Cognitive Psychology 85 (2016) 30–42

Contents lists available at ScienceDirect

Cognitive Psychology

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

Resources masquerading as slots: Flexible allocation of visual working memory ☆

Chris Donkin *, Arthur Kary, Fatima Tahir, Robert Taylor

University of New South Wales, Australia

ARTICLE INFO

Article history:
Accepted 5 January 2016

Keywords:
Visual working memory
Mathematical modeling
Hierarchical Bayesian models
Change detection

ABSTRACT

Whether the capacity of visual working memory is better characterized by an item-based or a resource-based account continues to be keenly debated. Here, we propose that visual working memory is a flexible resource that is sometimes deployed in a slot-like manner. We develop a computational model that can either encode all items in a memory set, or encode only a subset of those items. A fixed-capacity mnemonic resource is divided among the items in memory. When fewer items are encoded, they are each remembered with higher fidelity, but at the cost of having to rely on an explicit guessing process when probed about an item that is not in memory. We use the new model to test the prediction that participants will more often encode the entire set of items when the demands on memory are predictable.

1. Introduction

Whether the capacity of visual working memory is better characterized as a limited number of discrete slots or as a continuous resource is a source of ongoing debate (Luck & Vogel, 2013; Ma, Husain, & Bays, 2014). According to the slots view, a limited number of items are stored with high precision (Luck & Vogel, 1997). Resource-based accounts rather propose that memory can be allocated more

☆ Chris Donkin's contribution to this research was supported by the Australian Research Council (DP130100124; DE130100129).

* Corresponding author at: University of New South Wales, Psychology, Matthews Building, Kensington, New South Wales 2052, Australia.

E-mail address: christopher.donkin@gmail.com (C. Donkin).

http://dx.doi.org/10.1016/j.cogpsych.2016.01.002

flexibly among items, with no necessary constraint on the number of items that can be held in memory (Wilken & Ma, 2004).

The change-detection task has long been used to understand the capacity of visual working memory (Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974). In a change detection task, observers are presented with an array of items to remember followed by a test array given a short period later. Observers are asked to determine whether the study and test arrays are the same, or if they have changed. In their seminal paper, Luck and Vogel (1997) showed that performance suffered a steep drop-off in performance once participants had to remember more than about 3–4 items. A slots-based account of this data proposes that the capacity of memory is around 3–4 items, and performance worsens once this capacity is exceeded (Awh, Barton, & Vogel, 2007; Barton, Ester, & Awh, 2009; Vogel, Woodman, & Luck, 2001).

Rouder et al. (2008) found explicit evidence for a slots model account of choice probability data in change detection tasks. In their study, the proportion of correct change responses (hits) and incorrect change responses (false alarms) were plotted using receiver operating characteristic (ROC) curves. Rouder et al. (2008) proposed a high-threshold version of a slots model, which assumed that items in memory were stored with enough precision that change/same discrimination for remembered items is perfect. Such a model has been long known to predict straight-line ROC curves (Green & Swets, 1966). Human performance in their experiment was well captured by this slots model, which outperformed a signal detection model representing the class of resource-based models.

Donkin, Tran, and Nosofsky (2014) replicated the results from Rouder et al. (2008). They also showed that the slots model was preferred in an experiment featuring a more stringent bias manipulation, providing a stronger test of the slots model's predictions. Donkin, Nosofsky, Gold, and Shiffrin (2013) also developed slots and resource models that account for both choice proportion and response time data, and showed that data from the Rouder et al. (2008) paradigm are better characterized by slots models.

However, some of the data from change-detection tasks have been shown to be more consistent with a resource-based account of capacity. For example, Keshvari, Van den Berg, and Ma (2013) showed that resource models outperformed slots models in a change-detection task. Further, Van Den Berg, Ma, and colleagues have repeatedly demonstrated that a particular type of resource-based model, one that assumes a flexible and stochastic allocation of memory, consistently outperforms traditional slot-based models in other visual working memory paradigms (e.g., Van den Berg, Awh, & Ma, 2014; Van den Berg, Shin, Chou, George, & Ma, 2012).

The literature on the capacity of visual working memory is inconsistent, with evidence that supports both slot-like and resource-like capacity. Our aim here is to provide a potential resolution to why we see evidence for both slots and resource models. We begin by noting that although resource models can mimic slots models (by allocating memory to only some of the items in the display), slots models are unable to mimic resource models.[1] As such, one is forced to conclude that visual working memory is a flexible resource. However, we propose that the flexible resource of working memory is allocated in a slot-like fashion in certain environments.

We now pursue our conjecture that in certain environments, the resource-based capacity of memory is allocated such that it appears to be slots-based. We first describe the series of previous experiments that led us to this proposition, and outline a pair of new experiments that test our prediction. We then introduce a new flexible-resource computational model that incorporates both slot-like and resource-like encoding of items; either dividing its memory among all of the items in a display, or encoding a smaller subset of items (but with higher resolution). We show that under different environments, participants do indeed show more or less slot-like encoding of stimuli.

1.1. Old experiments: why we think that certain environments lead to slot-like encoding

Donkin et al. (2014) replicated and extended the results from Rouder et al. (2008) across a series of four experiments. All experiments used the change detection task, but differed in the exact manipulation of two primary independent variables; set size and change proportion. As set size – the number

[1] Note that the slots + averaging model in Zhang and Luck (2008) can mimic resource models when the number of items to remember is smaller than capacity, because a single item can be stored in multiple slots, thus improving its resolution in memory.

Table 1
A survey of the design of the Old Experiments in Donkin et al. (2014), and the New Experiments we present here (indicated by *). The best-fitting model from a comparison between slots and resource models is also reported.

Experiment	Set size		Change proportion	Model
	N	Varies		
Original	3, 5, 8	Trials	0.3, 0.5, 0.7	Slots
Extended	2, 5, 8	Trials	0.14, 0.3, 0.5, 0.7, 0.86	Slots
Set Size Only	1, 2, 3, 4, 6, 8	Trials	0.5	Slots
Change Proportion Only	6	N/A	0.14, 0.3, 0.5, 0.7, 0.86	Resources
Blocked Set Size*	2, 4, 6, 8	Sessions	0.14, 0.3, 0.5, 0.7, 0.86	Resources
Unblocked Set Size*	2, 4, 6, 8	Trials	0.14, 0.3, 0.5, 0.7, 0.86	Slots

of items to remember – increases, performance worsens as hit rates decrease and false alarms increase. The change proportion manipulation involves altering the proportion of trials on which an item changes from study to test. As the proportion of change trials increases, participants tend to increase the probability of responding change, and so both hit rates and false alarms increase. Manipulating both set size and change proportion yields a distribution of hit and false alarm rates that covers the ROC space, and allows for contrast between the slots and resource models of visual working memory capacity.

Table 1 details the manipulations involved in each of the experiments conducted by Donkin et al. (2014). The table also reports the model that best fit those data. The critical result is that the slots model provides a better account of the data than the resources model for all but one experiment.

The experiment for which the resources model gave the best fit had only one set size condition, such that participants were presented with 6 items to remember on all trials. We conjecture that when set size was constant, and thus predictable, participants were more able to allocate their memory more flexibly among all of the items in the display. That is, because participants knew how many items they were going to have to remember, they were better able to encode a smaller amount of information about all of the items in the display, and thus appear to have a more resource-like capacity. In all other experiments in Table 1, the number of items to remember changed from trial-to-trial, making it impossible to predict the upcoming demands on memory. In such environments, we propose that observers encoded a smaller number of items, but with high precision, and thus appeared to have a slot-like capacity.

2. New experiments

We ran two experiments to test whether observers distribute their memory over all items when they can predict their upcoming memory load. In both experiments we varied the number of items to remember. In Experiment 1, the memory set size was held constant for the entire one-hour session, but varied across sessions. In Experiment 2, the memory set size varied from trial-to-trial, as per Rouder et al. (2008) and our previous experiments. We expect that participants in Experiment 1 will divide their memory among all of the items they are given to remember, since they know the number of items they will have to remember. On the other hand, we expect that because participants in Experiment 2 are unable to predict the number of items they will have to remember, they will tend to encode only a smaller subset of items.

2.1. Participants

In Experiment 1, 33 participants completed four one-hour sessions of a change-detection experiment. Participants were recruited from a paid-participant register at UNSW, Australia, and were reimbursed $15 per session. Fifteen new participants were recruited for Experiment 2.

2.2. Stimuli

The stimuli used in both experiments were identical to those used in Donkin et al. (2014), and were intended to replicate Rouder et al. (2008). The stimuli were a set of 10 highly distinct color squares (white, black, red, blue, green, yellow, orange, cyan, purple, and dark blue-green) presented on a gray background. We used 24-in. LCD monitors. Each color square was 0.75° by 0.75° in size. Stimuli were presented within an imaginary rectangle of size 9.8° by 7.3°. Items were presented in random locations, with the constraint that they could not be within 2°s of another stimulus, or from the center of the display. A cue indicating the position of the target stimulus was a black, 1 pixel thick, 1.5° diameter circle that surrounded the target color square.

2.3. Design

The number of items in a study array was manipulated to be either $N = 2, 4, 6$, or 8 items. In Experiment 1, the set size, N, was constant for any given one-hour session, but varied across the four sessions. The order of presentation of the levels of N was randomized across participants. In Experiment 2, N varied from trial-to-trial, with the constraint that each set size was presented equally often in a given block. The proportion of trials on which an item changed from study to test was also varied across blocks of trials. In a given block of trials, the proportion of change trials was either 0.14, 0.3, 0.5, 0.7, or 0.86. The order of change proportion conditions was random, with the constraint that all change proportion conditions were presented before any one condition would repeat. Participants completed each change proportion condition twice per one-hour session. Finally, the order of change and same trials was random within a given block.

2.4. Procedure

A fixation cross was presented for 1000 ms at the start of each trial. A study array of N color squares was then presented for 500 ms. The stimuli were then removed and the screen remained blank for 500 ms, and a multicolored mask was presented for 500 ms at each location in which a study item had been presented. A single test color was then presented in one of the study locations, and was either the same as the item previously presented in that location (a *same* trial), or was an item not previously presented in the study array (a *change* trial). The participant was asked to indicate whether the test item was the same or had changed by pressing either the "F" or "J" keys on a standard keyboard, respectively. The participants then received feedback on the accuracy of their response for 1000 ms, and the next trial began after a 1000 ms blank screen.

At the start of each block of trials, participants were explicitly informed of the proportion of change trials in the upcoming block. In Experiment 1, participants were told at the start of each session how many stimuli they would be presented with on all trials. Each session also began with 6 practice trials with the given set size. In Experiment 2, participants were told that the number of items they had to remember would vary from trial-to-trial. Participants completed 10 blocks of 60 trials in each session, yielding between 18 and 102 trials per set size per change proportion condition for each participant (note that the range in sample size occurs because of the change proportion manipulation).

3. Computational model

We now outline a flexible-resource model that allows us to measure the extent to which the experimental design influences the encoding strategy of the participant. Though our model assumes a flexible allocation of memory, it differs from existing variable-precision models (Van den Berg et al., 2012), but is more similar to the model proposed by Bays, Catalao, and Husain (2008). In variable-precision models, the precision with which a given item is encoded is assumed to be a stochastic process. Here, instead of assuming that the allocation of memory is random, we allow for two ways in which a participant may allocate their memory. As outlined in Fig. 1, we assume that participants either allocate memory to all items they are asked to remember, or to only a subset of those items. When the model encodes only a subset of items, then performance for the remembered items will

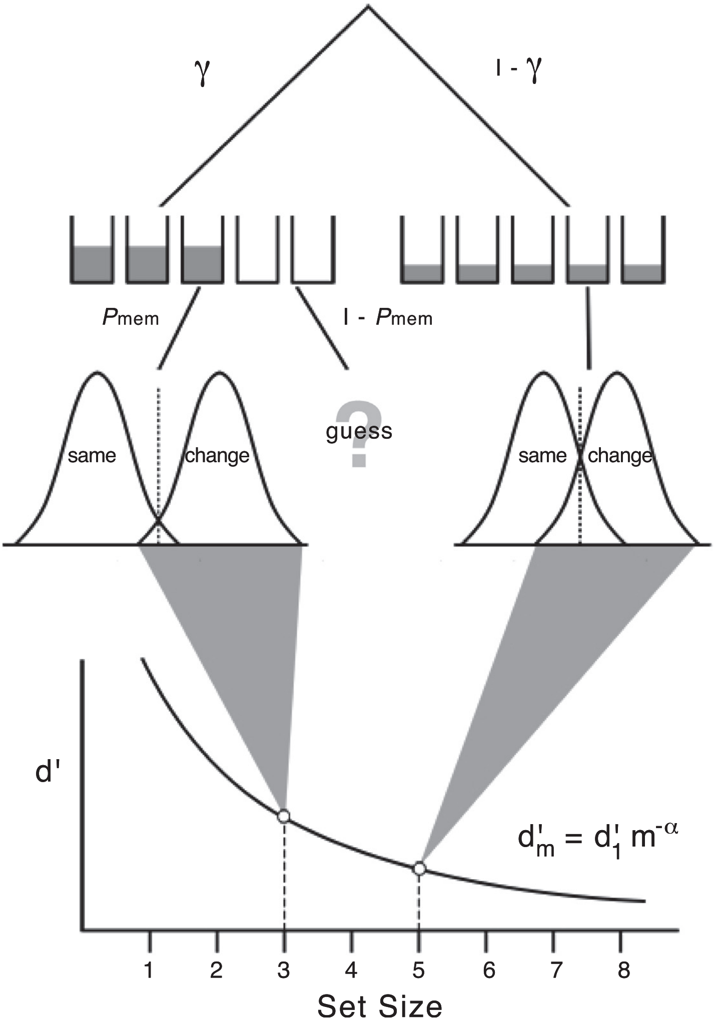

Fig. 1. A graphical depiction of our flexible-resource model. With probability γ, only a subset of items are encoded, or else all items are encoded. If slot-like encoding is used, then with probability p_{mem}, an item that is in memory is probed and the change-detection decision is made with high discriminability, or else the item probed is not in memory and the participant guesses. If memory is given to all items, then the change-detection decision is made with worse sensitivity, but there is no guessing process.

be good, but the observer will be forced to guess when probed with a test item that is not in memory. On the other hand, when all items are encoded the model never needs to guess, but has a less precise representation for each item. Further, for simplicity, we assume that memory is divided evenly amongst the stimuli. We assume that observers are capable of moving between these two possible encoding strategies from trial-to-trial, and our focus in what follows will be the probability that the observer uses a slot-like encoding strategy during each experiment.

We use a signal-detection framework to model the way in which the observer makes their change-detection decision, regardless of their encoding strategy. We assume that when a test item is presented in a particular location, it will evoke a particular level of familiarity, x. The level of familiarity is then compared to a criterion, which the observer uses to decide whether the test item and the study item are the same, or if the test item has changed from study.

In the signal-detection process, the familiarity evoked by a test item is assumed to vary from trial-to-trial. The exact sources of this variability is unknown, but we assume that it arises out of two process. The first is a noisy decision-making process (cf. Rae, Heathcote, Donkin, Averell, & Brown, 2014). The second is that the amount of mnemonic resource given to any one item on a particular trial is a stochastic process. Note that unlike Van den Berg et al. (2012), we do not explicitly model the stochastic allocation of memory to items. Rather, we assume that the individual will divide their memory evenly across encoded items, on average. Any additional variability in memory encoding is assumed to be contained within the signal-detection process.[2]

We use a standard normal distribution to represent the distribution of familiarity evoked by a test item that is the same as the study item presented earlier. When the test item has changed from study, then the familiarity follows a normal distribution with a mean of d' and a variance of 1. As such, when a test item evokes an amount of familiarity, x, the participant compares the likelihood of that familiarity under those two distributions: $LR(x) = \frac{\phi(x-d')}{\phi(x)}$ where ϕ is the probability density function of the standard normal distribution. If the likelihood ratio is greater than criterion β, then the observer responds "change" and otherwise responds "same". The predicted hit rate under such conditions is given by $\Phi\left(\frac{d'}{2} - \frac{\log\beta}{d'}\right)$ and the predicted false alarm rate is $\Phi\left(-\frac{d'}{2} - \frac{\log\beta}{d'}\right)$, where Φ is the cumulative density function of the standard normal distribution.

We assume that observers have a fixed amount of mnemonic resource. As the number of items in memory increases, any one item will be given less memory, and as such it will become more difficult to determine whether that item has changed or is the same. As such, d' will decrease as the number of items in memory increases. We assume that d' decreases as a power-function with the number of items in memory, m,

$$d'_m = d'_1 m^{-\alpha} \tag{1}$$

where α and d'_1 are free parameters. The power-law assumption is common in resource-based models of memory (Bays, Catalao, & Husain, 2009; Donkin & Nosofsky, 2012; Van den Berg et al., 2012). It is worth noting that the results we present are qualitatively identical when we assume a sample-size model for the relationship between d' and the number of items in memory (i.e., $d'_m = \frac{d'_1}{\sqrt{m}}$), as in Smith and Sewell (2013), Sewell, Lilburn, and Smith (2014), and Palmer (1990).

The model allows for two encoding strategies: Either all of the items in the display are allocated memory, or only a subset of those items are encoded. To model this, we assume that there is some probability, γ, that the observer uses a slot-like encoding strategy on a given trial, or else uses a resource-like encoding strategy. When using a slot-like encoding strategy, the observer encodes only a subset of k items from the set of N items presented. If capacity, k, exceeds the number of items to be remembered, N, then all items enter memory. As such, the probability that an item in memory is the test item is given by $p_{mem} = \min\left(1, \frac{k}{N}\right)$. When using a resource-like encoding strategy, the observer allocates memory to all items in the study array.

We assume that the precision with which an item is encoded into memory is based on the number of items in memory. Note that this is different from a high-threshold slots model, which assumes that participants will respond with perfect accuracy when an item is in memory (cf. Donkin et al., 2014; Rouder et al., 2008).[3] Instead, the encoding strategy used by the individual simply determines the number of items that are in memory. When participants encode only a subset of items, then the mnemonic resource must be divided among fewer items, but at the cost of having to guess on some proportion of trials. On the other hand, when encoding all items, participants will never have to guess, but will have spread their mnemonic resource across a larger number of items.

[2] If we were to explicitly model the variability in allocation of memory to items, then we would no longer have a closed-form expression of the model's likelihood. Such an extension is likely to be useful, but would make it almost impossible to implement the model into a hierarchical Bayesian framework.

[3] Note that our model, unlike the high-threshold slots model of Rouder et al. (2008), only predicts a straight-line ROC curve when a participant uses a slot-based encoding strategy on every trial, and encodes each item with a ceiling-level d'.

When using slot-like encoding, the test item is either in memory or not in memory. When the test item is in memory the d' for that item is given by Eq. (1), setting m to be $m_S = \min(N, k)$. When the test item is not contained in memory, then the participant guesses and responds "change" with probability g. When using a resource-like encoding strategy, then the participant encodes the full set of N items, and so $m_R = N$ in Eq. (1).

We assume that the probability of guessing "change", g, depends on the proportion of trials on which the test item changes from study. We expect that the observer will guess "change" more often when there are more change trials. Similarly, we expect that the criterion amount of familiarity required to make a "change" response, β, will depend on the proportion of change trials in a given block. Since we do not have any expectation about the functional form relating the g and β parameters to the level of change proportion condition, these parameters are simply estimated from data.

Taken together, the predicted hit rate in the ith set size condition (with N items), and the jth change proportion condition, θ^H_{ij}, is given by

$$\theta^H_{ij} = \gamma\left[p_{mem}\Phi\left(\frac{d'_1 m_S^{-\alpha}}{2} - \frac{\log\beta_j}{d'_1 m_S^{-\alpha}}\right) + (1-p_{mem})g_j\right] + (1-\gamma)\Phi\left(\frac{d'_1 m_R^{-\alpha}}{2} - \frac{\log\beta_j}{d'_1 m_R^{-\alpha}}\right)$$

and the predicted false alarm rate, θ^F_{ij}, is

$$\theta^F_{ij} = \gamma\left[p_{mem}\Phi\left(-\frac{d'_1 m_S^{-\alpha}}{2} - \frac{\log\beta_j}{d'_1 m_S^{-\alpha}}\right) + (1-p_{mem})g_j\right] + (1-\gamma)\Phi\left(-\frac{d'_1 m_R^{-\alpha}}{2} - \frac{\log\beta_j}{d'_1 m_R^{-\alpha}}\right)$$

where $p_{mem} = \min\left(1, \frac{k}{N}\right)$, $m_S = \min(N, k)$, and $m_R = N$. The free parameters in this model are γ, k, d'_1, α, g_j, and β_j.

For the ith set size and the jth change proportion condition, the lth individual in the mth experiment produces h_{ijlm} hit responses out of n^c_{ijlm} change trials, and f_{ijlm} false alarm responses out of n^s_{ijlm} same trials. We use a Binomial model to characterize the generation of the hit and false alarm responses, such that $h_{ijlm} \sim Binomial\left(\theta^H_{ijlm}, n^c_{ijlm}\right)$ and $f_{ijlm} \sim Binomial\left(\theta^F_{ijlm}, n^S_{ijlm}\right)$.

We implement this model in a hierarchical Bayesian framework, assuming that the parameters of each individual are drawn from population-level distributions. For example, we assume that the capacity k of the lth individual in the mth experiment comes from a Normal distribution, with $k_{lm} \sim N(\mu^k, \sigma^k)$.[4] For all parameters, except the probability of using a slot-like encoding strategy, γ, we assume that the population-level distribution is identical across experiments. In other words, we assume that an individual's capacity, their mnemonic resource, their guessing behavior, and their criteria settings *do not vary systematically across experiments*. That is, we allow only the encoding strategy to vary systematically across experiments, such that $\gamma_{lm} \sim N(\mu^{\gamma_m}, \sigma^{\gamma})$.[5]

Since we are interested in posterior distributions, and not in model selection, we placed vague priors on all parameters. In particular, for the means of the population-level distributions we set $\mu^{d'_1} \sim U(0, 10)$, $\mu^{\alpha} \sim U(0.01, 1)$, $\mu^k \sim U(0, 8)$, $\mu^{g_{0.15}} \sim U(0, 1)$, $\mu^{\beta_{0.15}} \sim U(1, 5)$, and $\mu^{\gamma_m} \sim U(0, 1)$. We assumed that the guessing parameters followed an order constraint. For example, we assumed that the average probability guessing "change" in the 30% change proportion condition was larger than in the 15% condition: $\mu^{g_{0.3}} \sim U(\mu^{g_{0.15}}, 1)$. We made the equivalent assumption for the criteria parameters (e.g., $\mu^{\beta_{0.3}} \sim U(0.001, \mu^{\beta_{0.15}})$). All population-level standard deviation parameters were given the same prior: all $\sigma \sim U(0.01, 10)$.

[4] We allow capacity to be a continuous value, despite it being a discrete quantity. We make this assumption as a crude approximation for trial-to-trial variability in the number of items encoded. For example, a value of $k = 3.5$ is interpreted as encoding 3 or 4 items on an equal proportion of trials.

[5] We also fit a version of the model in which k, d', and α were free to vary across experiments, and found the same pattern of results in *gamma* parameters.

4. Results

We used the data from all four of the experiments reported in Donkin et al. (2014) and the two new experiments to estimate our model parameters. Together, the data from 238 participants inform these estimates. We used JAGS to obtain posterior distributions, running 24 chains of 55,000 samples, discarding 5000 samples to burn-in, and thinning every 100th sample (Plummer, 2003). In support of open science, the code to apply the model, all of our raw data, and the code and stimuli used in our experiments, is available on the first author's website.

Our analysis focuses on the way in which the probability of using a slot-like encoding strategy, γ, changes across the different experiments. We will use Bayes factors to quantify the reliability of the differences between μ^{γ} parameters. We evaluate Bayes factors using the Savage-Dickey procedure (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). Bayes factors quantify the likelihood that the observed data would occur if there is a difference between the two parameters relative to if there is no difference between the two parameters. We report Bayes factors such that positive values indicate that the alternative hypothesis is more likely than the null.

4.1. Old experiments

The left panel of Fig. 2 contains the posterior distributions of the probability of using slot-like encoding, μ^{γ_m}, for each of the four old experiments, and the two new experiments. Focusing first on the results of the old experiments, we see that the results are consistent with previous attempts to compare slots and resource based models of these data. For the Original experiment, we see that the probability of using slot-like encoding is close to 1. We see similar levels of slot-like encoding in the Set Size Only experiment, in which set size varied from trial-to-trial and change proportion was held constant. The Bayes factor for the difference between μ^{γ} for the Original and Set Size Only experiments, *BF*, is 0.05. This Bayes factor tells us that the null hypothesis of no difference between the parameters is roughly 20 times more likely than the alternative hypothesis in which there is a difference.

The probability of slot-like encoding looks to be somewhat lower for the Extended experiment. The design of the Extended experiment was very similar to the original experiment, but included more extreme change-proportion conditions. The difference between the Extended experiment and the

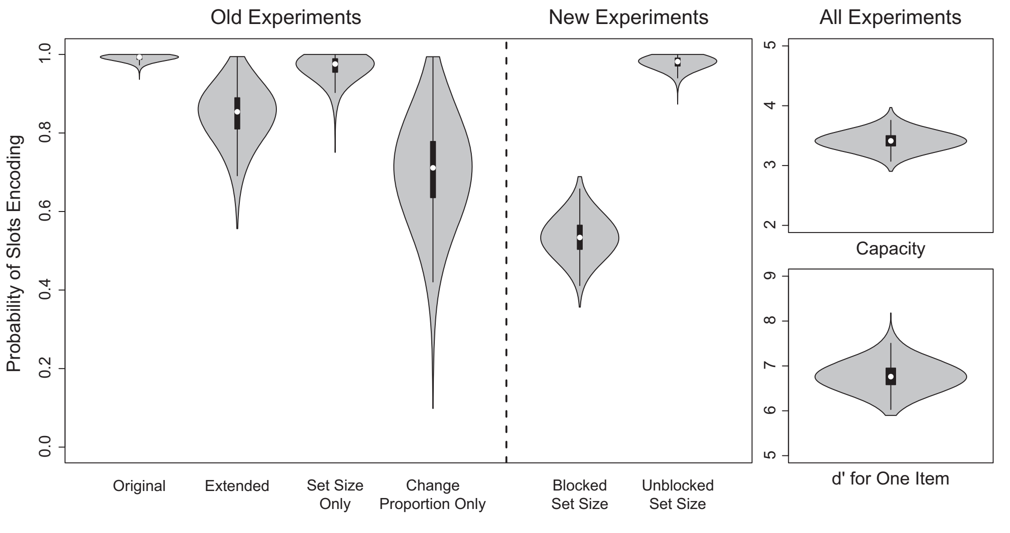

Fig. 2. Posterior distributions for the population-level parameters in the model. The left panel contains the likely values for the probability of using slot-like encoding in each of the experiments, μ^{γ_m}. The posterior distributions for the capacity parameter, μ^{k}, and the sensitivity for one item, $\mu^{d'_1}$, are plotted in the right panels.

Original experiment is reliable ($BF = 25$), though it is unclear whether there is a difference between the Extended and Set Size Only experiments ($BF = 0.8$).

Finally, we see that the probability of using slot-like encoding is smallest for the Change Proportion Only experiment. This is consistent with the conclusions in Donkin et al. (2014), where the resource models outperformed the slots models in standard model comparison processes. The probability of slot-like encoding in the Change Proportion Only experiment looks to be reliably smaller than that in the Original and Set Size Only experiments ($BF = 139$ and $BF = 10.7$, respectively), though it is unclear if there is a difference between Change Proportion Only and Extended experiments ($BF = 0.6$).

It is worth noting that we are much less certain about what values of μ^{γ} are likely for the Change Proportion Only experiment. We are uncertain about μ^{γ} in this experiment because the set size is held constant. We are presumably only able to make inferences about μ^{γ} at all because of the constraint on the model's parameters that come from other experiments. For example, because we know what values μ^{k} and $\mu^{d'_1}$ should take from other experiments.

4.2. New experiments

Turning now to the new experiments, it is clear that the probability of using slot-like encoding was different for the Blocked and Unblocked Set Size experiments ($BF > 9 \times 10^{7}$). Participants in the Blocked Set Size experiment seemed to only use slot-like encoding on about half of the trials, much less often than is typically observed in our other experiments. This probability of slot-like encoding is reliably smaller than the Original, Extended, and Set Size Only experiments ($BF = 1 \times 10^{10}$, $BF = 141$, and $BF_{10} = 2 \times 10^{5}$, respectively). However, it was not clear whether there was a difference between the Change Proportion Only condition and the Blocked Set Size experiment ($BF_{10} = 0.9$).

In contrast, participants in the Unblocked Set Size Experiment appear to behave much like the participants in the Original and Set Size Only experiments, using a slot-like encoding strategy on the vast majority of trials ($BF = 0.04$ and $BF = 0.06$, respectively). The value of μ^{γ} in the Unblocked Set Size experiments was maybe larger than that in the Extended experiment ($BF = 5$), and reliably larger than in the Change Proportion Only experiment ($BF = 34$).

4.3. Is the γ parameter interpretable?

The values of γ across experiments that we observe are consistent with our expectations. However, these conclusions assume that the model provides a reasonable account of the observed data. We have two sources of evidence for this claim: whether the interpretable parameters of the model are sensible, and whether the model gives a good fit to the data.

First of all, the parameters of the model look sensible. The right panel of Fig. 2 plots the posterior distribution for the population-level capacity parameter, μ^{k}. Across all experiments, we infer that the average capacity of individuals is between 3 and 4 items. Such an estimate of capacity is perfectly consistent with the standard estimate for this quantity (Cowan, 2001). We also see that the sensitivity when just one item is presented, $\mu^{d'_1}$, is large. A large value suggests that people would be extremely accurate if presented with just one item to remember, which seems likely given that we used a set of 10 highly-discriminable colors for our stimuli.

Second, and most importantly, Fig. 3 indicates that the model provides a good account of the data from all six experiments. The data from each experiment are plotted using ROC plots, with the hit and false alarm rates shown as a function of set size and change proportion conditions. To create the plot, we first fit a "data model"; a hierarchical Bayesian model that freely estimated the hit and false rates within all conditions for each individual in each experiment. The model assumed a Binomial data-generating process, and thus avoided the issues with treating count data as though it was normally distributed. Further, the model is hierarchical, and assumes that the rate parameters from individuals within a given experiment are distributed normally. We use a hierarchical model for this data because it automatically deals with atypical participants, shrinking their parameter estimates towards the rest of the group. Note that in the data model, all parameters were estimated independently for each experiment.

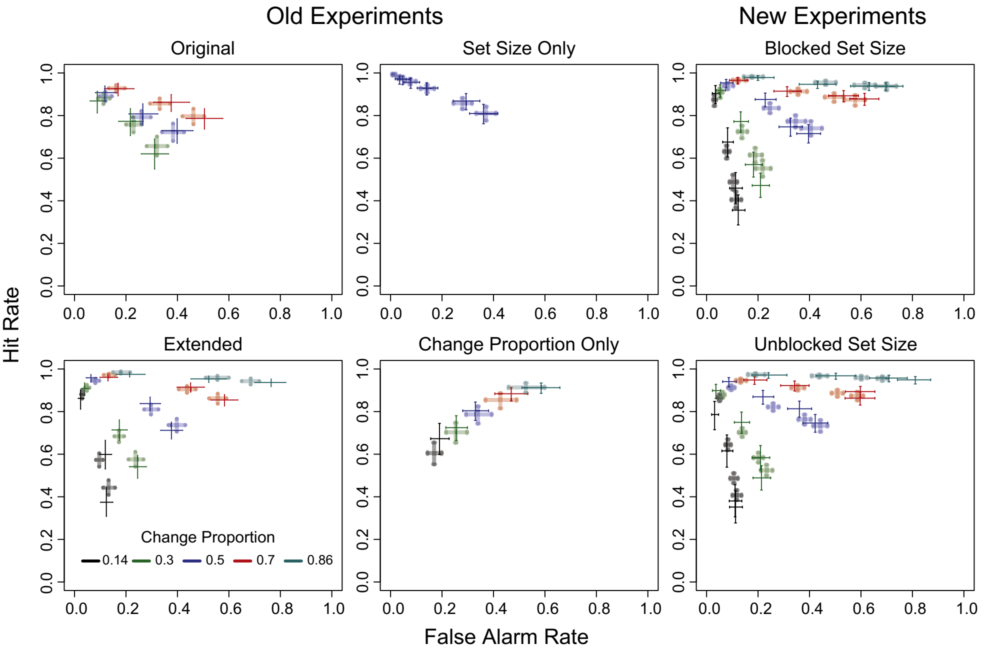

Fig. 3. The central 60% of the posterior distributions for the observed and predicted hit and false alarm rates from each of the four old experiments from Donkin et al. (2014), and the two new experiments. The posterior distributions were aggregated over individuals. The predictions come from the flexible resource model (thick lines). The posteriors for the observed data come from a "data model" (thin lines, see text for details).

We now have posterior distributions for the hit and false rates from a "data model", and posterior distributions for the predicted hit and false alarm rates from our cognitive, flexible-resource model. In Fig. 3, we plot the central 60% of the posterior distributions of the rate parameters from both the data model and the cognitive model, aggregating across the individuals within each experiment. The degree of overlap between the posterior distributions from the data and cognitive model reflects the agreement between the model and the observed data.

The cognitive model provides an impressive account of the data from the six experiments, especially considering that the population-level distribution of all but one parameter in the model, μ^{γ}, was assumed to be invariant across experiments. The only place in which the model appears to be systematically failing to account for the data is in the Unblocked Set Size experiment. This misfit is largely driven by atypical data. For example, there appear to be an equal number of false alarms in the set size 6 and 8 conditions for the 30, and 70% change proportion conditions. In general, the model seems to account for the data well enough that we can safely interpret the μ^{γ} parameters, though we return to a discussion of μ^{γ} for the Unblocked Set Size experiment in Section 5.

5. Discussion

The probability that memory was allocated to all of the to-be-remembered items was largest for experiments in which the memory set size was held constant for some time, and was thus predictable. Our conclusions came from a new computational model that flexibly allocates memory to study items. When we fit the model to a pair of new experiments, we saw a dramatic drop in the rate at which people used slot-like encoding when the number of items to remember was constant for an entire session, but varied across sessions. According to our model, when the number of items to remember did not change from trial-to-trial, people attempted to encode all of the stimuli on about 45–50% of trials.

On the other hand, when all aspects of the experiment were the same, except that the number of stimuli differed across trials, people appeared to almost never encode all of the items.

5.1. A mixture of encoding strategies?

We predicted that when set size was fixed, participants would encode all of the items in the display. Although the rate of slot-like encoding did decrease, it did not disappear entirely. Apparently, participants in the fixed set size experiment equally often used slots- and resource-like encoding strategies. We now speculate over why we may have observed such a mixture of strategies.

In short, we suspect that participants in the fixed set size condition are not always able to distribute their memory across all items in the display. Our fundamental assumption is that an attentional window allows for the encoding of items into visual working memory. When attention lingers over an item, or a subset of items in close proximity, then the features of those items are encoded into working memory. We also assume that more time spent encoding the features of an item will lead to a higher resolution representation of that item.

In the variable set size condition, observers may focus their attention on items such that they are encoded with high precision. Since high-quality encoding takes time, and study arrays are presented for a short time, relatively few items can be encoded. Empirically, it seems that in variable set size conditions, such a way of allocating memory leads to a rich representation of about 3–4 items, while any remaining items receive little to no representation. That is, for certain environments it appears that people allocate their flexible mnemonic resource in such a way as to appear that they have discrete slots.

When set size was fixed, we expected that observers would attempt to distribute their attention across all of the items in the display when the number of items to remember exceeds 3–4 items. When the number of stimuli to encode is predictable, the observer could learn to allocate attention such that, on average, information about all of the items could be encoded. Encoding information about all of the items requires that observers spend less time encoding the features of each item.

The mixture of resource- and slot-like encoding strategies we observe in the fixed set size condition is likely a failure to employ this strategy correctly on all trials. For one, it must take practice to learn how much time must be allocated to each item to divide one's resources adequately. Further, since it must take time to move attention, sometimes items will be arranged such that it is impossible to deploy attention to all items in a 500 ms presentation window. On such trials, observers will be probed about items for which they have no memory, and thus appear to use slot-like encoding.

5.2. Can people change their encoding strategy?

Our theory of encoding in visual working memory also may help understand the discrepancy between our results and those reported in Zhang and Luck (2011). Their key result was that participants were unable to increase the number of stimuli stored in memory by reducing the quality with which items were encoded. For example, participants in one experiment performed identically regardless of whether the task required a low or high precision representation to perform well. In Zhang and Luck's experiments, participants were shown the study display of four items for only 200 ms (one experiment used 6 stimuli). It is possible that such a short presentation does not permit participants to strategically allocate their attention to additional items, thus limiting participants ability to demonstrate a flexible deployment of memory.

Interestingly, Bengson and Luck (in press) recently showed that participants who were instructed to remember all items in a display outperformed those who were given neutral instructions or were asked to focus on a subset of items. Bengson and Luck's study was similar to our experiments, in that they used a change detection task, and also had participants remember up to 8 items, exceeding the number of items to be remembered in the Zhang and Luck (2011) study. It remains a possibility that larger set sizes are required to demonstrate a flexibility in visual working memory. Alternatively, it may be that change detection tasks are more sensitive to changes in capacity and resolution than the reproduction task used in Zhang and Luck, perhaps due to measurement issues or because participants are unwilling to encode with poor precision when they are required to produce stimuli. Clearly, further work is required to disentangle these possibilities.

If people are truly unable to demonstrate multiple encoding strategies within a given experiment, then we need an alternative explanation for our results. One possibility is that any given individual always uses just one of either a slot- or resource-based strategy, and that our results reflect different mixtures of individuals across experimental designs. For example, we can carry out a 'standard' model selection technique, using Bayes factors to identify individuals as being better fit by resource- or slot-based models. The results of this comparison mirror our primary conclusions: In the fixed set size experiment, half of the participants were better fit by resource models, while almost all participants were better fit by slot models in the variable set size experiment. We also implemented a hierarchical version of the individual-differences model to all six experiments, and found that it too yields the same general conclusion while also giving a good fit to the data from our experiments at the aggregate level. At present there exists no reliable means of selecting between hierarchical models of the kind we use here, though it is theoretically possible, and so we must leave it to future work to determine whether the mixture of encoding strategies occurs within or across individuals.

5.3. Extensions

Van den Berg, Ma, and colleagues provide considerable evidence that a resource of variable precision dictates the capacity of visual working memory (Ma et al., 2014; Van den Berg et al., 2014; Van den Berg et al., 2012). Such a model begs the question of whether the way in which memory is allocated to items will depend on the demands of the task. Our results suggest that encoding strategy does depend on the task; a result that itself raises a host of new questions. For example, we consider only two possible ways in which participants can allocate their memory to study items – by encoding all items, or encoding only a small subset. Future work could investigate whether more complex encoding strategies exist, which they almost certainly do. It seems likely that our model is a subset of the true range of encoding strategies that humans can perform. A more encompassing model could assume a full distribution over the number of items to be remembered, with the parameters of this higher-level distribution varying as a function of experimental design. Comparison between such a model, and existing variable-precision models, which lack such assumptions, would be an interesting avenue for future work.

There are, of course, a large number of ways in which participants may choose to encode stimuli. Some such strategies will require considerable extensions to our model. For example, Brady and Alvarez (2015) show that people can encode and use information about the global ensemble of stimuli to help with their change detection decisions. Similarly, participants are presumably able to chunk together visual stimuli, potentially increasing the efficiency with which information is stored in memory (Cowan, Rouder, Blume, & Saults, 2012). Whether flexible encoding strategies are observed in other paradigms, such as delayed-estimation or change localization tasks, is also an open research question.

The challenge, moving forward, will not be to simply generate potential encoding strategies, but rather to provide an explanation of the environmental factors that govern such strategies. Here, we propose that individuals appear to have adopted slot-like encoding because the number of items to encode was not predictable. However, Keshvari et al. (2013) found that resource-based models outperformed slots-based models in a task in which the memory set size varied from trial-to-trial. Their experiment differed from ours such that participants were asked to discriminate between items that were not always highly discriminable. It is possible that we observed slot-like encoding because we always used highly discriminable items. Such questions will require a lot of work to answer, but the computational modeling approach that we developed here seems to be a worthwhile start to such a pursuit. Our work also provides a clear proof-of-concept that such an approach is tractable. In our view, progress in the area of visual working memory will be accomplished by moving away from the debate over whether capacity is slots- or resource-based, and rather in developing a process model for the ways in which items are encoded (i.e., a model of the variable precision process that van den Berg, Ma and colleagues have proposed).

References

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*, 622–628.

Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1359–1367.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851–854.

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, 7*, 1–11.

Bengson, J. J., & Luck, S. K. (in press). Effects of strategy on visual working memory capacity. *Psychonomic Bulletin* & *Review*. http://dx.doi.org/10.3758/s13423-015-0891-7.

Brady, T. F., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*(3), 921–929.

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., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review, 119*, 480–499.

Donkin, C., & Nosofsky, R. M. (2012). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science, 23*, 625–634.

Donkin, C., Nosofsky, R. M., Gold, J., & Shiffrin, R. M. (2013). Fixed slots models of visual working memory response times. *Psychological Review, 120*, 873–902.

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Keshvari, S., Van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology, 9*.

Luck, S. K., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

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

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

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 332–350.

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

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*, 283–290.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using gibbs sampling.

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. D. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory and Cognition, 40*, 1226–1243.

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*, 5976–5979.

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 2214–2242.

Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attention and decision-making in brief multielement displays. *Psychological Review, 120*, 589–627.

Van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121*, 129–149.

Van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, 109*, 8780–8785.

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

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. P. P. P. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the savage-dickey method. *Cognitive Psychology, 60*, 158–189.

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

Zhang, W., & Luck, S. K. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233–235.

Zhang, W., & Luck, S. K. (2011). The number and quality of representations in working memory. *Psychological Science, 22*, 1434–1441.