An Interference Model for Visual Working Memory: Applications to the Change Detection Task

Hsuan-Yu Lin & Klaus Oberauer

University of Zurich

Author Note

Hsuan-Yu Lin and Klaus Oberauer, Department of Psychology, University of Zurich. This research was supported by grants from the Swiss National Science Foundation (project n° 100014\_135002 and 100014\_179002) to Klaus Oberauer. We thank Wei-Ji Ma for helping us extending the Bayesian Interference rule on IM and Jasmin Stöckli for collecting the data.

Correspondence should be addressed to Hsuan-Yu Lin, Department of Psychology, Cognitive Psychology Unit, University of Zurich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: [h.lin@psychologie.uzh.ch](mailto:h.lin@psychologie.uzh.ch)

The data and the analysis is available on Open Science Framework at: https://osf.io/xfdsb/

Abstract

Most studies of visual-working memory employ one of two experimental paradigms: change-detection or continuous-stimulus reproduction. In this study, we extended the Interference Model (IM; Oberauer & Lin, 2017Oberauer & Lin, 2017), which was designed for continuous reproduction, to the single-probe change-detection task. In continuous reproduction, participants occasionally report the non-target items instead of the target. The presence of non-target response is uniquely predicted by the IM, which relies in part on the interference of non-target items to explain the set-size effect. By presenting a probe from a non-target item, we can investigate the amount of interference from non-target items in change detection. As predicted by the IM, we observed poorer performance in rejecting a probe from a non-target item compared to a new probe (i.e., a cost due to intrusions from non-targets). We fitted the Interference Model along with the Variable Precision and the Slot-Averaging model to the data from two change-detection experiments. The models were equipped with a Bayesian inference rule based on the one used in Keshvari, van den Berg, & Ma (2013). The Interference Model successfully predicted the set-size effect and the non-target intrusion cost, whereas the Variable Precision (VP) and Slot-Averaging (SA) models failed to predict the intrusion cost at all. Even with additional assumptions enabling VP and SA to produce intrusion cost, the IM still performed better than VP and SA quantitatively.

Keywords: Visual-working memory, Interference Model, Change-detection task, Modeling

An Interference Model for Visual Working Memory: Applications to the Change Detection Task

Visual working memory has been studied primarily with two tasks: Change detection (Awh, Barton, & Vogel, 2007; Cowan, Blume, & Saults, 2013; Luck & Vogel, 1997; Rouder, Morey, Morey, & Cowan, 2011) and continuous reproduction (Bays, Catalao, & Husain, 2009; Souza & Oberauer, 2016; van Moorselaar, Olivers, Theeuwes, Lamme, & Sligte, 2015; Wilken & Ma, 2004; Zhang & Luck, 2008). In both tasks, a few items (i.e., squared color patches) are displayed at the beginning of the trial and participants are asked to remember the items. After a brief delay, the items disappeared from the screen, and one of item location is probed. In continuous reproduction task, participants are asked to reproduce the color at the probed location by selecting it on a continuous color wheel. For the change-detection task, specifically, the single-probe change-detection task, a color is presented in the probed location, and participants are asked to judge whether the color of the probe is the same as the original color presented at the same location. Whereas there are many computational models built for the continuous reproduction task (Bays, 2016; Bays et al., 2009; Oberauer & Lin, 2017; Zhang & Luck, 2008), there are only a few models designed for change-detection tasks (Donkin, Tran, & Pelley, 2015; Keshvari, van den Berg, & Ma, 2013). In this study, instead of creating new models specifically for the change-detection task, we adapted some of the existing models designed for the continuous-reproduction task to explain the data from the change-detection task[[1]](#footnote-1).

Our investigation builds on one previous study that adapted models initially designed for the continuous reproduction task to change detection. Keshvari et al. (2013) adapted the Slot-Averaging (SA; Zhang & Luck, 2008) model and the Variable-Precision (VP; van den Berg, Shin, Chou, George, & Ma, 2012) model to the change-detection task. In the meantime, we have proposed the Interference Model (IM) as a third alternative that performed well in comparison to SA and VP in explaining findings form the continuous-reproduction task (Oberauer & Lin, 2017). In the present study, we compare the IM with the SA and the VP for their ability to explain behavior in the change-detection task. Keshvari et al. (2013) compared the SA and VP models with regard to the full-array change-detection task, in which memory for an array of items is compared to a full test array. This task requires as many comparisons between memory items and test items as there are in the memory set, which entails two disadvantages: The comparison process is complicated and therefore difficult to model, and the memory set size (i.e., the number of items to be remembered) is confounded with the number of comparisons, making it difficult to separate the effects of the two variables. Therefore, we concentrate our efforts on the single-probe change-detection task, in which a single item selected at random from the array is presented at test.

The three models included in the present comparison have originally been designed to account for behavior in the continuous-reproduction task. Each model represents a different theory about the nature of visual working memory capacity. The Slot-Averaging model builds on the assumption that visual-working memory has a limited number of discrete slots, and a slot can only store one item or chunk. Once all the slots are used, the remaining items in the array will not be remembered at all. In the continuous reproduction task, if the target is remembered in a slot, the target will be recalled with the precision afforded by one slot. If the target is not remembered, however, the participant has to guess. One additional assumption in the Slot-Averaging model is that an item can be stored in multiple slots if there are free slots available. When recalling the items stored in multiple slots, the participant will retrieve the item from all the slots and respond with the average of the retrieved memory, which increases the precision of recall.

The Variable-Precision model incorporates the assumption that visual working memory is limited by a continuous resource. The resource can be continuously divided and distributed among all array items, and the precision of the memory representation increases when more resource is allocated to an item. The Variable-Precision model also includes the assumption that the resource is not evenly distributed among all the items, and the amount of resource allocated also varies from trial to trial. Hence, the precision of memory for the target varies from trial to trial.

The Interference Model is based on the assumption that the limit of visual working memory capacity is caused by interference between remembered items. Although participants can represent an unlimited number of items in working memory, these representations interfere with each other, and the precision of recall decreases as a result. In the Interference Model the recall process is driven by a retrieval cue that activates the response candidates (e.g., all the colors on the color wheel) to different degrees, and response candidates with higher activation are more likely to be recalled. The activation of the response candidates comes from three sources. The first source is the context-independent activation, which arises from persistent activation of all items encoded in the current trial. The second source is the context-based activation, which arises from using the context information by which the target is identified (e.g., its location given at test) as a cue to retrieve the content bound to it in memory (e.g., the target color). The last source of activation is the background noise, which arises during the encoding and retrieval process, and activates all response candidates uniformly.

In the Interference Model, several mechanisms are responsible for the capacity limit, that is, the decline of the precision of recall as memory set size increases: First, with larger set size more response candidates receive activation from the context-independent activation of all array items. Second, because the context representations (e.g., location in the array) have limited precision, a contextual retrieval cue activates not only the target item but also other items in nearby contexts (e.g., spatial neighbors). With increasing set size, it is more likely that several other items are near neighbors of the target on the context dimension. Third, because each item representation includes a constant amount of background noise, the amount of background noise encoded into working memory increases linearly with set size. Finally, an additional assumption in the IM is that one item can be stored in the focus of attention, and that item has higher precision and is resistant to the interference from the other items. As set size increases, the chance that the target is the item held in the focus of attention decreases.

The main difference between the continuous reproduction task and the change detection task is that in the latter, the information retrieved from memory has to be compared to the probe, and the person has to make a decision as to whether to respond "same" or "change". To adapt the models designed for the continuous reproduction task to the single-probe change detection task, we therefore need to model this comparison and decision step. To this end we implemented the optimal Bayesian inference rule as pioneered by Keshvari et al. (2013). This rule has the advantage that it requires no additional parameters, so that it provides no additional flexibility to the models. As such, we use it as a neutral common ground for our model comparison. Our choice of the optimal Bayesian decision rule does not reflect a commitment to the assumption that people actually make change-detection decisions in an optimal way. Nevertheless, we believe that the assumption of optimality is a useful and parsimonious starting point, from which we should depart only if the data force us to do so.

The Bayesian inference rule starts from the assumption that participants first try to retrieve the target feature (e.g., color) at the probed context (e.g., location) in the same way as in the continuous-reproduction task. Because in continuous-reproduction tasks, the to-be-remembered features are usually drawn from a circular dimension (e.g., orientations, or colors on a color wheel), and the response options are on a circular scale, we develop the decision rule for features on a circular dimension (as in Keshvari et al., 2013). Whereas in the continuous-reproduction task, that retrieved feature directly determines the chosen response (bar some motor noise), in the change detection task, the person compares it to the probe feature. Specifically, the person evaluates the probability that the retrieved color and the probe came from the *same* condition, and the probability that both came from the *change* condition. The response is given according to the option with the higher posterior probability. In the models, we compute a decision variable *d*, as

|  |  |
| --- | --- |
|  | (1) |

where the is the retrieved color and is the probe color. If the log posterior odds of the two conditions, , is larger than zero, the response will be *change*; otherwise, the response will be *same*. With the help of Bayes' theorem, Equation 1 can be rewritten as:

|  |  |
| --- | --- |
|  | (2) |

The values of and depend on the experiment design. When they are identical, as they are in most experiments, both cancel out in the equation.

Next we need to determine and . A first step is to decompose the joint conditional probabilities of the retrieved feature *x* and the probe :

|  |  |
| --- | --- |
|  | (3) |

For the probe is chosen at random from a uniform distribution on the circle, so that every possible probe has probability density The retrieved feature *x* comes from a distribution centered around the target feature, and the target feature is also selected at random from a uniform distribution independently of the probe. Therefore, the probability distribution of the retrieved feature is again a uniform distribution over all possible feature values on the circle. The probability of retrieving in the *change* condition is

|  |  |
| --- | --- |
|  | (4) |

Taking Equation 3 and Equation 4 together, we obtain

As for the , because the probe condition is *same*, the probe feature is identical to the target feature. As the target feature is drawn at random from a uniform distribution, is the probability of retrieving feature , given is the target feature. This probability depends on the model in question: Every model of continuous reproduction predicts the probability distribution of retrieving feature , given target feature *y*, based on two sets of information: the model parameters, and some additional information about the current trial. In the case of the SA model, the additional information is whether on the current trial the target is in memory (so that the probability distribution is a von-Mises distribution centered on the target feature) or is not in memory (so that the person has to guess, and the probability distribution is uniform). In the case of the VP model, the additional information is the precision of the target representation on the given trial, which varies randomly from trial to trial. In the case of the IM, the additional information consists of (1) the features of all other (non-target) items in the array, because their activation contributes to the probability distribution of retrieval, and (2) whether or not the target is in the focus of attention.

Whereas both the model parameters and the additional information about the current trial determine the actual probability of retrieving feature , given target feature *y*, predicted by a model, here we need to ask which information influences the *subjective* probability of retrieving feature , given target feature *y*, that is used by the person's decision process. We can reasonably assume that the decision process has a good estimate of the model parameters because these are person parameters that are relatively constant over time, and certainly constant across all trials in an experiment, so the person has ample opportunity to learn about them. It is not clear, however, whether the decision-making process has access to the additional information about each specific trial. For instance, for the SA model we must ask whether the decision-making process should be assumed to know whether, on a given trial, the target feature is represented in a slot, or whether it should only have knowledge about the probability of having the target in a slot for a given experimental condition. For the VP model we need to ask whether the decision-making process should be assumed to know the precision with which the target is represented on the current trial (as opposed to the average precision across trials for a given memory set size). For the IM we need to ask whether the decision process should be assumed to know all the non-target features and their locations in the array, and whether it should be assumed to know whether the target is in the focus of attention. Nosofsky and Donkin (2016) discuss these questions, distinguishing between *knowledge-limited* and *knowledge-rich* versions of recognition models, and they argue that knowledge-rich model versions, which are imbued with knowledge about each individual trial, strain psychological plausibility. Here we do not want to make an a-priori decision on how much knowledge should be attributed to the decision process, and therefore we compared all the models with different level of knowledge in the inference rule. Regardless the level of knowledge of the inference rule, the retrieval probability of feature *x* in the *same* condition can be generally expressed as a weighted mixture of a von-Mises distribution centered on the target, , and a uniform distribution. The models differ in how they determine the mixture weight *Ps*.

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Taking Equation 5 together with Equation 3, we obtain

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

Here, is the probability of retrieving the target item, and is the precision of retrieval. The precision of retrieval, , is a parameter in all three models we aim to compare. It turns out that can be eliminated from the decision equation in the end, as explained in Appendix A.

Assuming equal priors, the decision variable can now be computed as the log-odds ratio of the two likelihoods:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

To determine the probability of a "change" response, we integrate over all possible retrieved features , as

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

where the is the probability of making a “change” response when given the array of memory items and the probe .

The advantage of using the Bayesian inference rule is that, as long as a model provides and , the model can be adapted to the single-probe change-detection task without an additional parameter. All the models under consideration here predict because this is the likelihood of responding *x* in the continuous-reproduction paradigm.

We used the optimal Bayesian decision rule to make predictions for change-detection behavior from the three models (SA, VP, and IM) with different levels of knowledge in the inference rule. The details of the models and how the inference rule was applied to the models are explained in Appendix B. We tested these predictions with two experiments, described next.

# Experiments A and B

The goal of Experiment A and B was to acquire the data for model comparison using the single-probe change-detection task, varying memory set size from 1 to 6. The inference rule (with equal priors) is based on the the assumption that 50% of the probes are exact matches, and the other 50% are drawn at random from a uniform distribution on the circle. In Experiment A we implemented exactly that sampling scheme. This probe distribution has two potential drawbacks. One is that some of the *change* trials will have probes very similar to the target, so that participants are very likely to judge them as "same". As a consequence, they might make more than 50% "same" responses, which could lead them to expect that there are more *same* than *change* trials. Hence, although the objective prior probability of both trial types is identical, the subjective one would not be. The second drawback is that sampling *change* probes from a uniform distribution results in very few probes that match a non-target item in the array. These so-called *intrusion probes* are of interest because they are harder to identify as change probes than *new probes* not matching any array item (Donkin et al., 2015; Rerko, Oberauer, & Lin, 2014). Donkin et al. (2015) have argued that this intrusion cost – and in particular its interaction with set size – presents a challenge for slot models, and therefore we were interested in how well the SA model can accommodate them.

To mitigate these two shortcomings of the probe sampling scheme of Experiment A, we ran Experiment B, in which the probes from the *same* condition (50%) were sampled from a region of feature space around the target color for which we expect participants to perceive them as indistinguishable from the target. In addition, the probes from the *change* condition were subdivided into two categories: *Intrusion probes* (25%) were sampled from regions around the non-target colors, whereas *new probes* (25%) were sampled from the colors what are not in any region around the target or the non-target colors.

## Method

### Participants

Experiment A and Experiment B both included twenty participants recruited from the student population of the University of Zurich. Participants had normal or corrected-to-normal vision and no color blindness. Participants were rewarded with course credits or 60 Swiss Francs after completing the Experiment A, and 45 Swiss Francs after completing Experiment B.

### Material

Items in the memory arrays were color patches selected from a color wheel, which was created in CIE L\*a\*b\* color space with a radius of 60 and centered at luminance set to 70, *a* set to 20, and *b* set to 38. The minimum distance between selected colors was 1 degree, so there were 360 distinct colors. The stimuli were displayed in RGB values with Gamma correction for IEC 61966-2-1 standard.

### Procedure

Experiment A consisted of four identical sessions, and Experiment B consisted of three identical sessions. All the sessions were conducted on different days, and each session took about one hour to complete. Experiment A consisted of 2640 trials, and Experiment B consisted of 1260 trials. The procedure was identical for both Experiment A and B.

At the beginning of each trial, between 1 and 6 color patches were displayed simultaneously on the screen for 500 ms, which was followed by a blank screen for 500 ms. The locations of the color patches were randomly selected from 13 possible locations equidistantly distributed on an invisible circle which was centered on the center of the screen with 5 degrees visual angle when viewed from 70cm away from the screen. After the blank screen, a probe was displayed in one of the stimulus locations, and empty frames were displayed in the remaining locations that had been occupied by a stimulus in the current array. Participants were asked to judge if the probe was the same color as the color patch presented previously at the same location by pressing the left mouse button for “same”, or the right mouse button for “change.” After participants made their response, a blank screen appeared for 500 ms, and was followed by the beginning of the next trial.

In Experiment A, the probe matched the target color exactly in 50% of the trials. For the remaining 50% of the trial, the probe was selected randomly with equal probability from all possible colors other than the target color. In Experiment B, the probe color was selected within a boundary around the target color for 50% of the trials, which were defined as *same* trials. In 25% of the trials, the probe color was selected from within regions defined by boundaries around the non-target colors (*intrusion* trials). For the remainder of the trials, the probe color was selected from the remaining colors that were not within the boundaries of the target color and non-target colors (*new* trials). The only exception was in set size one because there was no non-target item, so all 50% of change trials were new trials. The boundaries for selecting the probe color around the target and non-target colors were chosen on the basis of error distributions from continuous-reproduction tasks using the same materials (Oberauer & Lin, 2017), so that they covered the range of colors that participants were likely to reproduce when trying to reproduce the selected color. Because the error distributions usually become broader with larger set sizes, these boundaries varied according to the set size of the trial, which was 10, 14, 16, 17, 18, and 18 degrees for set size one to six, respectively. The probe sampling schemes for Experiment A and Experiment B are illustrated in Figure 1.

### Results

To bring the result of Experiment A in line with the result of Experiment B, we relabeled the probes in Experiment A in a similar fashion as Experiment B. The *same* probes stayed the same[[2]](#footnote-2). The *change* probes were relabeled into *intrusion probe,* or *new probe*, where the former was any probe that was within the boundaries around a non-target color, and the latter was any probe not within the boundaries of any non-target color. The boundaries were the same as the boundaries used to sample probes in Experiment B.

After relabeling the probes, we compared mean accuracy for the three different probe types (same, intrusion, and new) and the six set sizes between experiments with a Bayesian linear model using the BayesFactor package for R (Morey & Rouder, 2013; R. Core Team, 2016) with default priors and a random intercept only. The Bayes factor suggests there was no difference between Experiments (BF01 = 1111.11). Therefore, the following analysis will be conducted with both experiments collapsed together. The Bayes factor showed evidence supporting both the main effects of set size and of probe type (BF10 = 5.92e+41 and BF10 = 3.23e+18, respectively). Evidence also supported the interaction between set size and probe type (BF10 = 1.59e+54). The statistic is summarized in Table 1, and these effects are visualized in the right panel of Figure 2.

# Model Comparison

We fitted the Interference Model, the Slot-Averaging Model, and the Variable Precision Model to the data of each experiment after embellishing each model with the Bayesian inference rule. The models were implemented in Python 3.6, and the parameters were estimated with the differential evolution algorithm in SciPy (Jones, Oliphant, Peterson, & others, 2001). The goodness-of-fit of the models was estimated by the deviance, which is –2\*log(Likelihood). The differential-evolusion is a global optimization algorithm, and the starting population was automatically chosen by the differential-evolution algorithm (Storn & Price, 1997). To compensate for the different number of parameters in the models, we used AIC, which penalizes models for their free parameters, to compare between models.

As mentioned above, there is some room for interpretation in implementing the optimal Bayesian inference rule (Nosofsky & Donkin, 2016): We need to make assumptions about what level of knowledge about memory is attributed to the inference rule. In the IM, we were unsure if the Bayesian inference rule has knowledge of whether or not the target is in the focus of attention on a given trial. If the target is in the focus of attention, the target would have higher precision and resistance to non-target interference. Having knowledge of whether the target item is in the focus of attention or not would affect the assumed precision of retrieval in the decision rule. We therefore compared two versions of the Bayesian inference rule. The winning model was the inference rule with the knowledge of whether the target is in the focus of attention or not. The following report was based on said version.

For the Slot-Averaging model, we tested two versions of the inference rule with different levels of knowledge. The first had the knowledge of whether or not the current target is in a slot in memory. If the target is in memory, the retrieved color always comes from the target color, i.e., , and is retrieved with precision . If the target is not in the memory, however, is a uniform distribution, and as such identical to . As a consequence, is 0, and the inference rule can only guess with 0.5 probability to respond "same" or "change". The second version of the inference rule does not use knowledge of whether the target is in the memory or not, and therefore relies only the general probability of retrieving the target for , and assumes precision for all trials. The fitting result showed that the second version of the inference rule performed much better than the first one, as shown in Figure 4. Therefore, for the following comparison, we used the inference rule which doesn’t have knowledge of whether the target is remembered or not on a given trial.

We also tested two versions of the inference rule for VP. The first one has the knowledge of the precision of the target item in the current trial, and the second version only has the knowledge of the average precision at the current set size. The fitting result showed that the version which only has knowledge of the average precision performed better, as shown in Figure 5, which was in contradict to the previous study (Keshvari et al., 2013). We used the VP model with the inference rule which only has knowledge of the mean precision of the current set size.

Overall, all the models were able to fit the similarity gradient effect for *change* probes, and the set-size effect for the *same* probes. However, only the IM was able to predict the worse performance of the *intrusion* probes compared to the *new* probes. The goodness-of-fit showed that IM was the best fitting model for both experiments. On average, the IM won over the SA and the VP by 29.55 and 23.89 AIC points per participant; see Table 2 for the summary of the goodness of fits. The SA and the VP fitted the data poorly mostly because they failed to accommodate the intrusion cost, and the fitting algorithm had to compromise between reproducing the performance on intrusion probes and new probes.

The failure of predicting the intrusion cost is due to the fact that the SA and VP models have no mechanism for generating an elevated chance of retrieving colors similar to the non-target colors. In both models the non-target features play no role in predicting the distribution of retrieved colors. This results in the same prediction for both intrusion probes and new probes. The IM, in contrast, predicts that non-target features are retrieved more often than new features because non-targets receive activation at retrieval. As a consequence, features matching an intrusion probe are more likely to be retrieved than features matching a new probe. This enables the IM to explain the intrusion cost.

To compensate for their lack of ability to predict the intrusion cost, we also fitted variants of the SA and the VP which were able to predict elevated retrieval probabilities for non-targets, namely: SA-Swap and VP-Binding. The SA-Swap model adds to the SA model the assumption that swap errors occur between array items at encoding (Donkin et al., 2015; Lin & Oberauer, 2019), and the probability of making such a swap error increases linearly with set sizes. The VP-Binding model builds on the assumption that – as in the IM – arrays are represented as color-location bindings with limited precision on two dimensions, the feature dimension (color) and the context dimension (location). At test, the location is used to retrieve the color bound to it. The precision of a binding on both dimensions is dependent on the resource amount allocated to it, which varies randomly as in the original VP. This version of the VP model can be thought of as a simplified version of the neural resource model proposed by Schneegans & Bays (2017) – although it does not incorporate all features of that model, the simplified version is more amenable to fitting. Please refer to Appendix B for more detail about the SA-Swap and VP-binding models.

Both the SA-Swap and the VP-binding model were able to capture the general trend in the data, including the intrusion cost as shown in Figure 6. However, AIC values – listed in Table 2 – indicated that SA-Swap and VP-binding still fit the data more poorly than the IM.

# Conclusion and General Discussion

In this study, we tested three models representing three different theories about visual working memory capacity on the single-probe change detection task. The Interference Model outperformed both the Slot-Averaging and the Variable-Precision model, even after we extended the SA and VP models to account for the intrusion cost we observed in the data.

The intrusion cost – an elevated false-alarm rate to probes matching a non-target compared to new probes not matching any array item – is consistently observed in change-detection task (Donkin et al., 2015; Rerko et al., 2014). The intrusion cost was also observed in verbal materials, for instance in the local-recognition task (Oberauer, 2008). The intrusion cost in the change-detection task is conceptually related to the above-chance tendency to report features of non-targets in the continuous-reproduction task: Both can be explained by a tendency to erroneously retrieve a non-target instead of the target. However, non-target related responses in continuous reproduction could also be explained as sophisticated guesses: A person not knowing the target feature could intentionally choose the feature of a non-target as response on the assumption that this might be a better guess than choosing any feature at random. The increased false-alarm rate to intrusion probes in change detection, in contrast, cannot be explained as a form of sophisticated guessing. If the participant did not remember the target color but remembered the non-target color, and the non-target color appeared as a probe in the target location, then the participant should reject the probe because they remembered that this probe had been in a different location in the array. As a consequence, we should observe an intrusion benefit. The intrusion cost can only be explained by either assuming that colors are occasionally remembered after losing their location information, so that people remember having seen the probe color but have no information about where in the array it was, or else by assuming that swap errors (i.e., retrieval of non-target features that are treated by the decision process as if they were the target feature) occur with above-chance probability.

In the presented study, the Slot Averaging model failed to explain the intrusion cost without including the assumption of swap errors at encoding. Moreover, the probability that a swap error occurs has to increase with set size to accommodate the observed intrusion cost. Donkin et al. (2015) has shown that a slot model with constant swap error probability would predict a decrease of the intrusion cost with increasing set size, which contradicted the findings in Donkin et al. (2015) and in the present study. However, with the assumption that the probability of swap error increases with set sizes, the SA model can explain the intrusion cost we observed, although it is difficult to explain why the swap error would increase with set sizes, especially when the set size exceeds the capacity limit. For instance, if the capacity limit is 3 slots, then 3 array items are stored regardless of whether set size is 3 or anything larger than 3, so there is no conceivable reason for the swap probability to increase beyond set size 3.

The Variable Precision model also cannot reproduce the intrusion cost without implementing an additional mechanism for generating non-target responses. Also, similar to the SA-swap model where the probability of swap error increases with set sizes, the VP-binding model also assumes that the probability of making the non-target response increases with set sizes. Different from the SA-swap model, this assumption is not made ad hoc but arises necessarily from the assumption that a limited resource governs the precision of bindings on the context dimension. With increasing set size, the target location becomes less discriminable from non-target locations, so that non-targets become more likely to be retrieved in response to the target location as cue.

Assuming higher degrees of trial-by-trial knowledge involved in the Bayesian inference rule has shown both benefits and harm to the model fitting depending on the models. For the IM, attributing to the decision process knowledge of whether the target is in the focus of attention or not improved the model performance. However, assuming knowledge of whether the target is in memory or not reduced the performance of the SA model and the VP model. The detriment to the SA model arises because when the decision process knows that the target is not in memory, it guesses, and both *same* and *change* are equally probable responses. They are equally probable because in the optimal decision rule guessing reflects the prior probability of the two probe types. As set size increases, the target is more often not in memory, and thus, the predicted performance for both *same* and *change* probes declines as the set size increases. This is contrary to the data of both experiments, in which set size predominantly affected accuracy on *same* trials. For the VP model, the higher degree of knowledge leads to similar predictions as the SA model with knowledge of the memory state of the target. Under lower precision in the current trial, the retrieval distribution from VP will be similar to a uniform distribution. If the inference rule is assumed to know that, on a given trial, target precision is very low, it will use a very low precision parameter for computing the likelihood . In that case, the inference rule can barely distinguish between the *change* and the *same* likelihood, and the response will be almost like random guessing. When the set size gets larger, the decision rule is more likely to encounter trials with low precision, and more guessing-like decisions will be made on both *same* and *change* trials, which is contrary to the observed data.

# References

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual Working Memory Represents a Fixed Number of Items Regardless of Complexity. *Psychological Science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Bays, P. M. (2016). A signature of neural coding at human perceptual limits. *Journal of Vision*, *16*(11), 4–4. https://doi.org/10.1167/16.11.4

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*(10), 7. https://doi.org/10.1167/9.10.7

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. https://doi.org/10.1037/a0029687

Donkin, C., Tran, S. C., & Pelley, M. L. (2015). Location-based errors in change detection: A challenge for the slots model of visual working memory. *Memory & Cognition*, 1–11. https://doi.org/10.3758/s13421-014-0487-x

Jones, E., Oliphant, T., Peterson, P., & others. (2001). *SciPy: Open Source Scientific Tools for Python*. Retrieved from http://www.scipy.org/

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No Evidence for an Item Limit in Change Detection. *PLOS Computational Biology*, *9*(2), e1002927. https://doi.org/10.1371/journal.pcbi.1002927

Lin, H.-Y., & Oberauer, K. (2019). How to say no in recognition tests of visual working memory: Testing unidimensional and two-dimensional models with continuous or discrete memory states. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000700

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Morey, R. D., & Rouder, J. N. (2013). *BayesFactor: Computation of Bayes factors for common designs*. Retrieved from http://CRAN.R-project.org/package=BayesFactor

Nosofsky, R. M., & Donkin, C. (2016). Qualitative contrast between knowledge-limited mixed-state and variable-resources models of visual change detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1507.

Oberauer, K. (2008). How to say no: Single- and dual-process theories of short-term recognition tested on negative probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 439–459. https://doi.org/10.1037/0278-7393.34.3.439

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59. https://doi.org/10.1037/rev0000044

R. Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Retrieved from http://www.R-project.org/

Rerko, L., Oberauer, K., & Lin, H.-Y. (2014). Spatial transposition gradients in visual working memory. *The Quarterly Journal of Experimental Psychology*, *67*(1), 3–15.

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 & Review*, *18*(2), 324–330.

Schneegans, S., & Bays, P. M. (2017). Neural Architecture for Feature Binding in Visual Working Memory. *Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, *78*(7), 1839–1860. https://doi.org/10.3758/s13414-016-1108-5

Storn, R., & Price, K. (1997). Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

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*, 201117465. https://doi.org/10.1073/pnas.1117465109

van Moorselaar, D., Olivers, C. N. L., Theeuwes, J., Lamme, V. A. F., & Sligte, I. G. (2015). Forgotten but not gone: Retro-cue costs and benefits in a double-cueing paradigm suggest multiple states in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1755–1763. https://doi.org/10.1037/xlm0000124

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Tables

Table 1

Summary of statistics

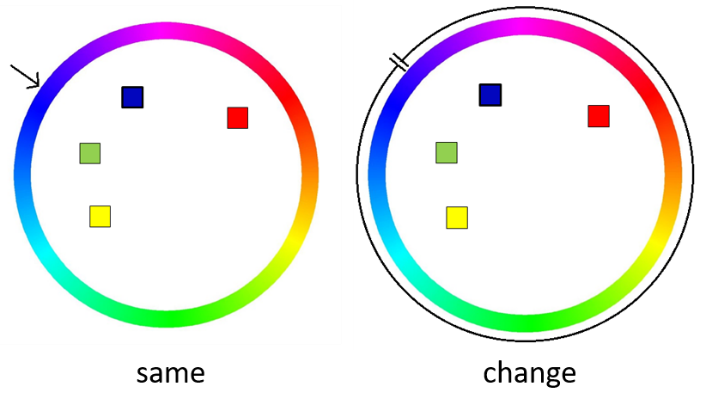
|  |  |
| --- | --- |
| Model | Bayes factor |
|  |  |
| Comparing to |  |
|  | 0.0009 |
| Comparing to |  |
|  | 5.92\*10^41 |
|  | 3.23\*10^18 |
| Comparing to |  |
|  | 1.59\*10^54 |

Table 2

Goodness-of-fit for Slot-Averaging model, Variable-Precision model, and Interference Model under different levels of knowledge attributed to the inference rule. The and Δ AIC represent the difference between each model’s AIC and the lowest AIC value of any model, averaged over participants; smaller values reflect better fit.

|  |  |  |
| --- | --- | --- |
| Model | Knowledge of the Inference Rule | ∆AIC |
| Slot-Averaging |  |  |
|  | Knowing whether the target is in a slot or not | 81.96 |
|  | Not knowing whether the target is in a slot or not | 29.55 |
| Variable-Precision |  |  |
|  | Knowing the precision of the current trial | 75.39 |
|  | Only knowing the average precision | 23.89 |
| Interference Model |  |  |
|  | Knowing whether the target is in the focus of attention or not | 4.89 |
|  | Not knowing whether the target is in the focus of attention or not | 0 |
| SA-Swap | Not knowing whether the target is in a slot or not | 17.15 |
| VP-Binding | Only knowing the average precision | 10.41 |

Figures



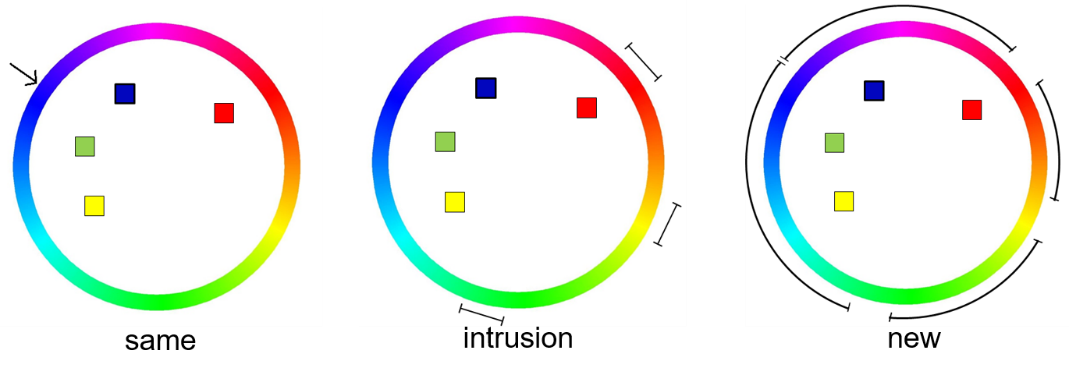


Figure 1. The probe sampling scheme for Experiment A and Experiment B. The four colors used in the trials are blue, red, green, and yellow, and the probe is presented at the blue color location. The top row is the probe sampling scheme for Experiment A, and the bottom row is the probe scheme of Experiment B.

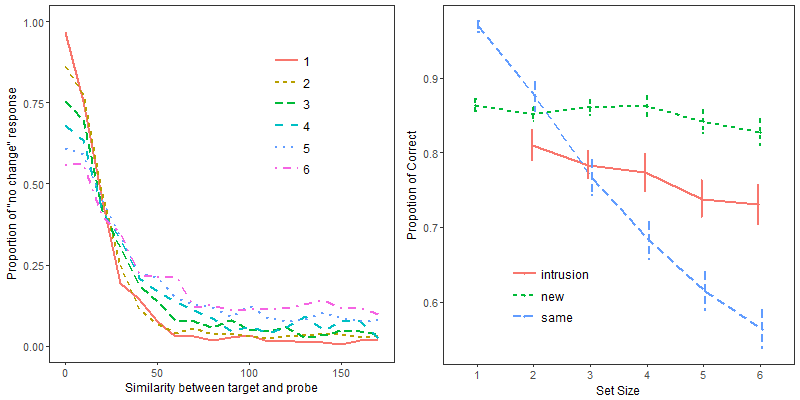


Figure 2. The proportion of "same" responses as a function of similarity of the probe to the target color, and the proportion of correct responses for the three probe types as a function of set size. Data are collapsed over Experiment A and B. The error bars indicate one standard error for within-subjects comparison.

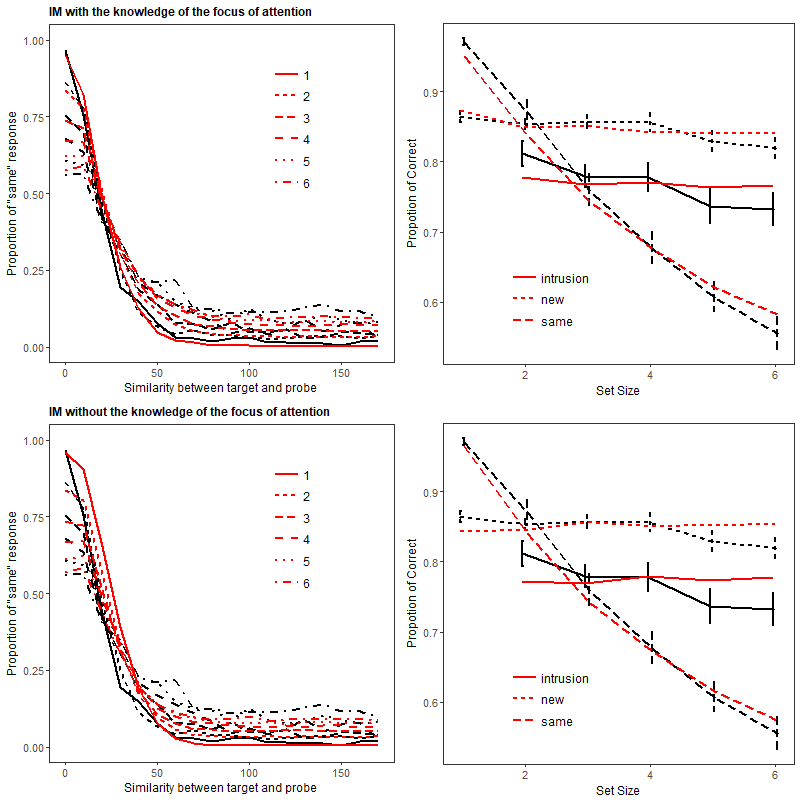


Figure 3. The model fitting from IM with different knowledge attributed to the inference rule. The black lines are the data, and the red lines are the model predictions from IM.

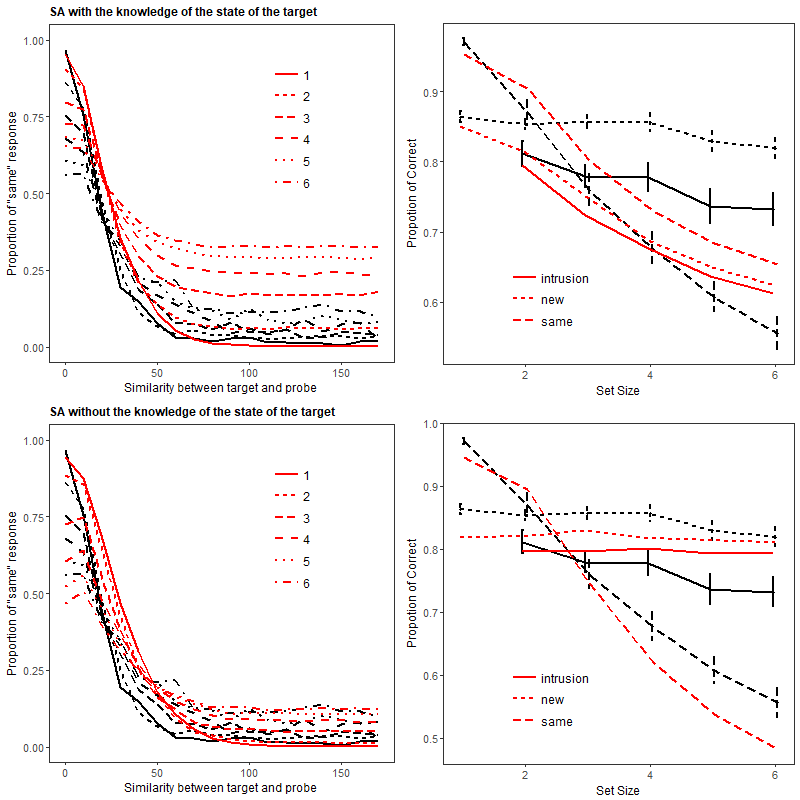


Figure 4. The model fitting from SA with different knowledge attributed to the inference rule. The black lines are the data, and the red lines are the model predictions from SA.

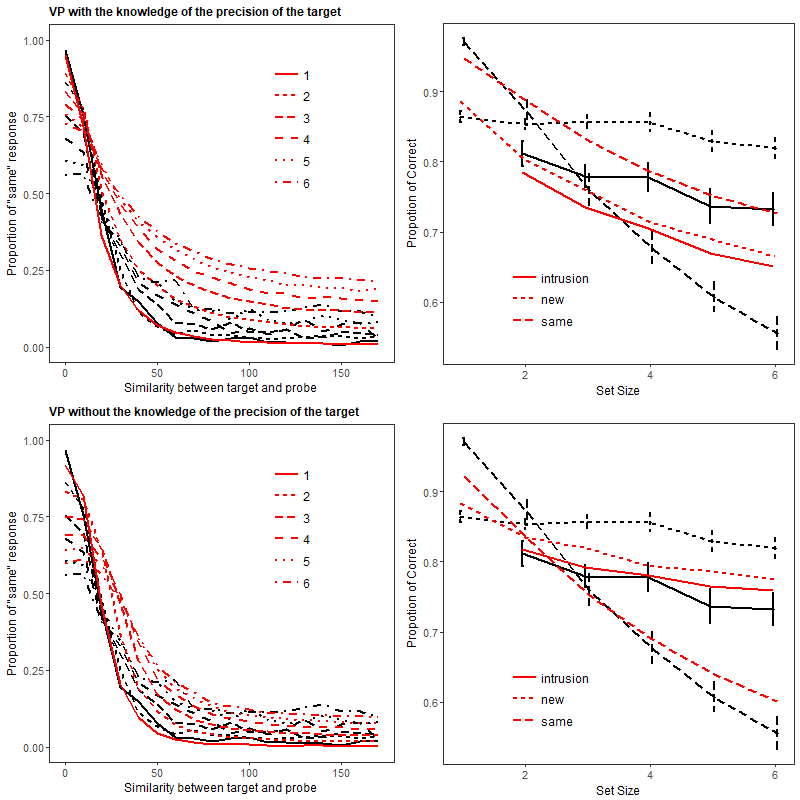


Figure 5. The model fitting from VP with different knowledge attributed to the inference rule. The black lines are the data, and the red lines are the model predictions from VP.

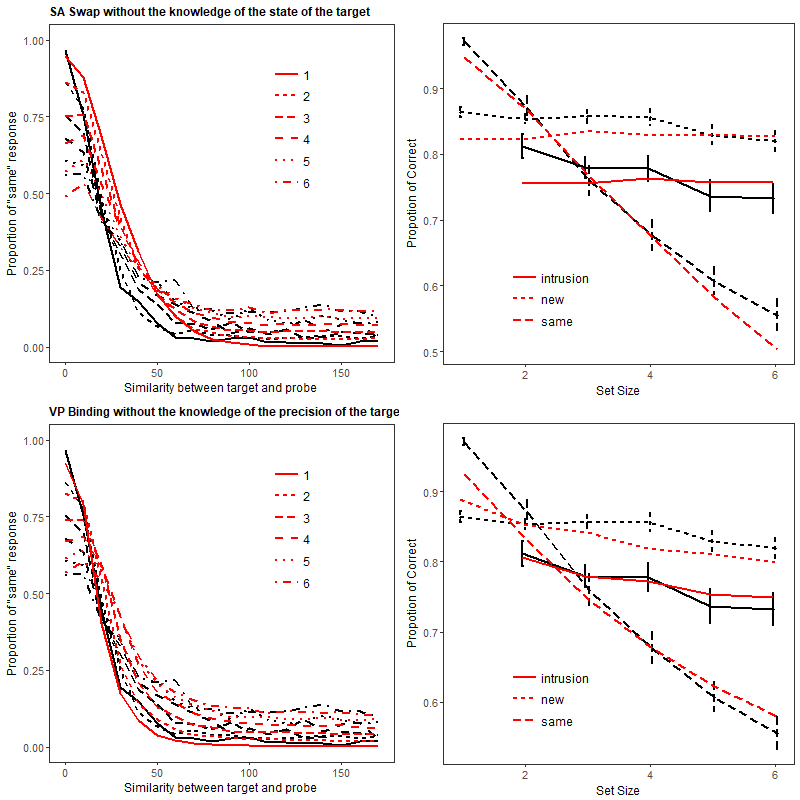


Figure 6 The model fitting results from SA-Swap, and VP-Binding. The top row is the response distribution and proportion of correct of the data (in black) and the model prediction (in red) from SA-Swap. The bottom row is the fitting result from the VP-Binding model.

# Appendix A

In the Bayesian inference rule, the is determined by both and . However, does not influence the expected outcome of the decision, as expressed in . In this Appendix, we explain the why drops out of the decision rule in the end.

According to Equation 8, the probability of change response, , is determined by the integral across of . We can interpret Equation 8 as the sum of the probabilities of recalling any color which is sufficiently similar to the probe color to decide *same*, and how similar the recalled color has to be for a *same* response is determined by . If the recalled color is similar enough to the probe color, which results in , a *same* response is given, and otherwise a *change* response is given. Thus, the minimum similarity for a recalled color to be given a *same* response is determined by . Because is the log-likelihood ratio between and , the exponential of is the likelihood ratio:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Setting = 0 implies , so we obtain:

|  |  |  |
| --- | --- | --- |
|  |  | (A.1) |

Thus, when simply calculating the minimum similarly for *same* responses, is no longer a factor in the equation: The minimum similarity of a probe to the target for a same response is the angular distance at which the von-Mises density (with precision ) crosses the density of the uniform distribution. Hence, we left out the when applying the inference rule to all the models.

# Appendix B

We tested three core models and their variants in the single-probe change-detection task. In this Appendix, we explain those models in detail and how the Bayesian inference rule was applied.

## Slot-Averaging model

The assumption of the SA model is that participants can only remember the number of items up to the capacity of working memory, . Thus, the probability of the target being stored in memory is

|  |  |  |
| --- | --- | --- |
|  |  | (B.1) |

the stands for set size of the current trial, and the is the probability of the target being in memory. If the target is in memory, the retrieved feature is distributed around the target feature. If the target is not in memory, the participant would have to guess, and the retrieved feature will be uniformly distributed among all the possible responses.

Assuming the target item is stored in a slot, the precision of retrieval is depended on how many slots are contributing to the retrieval. If each slot holds a unique item in the trial, which happens when , only one slot will contribute to the memory retrieval, and the precision of retrieval is the free parameter . However, when multiple slots hold the same item at once, which happens when , multiple slots will be sampled during the retrieval process, which results in higher precision of retrieval. The precision of retrieval is determined by how many slots are contributing during the retrieval process. We followed the equations used in Zhang & Luck (2008). When an item is stored in slots, the standard deviation of the retrieval distribution, , is

|  |  |  |
| --- | --- | --- |
|  |  | (B.2) |

and the is the standard deviation when the item is stored in one slot (which is a function of). Once the standard deviation of the retrieval is determined, we then convert it to the precision of retrieval by using the *sd2k* function provided by Bays et al. (2009).

For Slot-Averaging Swap, we assumed there is a certain probability that a swap error could occur during encoding, and the probability of a swap error increases linearly with set sizes. The probability that a swap error occurs between the target and any other item is at set size 2, which is the lowest possible set size for a swap error to occur. Thus, the probability of retrieving the target in SA-Swap is

|  |  |  |
| --- | --- | --- |
|  |  | (B.3) |

where is the probability of a swap error occurring at set size 2. When a swap error occurred, we assumed that all the non-target items have equal probability of being remembered instead of the target. Therefore, at retrieval, each individual non-target item has probability of being retrieved, regardless of set size.

To implement the inference rule for the SA and the SA-Binding model, we assumed that the inference rule has knowledge of the precision of the current trial, i.e., the value computed from . Depending on the level of knowledge, the inference rule either has knowledge of whether the target is in memory or not, or does not have knowledge regarding the memory state of the target. If the inference rule has knowledge of the memory state of the target, the probability of responding *change* is

|  |  |  |
| --- | --- | --- |
|  |  | (B.4) |

If the target item is known not to be in memory (which happened with probability ), the inference rule has to guess, and the optimal guessing is responding change 50% of the time. However, if the target is known to be in memory, the response is determined based on the retrieved memory and the probe with Equation B.4, with the probability of retrieving each possible feature given by a von Mises distribution centered on the true target feature , and precision = .

If the inference rule has no knowledge of the memory state, the probability of a "change" response is

|  |  |  |
| --- | --- | --- |
|  |  | (B.5) |

## with

## Variable Precision model

The assumptions in the VP model is that the total amount of resources is fixed, and the amount of resource put into an item varies from trial to trial. The mean precision at set size is

|  |  |  |
| --- | --- | --- |
|  |  | (B.6) |

is the mean Fisher information at set size 1, which determines the precision , and the parameter controls the steepness of decline in Fisher information as the set size increases. The relationship between and is . There is no analytic solution to compute directly from , thus we computed multiple s from a fine grid of s and interpolate the at given . The Fisher information *Ji* at each trial is drawn from a Gamma distribution with the mean of and scale of . For computational reason, instead of integrating across the Gamma distribution, we draw the *Ji* values from the 1000-quantiles of the Gamma distribution to simulate the integration.

The VP-binding model incorporated the assumption that participants used the location to retrieve the correct item, and the resource does not only affect the precision of the retrieved feature but also the precision of location information of all array items, and thereby the probability of retrieving the correct item, given the target's location. We assumed that the resources allocated to the binding between an item and its location is tied to the resources allocated to the memory item. The Fisher information of binding is the Fisher information of the memory item, multiplied with a scaling parameter , and the precision of binding is derived from . The probability of retrieving the th item is defined as

|  |  |  |
| --- | --- | --- |
|  |  | (B.7) |

The is the location of the probe.

For implementing the VP with the Bayesian Inference rule, is easy to determine in the VP and VP-binding model. If the inference rule has the knowledge of the precision of the current trial, the drawn from the Gamma distribution on is used in the inference rule. If the inference rule only has the knowledge of the average precision of the current set size, the average computed from is used instead.

## Interference Model

The assumption of the IM is that there are three sources of activation contributing to the retrieval process. Activation AC is the context-dependent information, which arises from using the location to re-activate the item bound to it,

|  |  |  |
| --- | --- | --- |
|  |  | (B.8) |

The is the color of item , and is the distance between the location of item and the location of the probe . The parameter is the rate of decline of the spatial generalization gradient of the probe location as a retrieval cue, that is, the precision of spatial-location memory.   
Activation Aa is the content-independent information, which arises from recent experience with every array item.

|  |  |  |
| --- | --- | --- |
|  |  | (B.9) |

Activation Ab is the background noise activation. Every response candidate *x* shares the same background-noise activation, and the amount of activation has a linear relationship with set sizes, as

|  |  |  |
| --- | --- | --- |
|  |  | (B.10) |

The activation of a response candidate is the sum of the three activations, with weights a, b, and c for each source of activation, as

|  |  |  |
| --- | --- | --- |
|  | , | (B.11) |

and we used the Luce's choice rule to determine the probability of retrieving based on the through

|  |  |  |
| --- | --- | --- |
|  |  | (B.12) |

where *N* is the number of possible response candidates (i.e., *N*=360 for the 360 colors on the color wheel).

We also assumed that the focus of attention is playing a role in the VWM. The focus of attention is able to maintain one item with higher precision, and that item is resistant to interference from the other items in the trial. However, with simultaneous presentation of the memory items, we cannot know which item is in the focus of attention. Therefore, we assumed that all the items have equal probability of being in the focus of attention. When the target is in the focus of attention, the feature precision is used instead of the usual , thus the activation Ac is

|  |  |  |
| --- | --- | --- |
|  | , | (B.13) |

The activation of the response candidate becomes

|  |  |  |
| --- | --- | --- |
|  | , | (B.14) |

where is a free parameter ranging from 0 to 1, reflecting the proportional reduction of the interference from Aa and Ab activations.

To implement the inference rule in the IM, we assumed that the used in the inference rule for the IM varies depending on the level of knowledge of the inference rule. If the inference rule does not have knowledge of whether the target item is in the focus of attention or not, the in the inference rule is the same as the in the IM without the focus of attention. However, if knowledge of whether the target is in the focus of attention or not is involved in the inference rule, we assumed that is used when the target item is in the focus of attention. Otherwise, the normal is used.

1. Specifically, we chose those models for which a likelihood is relatively easy to compute so we could fit them to data using a gradient-descent algorithm [↑](#footnote-ref-1)
2. It is important to note that even when a *change* probe is very similar to the target, the *change* probe was not relabeled into a *same* probe. We also tried relabeling the *change* probes within the similarity-region around the target into *same* probes, and the conclusion drawn from Bayesian statistic did not differ. [↑](#footnote-ref-2)