procedure used a process of elimination,

where knowledge of other features in the array could be used to rule out candidates at test (as features were sampled without replacement). Crucially, the mixture parameters in this model allowed for participants to use a combination of these modes of responding. Consistent with our modeling results, Hardman and Cowan (2016) found that more often participants could be characterized as ideal information users; but there was a great deal of variability.

One suggestion that participants in our experiments may be responding more optimally—that is, selecting the most likely response option more often—than accounted for by the informed guessing model comes from the estimates of u. These estimates invariably deviated from the base-rates provided to participants (see Table 1) and they did so in different ways depending on the probe used. In the whole display experiments this is reflected in a bias towards expecting *same* and (as noted above) in the single probe experiments, a bias towards *change*. As described in the Introduction, the different constraints on information use in these two tasks lead to informed guessing predicting an updating of u towards *same* for whole display and towards *change* for single probe. Therefore, one account of the aberrant uninformed guessing rates may be that observers are going with the more likely option more often than predicted by simple probability matching.

To explore this possibility further we evaluated two additional models that allow for variation in the extent to which guessing behavior matches that of a probability matcher or an optimal responder. The first model included a mixture parameter, P^{OG}, which determined the probability that the observer responds optimally. Optimal guessing always selects *change* if $g > 0.5$ and *same* if $g < 0.5$, whereas if $g = 0.5$ a random choice is made between the two options. The second model considered offers more gradation between inefficient and ideal information use via a single parameter that determines the 'noise' in the choice probabilities (see Friedman & Massaro, 1998). In this model the probability of guessing *change* is determined by $1/(1 + e^{-\text{logit}(g)\lambda})$, where $\text{logit}(g)$ is the log odds transformation of the informed guessing rate and the parameter λ controls how efficiently information is utilized. An estimate of 1 for this parameter is

identical to the informed guessing models assessed above, whereas a λ value less than 1 is indicative of inefficient information use and large values approximate optimal guessing (as it increasingly approaches a step function). Thus these models correct for a potential source of mis-fit in the informed guessing model in that they allow for variation, both within and between participants, in how efficiently information is used to update the base expectation of a change occurring between study and test.

These models and the resulting parameter estimates are described in full in the Supplementary Material, but it suffices to say that the two were in general agreement. Averaged across the four experiments the estimate of the mixture parameter, P^{OG}, suggested that participants guessed optimally approximately a quarter of the time (0.282 [0.145, 0.408]). Estimates of λ from the second model were also consistently greater than 1 (1.699 [1.348, 2.161]), which also points towards generally more efficient information use than is currently captured by the informed guessing model. However, these estimates also show that participants are far from responding wholly optimally in this task which, along with the reasonable fit displayed in the figures, suggests that the probability matching assumption made in the informed guessing model is not a gross misspecification. At the individual level there was evidence of variation in the extent to which participants used information to guide guessing behavior, in line with the findings of Hardman and Cowan (2016) (see the Supplement). Future work examining individual differences in change detection performance and how model parameters covary (see Morey, 2011) may benefit from these extended models.

Nevertheless, there are other important possibilities that should be addressed by future work. For example, the observed bias in responding, towards *same* with a whole display and towards *change* with a single probe, could be explained by a systematic bias in participants' evaluation of the number of items they have in memory. For example, an overconfident observer may think "if there was a change I would have detected it" with a whole display (producing a same bias), whereas with a single probe they would think "if there was a match I would have detected it" (producing the opposite bias). Specifically, if participants *overestimate* what proportion of the array they have in

mind, their base expectations of a change occurring will undergo a more extreme updating, given that the participant's *belief* about the number of items they have in memory is the determining factor in driving their guessing behavior. Cowan et al. (2016) found a tendency for such overestimation when participants were probed to report the number of items in memory, relative to model-based estimates. Including these judgements in further work will be important in distinguishing the extent to which observers use information optimally from the extent to which they under- or over-estimate their capacity.

General Discussion

Estimating the number of items that an observer could retain, on average, across a series of change detection trials requires a specification of the way in which information can be used to perform the discrimination (Rouder et al., 2011). In some cases the task parameters allow observers to use information in memory to guide responses even when they are in a state of uncertainty; that is, when the relevant information is not present in working memory. However, whether or not people actually engage in this kind of informed guessing is unclear (although see Hardman & Cowan, 2016). Across four experiments the present work varied 1) the base-rate of a change occurring between study and test, 2) the number of to-be-retained items, and 3) the number of changes that could occur, to contrast the predictions of uninformed and informed guessing models. In each case, with both whole display and single probe versions of the task, the data suggested that the informed guessing model gives a better account. Parameter estimates from the informed model, specifically those controlling participants' base expectation of a change, suggested that information use may be somewhat more efficient than allowed for in the model. Therefore, we outlined two additional models that allow for exploration of the efficiency of information use. Below we discuss the implications of not adequately accounting for the way in which observers utilize information for the interpretation of experimental effects in change detection tasks along with some limitations of the present work.

Our findings allow for a more precise specification of the simple processing models used to estimate the number of items in working memory and, consequently more principled estimation. Failure to account for this information use will bias estimation and may well change the pattern of experimental effects. To illustrate this we fit versions of the informed and uninformed guessing models with separate estimates of k for each set size to the data from Experiments 1–4. If the average number of items that can be retained is fixed, then these variable k models should not outperform the versions with a single k parameter (Rouder et al., 2008), whereas evidence for the variable version trumps the notion of a capacity limit. For Experiments 2 and 3 the uninformed variable k model had a lower DIC relative to the uninformed fixed k models used above (for Experiment 4 the difference was unconvincing, 0.57 and WAIC always favored the fixed k version). However, for all data sets the informed fixed k model outperformed the variable k version.[5]

We also applied the simpler models of Pashler (1988) and the 'reverse-Pashler' formula described by Cowan et al. (2013) to our whole display and single probe data, respectively, as these easily applied estimators do not account for informed guessing. These estimates of k yielded significant effects of set size in each of the four data sets (see Supplementary Material). This, along with the findings from the DIC comparison above, acts as a proof-of-concept that failure to appropriately account for information use in the change detection task can affect experimental effects (see Rouder et al., 2011, for a similar demonstration). If we apply the uninformed guessing model we may conclude that participants were able to encode more information, on average, from large arrays (8 vs. 5), whereas the better fitting informed model suggests that this is not the case and participants are just making better use of the information available to them (see also Hardman & Cowan, 2016).

[5]Readers may note that relaxing the fixed k assumption and allowing for the possibility that, on average, more items are encoded from larger set sizes allows the uninformed model to better predict the the shape of ROCs in Experiments 1, 3, and 4, and the dependence of false-alarm rate on possible number of changes in Experiment 2. Nevertheless, the fixed k informed guessing model still outperforms the uninformed variable k model for each experiment in terms of DIC and WAIC.

Our findings are in line with the current literature on participants' knowledge of the contents of working memory. Previous research has suggested that confidence ratings track the precision of subsequently recalled items (Rademaker et al., 2012) and that participant reports of the number of items in memory accord well with model based estimates of capacity (Cowan et al., 2016). The present work suggests that participants are able to use this knowledge of the contents of memory to update their base expectation of change occurring during both whole display and single probe change detection. As currently implemented the informed guessing model assumes that the participant has perfect knowledge of the number of items in memory. However, it may be that participants systematically overestimate their capacity, as recently suggested by the data of Cowan et al. (2016). As we note above the parameter estimates for participants' base expectation of change (u) were more extreme than expected. In the preceding section we outlined models that allow for more flexible information use that may be able to account for some of this. However, a tendency for participants to overestimate their capacity could also produce a similar pattern of results. This is because it is the participants' belief about the proportion of the array they have in memory that determines the extent to which prior expectation of a change is updated. If participants overestimate the proportion of the array they have in memory, this updating will be more pronounced than if they perfectly knew their capacity. Future work on the meta-cognition of working memory will be needed to distinguish more efficient use of information in memory from a general tendency to overestimate one's capacity.

A related question that we are unable to address with the present data is just what estimate of the number of items in memory participants are working with when they inform their guessing. Contrary to the notion that the number of items in working memory is fixed on every trial, recent work has suggested that either fluctuations of attention (Adam, Mance, Fukuda, & Vogel, 2015) or in capacity itself (Cowan et al., 2016; van den Berg, Awh, & Ma, 2014) result in different numbers of items in memory from trial to trial. The present models cannot capture this aspect (although fractional

values of k may be interpreted as mixtures; so for example 3.3 could be seen as a k of 4 30% of the time and 3 the rest of the time). Participants may use the average number of items they can keep in mind to update their guessing rate or, as the previous findings suggest, they may use trial-to-trial knowledge to do this updating. Future work including probes for introspection on the numbers of items in memory or using more information rich paradigms (e.g. recall) may be able to distinguish these two. Nevertheless, the present findings allow users of standard change detection paradigms to obtain more reliable estimates of the key parameters of interest.

Summary and Conclusions

Estimates of the number of items that observers can retain over brief intervals are of great interest to cognitive psychologists and the change detection paradigm is often used to obtain such estimates. To estimate this quantity in a principled manner a model of the way in which information is used to perform the discrimination is needed. For some versions of this task fairly simple assumptions can be made regarding what observers do when they are in an uncertain state. For other versions there is more information available to observers that may guide their guessing behavior. The four experiments reported here suggest that patterns of hits and false-alarms in whole display and single probe (not in location) change detection are better reconciled with the assumption that observers use knowledge of the number of items in memory to inform guessing. Exploratory modeling suggests that there is variability between individuals in the extent to which information is utilized, but other potential explanations remain to be ruled out.

Table 1

Table of parameter estimates (posterior medians and highest density intervals) of population-level mean parameters of the informed and uninformed guessing models for Experiments 1-4. p_{DIC} and p_{WAIC} are estimates of effective number of parameters for DIC and WAIC, respectively.

Exp	Model	k	a	$u_{0.3}$	$u_{0.5}$	$u_{0.7}$	DIC (p_{DIC})	WAIC (p_{WAIC})
1	Uninformed	3.52 [3.22, 3.80]	0.92 [0.89, 0.95]	0.13 [0.09, 0.17]	0.19 [0.14, 0.25]	0.28 [0.21, 0.35]	2269 (150)	2256 (116)
	Informed	3.58 [3.36, 3.81]	0.92 [0.88, 0.95]	0.24 [0.18, 0.30]	0.34 [0.27, 0.40]	0.44 [0.36, 0.51]	2417 (165)	2462 (174)
2	Uninformed	3.13 [2.70, 3.55]	0.92 [0.88, 0.95]	-	0.13 [0.10, 0.17]	-	2042 (80)	2080 (101)
	Informed	3.12 [2.81, 3.42]	0.88 [0.84, 0.92]	-	0.30 [0.24, 0.36]	-	2241 (114)	2344 (181)
3	Uninformed	3.66 [3.33, 3.99]	0.94 [0.91, 0.96]	0.81 [0.77, 0.86]	0.86 [0.82, 0.89]	0.88 [0.85, 0.91]	2182 (142)	2173 (113)
	Informed	3.71 [3.45, 3.97]	0.93 [0.91, 0.95]	0.67 [0.60, 0.73]	0.73 [0.68, 0.78]	0.77 [0.72, 0.81]	2320 (151)	2399 (180)
4	Uninformed	3.18 [2.81, 3.57]	0.93 [0.88, 0.96]	0.70 [0.63, 0.77]	0.81 [0.75, 0.86]	0.84 [0.79, 0.88]	2248 (138)	2266 (130)
	Informed	3.11 [2.69, 3.53]	0.93 [0.88, 0.96]	0.56 [0.48, 0.64]	0.69 [0.62, 0.76]	0.74 [0.68, 0.80]	2361 (146)	2432 (179)

Appendix

Model Details

Whole Display. For whole display change detection, if a change has occurred between study and test, the probability that an item in memory has changed, and thus the change is detected, for the i^{th} participant, j^{th} set size (N), and m^{th} possible number of changes (C) is given by:

$$d_{ijm} = 1 - \frac{\binom{N_j - k_i}{C_m}}{\binom{N_j}{C_m}},$$

provided $k_i < N_j - C_m + 1$ otherwise $d = 1$. This reduces to $d = \min(k/N, 1)$ for $C = 1$. Thus the probability of correctly detecting a change in the l^{th} base rate condition is:

$$h_{ijlm} = a_i(d_{ijm} + (1 - d_{ijm})\psi) + (1 - a_i)u_{il},$$

whereas, for trials on which there was no-change the probability of incorrectly indicating that a change has occurred is given by:

$$f_{ijlm} = \begin{cases} (1 - a_i)u_{il}, & \text{if } d_{ijm} = 1 \\ a_i\psi + (1 - a_i)u_{il}, & \text{otherwise} \end{cases}$$

In this formulation, ψ controls the model of guessing under consideration. For an uninformed guesser $\psi = u_{il}$, whereas assuming an informed (probability matching) observer,

$$\psi = g_{ijlm} = \frac{(1 - d_{ijm})u_{il}}{(1 - d_{ijm})u_{il} + (1 - u_{il})}.$$

Single Central Probe. With a single probe, the probability that a match is correctly identified for the i^{th} participant and j^{th} set size is:

$$d_{ij} = \min(k_i/N_j, 1).$$

Consequently, the probability of incorrectly indicating that a change occurred when in fact the probe matches an item from the study set in the l^{th} base rate condition is:

$$f_{ijl} = a_i(1 - d_{ij})\psi + (1 - a_i)u_{il},$$

whereas, when the single item probe truly was not one of the original studied set the probability of correctly identifying this was:

$$h_{ijl} = \begin{cases} 1 - (1 - a_i)(1 - u_{il}), & \text{if } d_{ij} = 1 \\ a_i\psi + (1 - a_i)u_{il}, & \text{otherwise} \end{cases}$$

Once again, ψ determines the type of guesser under consideration. For the uninformed model $\psi = u_{il}$, whereas, for the informed model:

$$\psi = g_{ijl} = \frac{u_{il}}{u_{il} + (1 - d_{ij})u_{il}}.$$

Priors. In fitting the models to each data set, each trial was assumed to be distributed as a Bernoulli with probability of success determined different parameters (appropriate to the task at hand) depending on whether the trial was a same or change trial.

In both the uninformed and informed models the free parameters were k_i, a_i, and u_{il} and hierarchical priors were placed on transformations of these parameters; for a and u, as these parameters are constrained to fall within $[0, 1]$, the logit transformation was used. For k we used the mass-at-chance transformation, $k_i = \max(\kappa_i, 0)$, where the prior is placed on κ (Morey, 2011).

Participant level parameters were sampled from a normal distributions,

$$\begin{aligned} \kappa_i &\sim \text{Normal}(\mu^{(\kappa)}, \sigma^2_{(\kappa)}) \\ \text{logit}(a_i) &\sim \text{Normal}(\mu^{(a)}, \sigma^2_{(a)}) \\ \text{logit}(u_{il}) &\sim \text{Normal}(\mu^{(u)}_l, \sigma^2_{(u)}), \end{aligned}$$

and weakly-informative priors were placed on both the μ and σ parameters. For μ parameters Normal priors were used with a mean that depended on the parameter in question and a standard deviation of 10 (broad on the k and logit scales). The mean of the $\mu^{(u)}$ priors was set to zero and $\mu^{(a)}$ was set to 3 to reflect the expectation that participants would attend on the majority (around 95%) of trials. For k, $\mu^{(\kappa)}$ mean was set to 1 for technical reasons associated with the generation of initial values and the implementation of choose functions in JAGS for the whole display models. Setting the

prior mean to 3 for the data from Experiment 1 (which is more in accordance with our prior expectation) did not change anything, owing to the broad standard deviations used. For σ parameters the same generic prior was used for each of the parameters;

$$\sigma \sim \text{Gamma}(1.01005, 0.1005012).$$

This broad gamma prior better controls shrinkage relative to other typically used priors on variance parameters (Kruschke, 2015).[6]

Hit/ False-alarm Rate Model

To estimate the hit and false-alarm rates depicted in the Figures we applied a hierarchical Bayesian model which freely estimated each participant's rates as draws from different population distributions for each condition (cf. Donkin et al., 2016). More concretely, for the i^{th} participant, j^{th} set size, l^{th} base rate, and m^{th} possible number of changes the probability of responding *change* for the n^{th} trial type, which was either change or same (no-change), was assumed to be distributed as a Bernoulli with probability of success;

$$\text{logit}(P_{ijlmn}) \sim \text{Normal}(\mu^{(P)}_{jlmn}, \sigma^2_{(P)}).$$

Thus the hit and false-alarm rates for each condition were sampled from independent normals with different means but a shared standard deviation, providing an important shrinkage constraint on our estimation. Priors on the $\mu^{(P)}$ components were all normals with a mean of 0 and SD of 10 and the same gamma prior, described above, was placed on $\sigma^{(P)}$.

Model Estimation

Samples from the joint posterior distribution of parameters for each of the guessing models and the hit/ false-alarm rate model were obtained using JAGS (Plummer, 2003) and the `R` package `R2jags` (R Core Team, 2015; Su & Yajima, 2015).

[6] see also `http://doingbayesiandataanalysis.blogspot.com/2012/04/improved-programs-for-hierarchical.html`

We ran 4 chains of 5000 samples each following a burn-in period of 5000 samples. Convergence of the hierarchical parameters was identified when all values of the Brooks-Gelman-Rubin potential scale reduction factor fell below 1.1 (Brooks & Gelman, 1998; Gelman & Rubin, 1992). Where possible we did not thin chains and retained the full set of 20000 samples (Link & Eaton, 2012), however occasionally it was necessary to thin and retain every 5^{th} or 10^{th} sample to obtain convergence (20000 samples were still retained). The code written to fit the models is provided here `https://github.com/stephenrho/Guessing`.

References

Adam, K. C., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience*, *27*(8), 1601–1616.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455.

Chen, Z., & Cowan, N. (2013). Working memory inefficiency: Minimal information is utilized in visual recognition tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1449–1462.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 731–747.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100.

Cowan, N., Hardman, K., Saults, J. S., Blume, C. L., Clark, K. M., & Sunday, M. A. (2016). Detection of the number of changes in a display in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 169–185.

Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. *Cognitive Psychology*, *85*, 30–42.

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review*, *120*(4),

873–902.

Donkin, C., Tran, S. C., & Nosofsky, R. (2014). Landscaping analyses of the ROC predictions of discrete-slots and signal-detection models of visual working memory. *Attention, Perception, & Psychophysics*, *76*(7), 2103–2116.

Friedman, D., & Massaro, D. W. (1998). Understanding variability in binary and continuous choice. *Psychonomic Bulletin & Review*, *5*(3), 370–389.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 3). Chapman & Hall/CRC Boca Raton, FL, USA.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.

Gibson, B., Wasserman, E., & Luck, S. J. (2011). Qualitative similarities in the visual short-term memory of pigeons and people. *Psychonomic Bulletin & Review*, *18*(5), 979–984.

Gilchrist, A. L., & Cowan, N. (2014). A two-stage search of visual working memory: investigating speed in the change-detection paradigm. *Attention, Perception, & Psychophysics*, *76*(7), 2031–2050.

Hardman, K. O., & Cowan, N. (2016). Reasoning and memory: People make varied use of the information available in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 700–722.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 683–702.

Johnson, M. K., McMahon, R. P., Robinson, B. M., Harvey, A. N., Hahn, B., Leonard, C. J., ... Gold, J. M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology*, *27*(2), 220–229.

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Link, W. A., & Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, *3*(1), 112–115.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356.

Morey, R. D. (2011). A Bayesian hierarchical model for the measurement of working memory capacity. *Journal of Mathematical Psychology*, *55*(1), 8–24.

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics*, *44*(4), 369–378.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125).

Plummer, M. (2008). Penalized loss functions for bayesian model comparison. *Biostatistics*, *9*(3), 523–539.

R Core Team. (2015). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

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*(13), 21–21.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975–5979.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic*

Bulletin and Review, *18*, 324–330.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

Su, Y.-S., & Yajima, M. (2015). R2jags: Using R to Run 'JAGS' [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=R2jags` (R package version 0.5-7)

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149.

Van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological Review*, *124*(2), 197–214.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48–64.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135.