Are Recognition Decisions in Visual Working Memory Different from Recognition Decisions in Long-Term Memory?

Eda Mızrak (eda.mizrak@gmail.com)[1,2] & Henrik Singmann (singmann@gmail.com)[3]
[1]Department of Experimental Psychology, University of Oxford
[2]Department of Psychology, Queen Mary University of London
[3]Department of Experimental Psychology, University College London

Abstract

What are the evidence distributions underlying recognition decisions in visual working memory, and how do they compare to those in long-term memory? A recent critical test shows that long-term recognition memory is best described by a Gumbel$_{\text{min}}$ signal-detection model. Our study applies this test to visual working memory. Participants studied shape-colour bindings and completed a recognition task with varying set size but an equal number of studied and non-studied items. Half of the participants had to select one studied item; half had to select one non-studied item. The Gumbel$_{\text{min}}$ model predicts that accuracy should improve with set size when selecting a studied item but remain unchanged when selecting a non-studied item. Our results are somewhat ambiguous but suggest that accuracy increased with set size in both task variants, which would provide evidence against the Gumbel$_{\text{min}}$ signal-detection model for visual working memory. Future research needs to replicate our results in a design without potential experimental confounds.

Keywords: visual working memory; recognition memory; signal detection theory; critical test

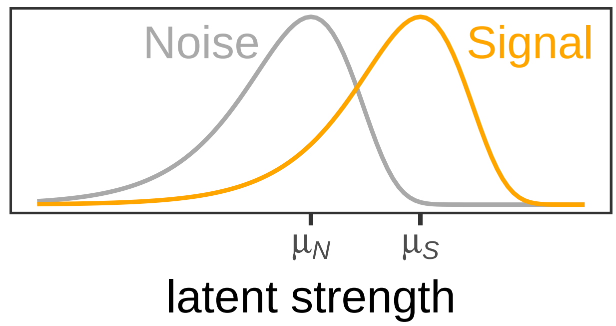

Figure 1: Illustration of the Gumbel$_{\text{min}}$ signal detection model. The two evidence distributions (noise corresponding to not-studied items and signal corresponding to studied items) are shown in different colours. The x-axis represents the latent strength axis, which is often assumed to represent familiarity in memory. The means of the noise distribution, μ_N, and the signal distribution, μ_S, are shown as well.

Introduction

What is the nature of the representations underlying recognition decisions in memory? This long-standing research question is often framed as one between continuous-resource models on the one hand and discrete-slots or threshold models on the other hand. According to the continuous resource models, memory exists in graded form, can be flexibly allocated to studied items, and items can be remembered partially. In contrast, according to discrete-slots or threshold models, studied items can be either in memory or not; if the memory information available for an item is below the memory threshold the decision-maker acts as if no memory information is available.

If we look at the literature from just a few decades back, this question seems to have been answered already: Long-term memory is best described by a continuous-resource model, specifically the signal detection model assuming unequal-variance Gaussian evidence distributions (e.g., Pazzaglia et al., 2013; Wixted, 2007), in contrast, visual working memory is best described by discrete-slots models (e.g., Luck & Vogel, 1997; Rouder et al., 2008; Zhang & Luck, 2008).

In more recent years this clear delineation between long-term memory and visual working memory seem to have disappeared. In particular, based on a combination of developments in experimental design, model development, and model comparison, results now suggest that visual working memory is also governed by a continuous-resource process (e.g., Donkin et al., 2016; Ma et al., 2014; Robinson et al., 2020; Williams et al., 2022). Importantly, most of this more recent work follows the same general paradigm as the earlier work and derives its evidence from fits of specific parametric implementations of continuous-resource models, discrete-slots models, and the comparison between both.

The goal of the present work is to evaluate how far the similarity in the representations underlying working memory and long-term memory recognition decisions really goes. Furthermore, we go beyond the general paradigm used in most of the previous work that relied on fitting specific parametric versions of continuous-resource models. Instead, we use a recently developed critical test that provided evidence on the shape of the evidence distribution underlying recognition memory decisions in long-term memory (Meyer-Grant et al., in press). In particular, contrary to the prevailing assumption that the evidence distribution in long-term recognition memory is Gaussian, this test provided strong evidence that it is a minimum-extreme Gumbel – that is, the Gumbel$_{\text{min}}$ – distribution (illustrated in Figure 1). Thus, our goal is to test whether recognition decisions in visual working memory also follow a Gumbel$_{\text{min}}$ distribution.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Critical Tests in Recognition Memory

Even though continuous-resource and discrete-slots models are often described as two very different models, Kellen et al. (2021) showed that both models can be subsumed in the class of models satisfying *random-scale representation* (Falmagne, 1978). The shared assumption of this model class is as follows: Consider a memory task in which participants are presented with two or more items at test, one of which was studied before, and the task is to select the studied item. Models with random-scale representation assume that participants pick the item with the largest latent-strength value on a latent-strength dimension, where both studied and non-studied item are represented by some random variable. Both the continuous-resource and discrete-slot models share this assumption. Where they differ is in the shape of the underlying evidence distributions.

The theoretical result that the main models for recognition memory are members of the same model class led to a series of exciting new empirical results in the long-term memory domain. The first set of results comes from Kellen et al. (2021) who, using a critical test approach (based on formal results of Block & Marschak, 1974; Falmagne, 1978), focussed on the question if there is evidence for *any* model assuming random-scale representation. In two experiments they showed that long-term memory recognition judgements satisfy the constraints imposed by random-scale representation. Thus, their work provides empirical evidence that *some* model assuming random-scale representation *must* underlie recognition judgements in long-term memory. What Kellen et al.'s work did not provide was a direct answer to the question of what the shape of the evidence distributions is. The only direct evidence for this question comes from their Experiment 3, which ruled out threshold model variants, favouring evidence distribution shapes consistent with continuous-resource models.

Building on the results of Kellen et al. (2021), Meyer-Grant et al. (in press) investigated the shape of the evidence distributions in long-term memory. One part of their results was based on parametric model comparisons and showed that much of the existing long-term memory data is well described by a continuous-resource or signal detection model assuming $\text{Gumbel}_{\text{min}}$ distributions; at least as well as the prevailing unequal-variance Gaussian signal detection model.

In addition to parametric model comparisons, Meyer-Grant et al. (in press) also developed a new critical test for the $\text{Gumbel}_{\text{min}}$ distribution.[1] Their test is based on the behavioural principle of "invariance under uniform expansions of the choice set" (Yellott, 1977) for the $2M$-task. The $2M$-task, illustrated in Figure 2, is a recognition task where always the same number (i.e., M) studied and non-studied items are presented: 1 studied and 1 non-studied item (i.e., $M = 1$), 2 studied and 2 non-studied items (i.e., $M = 2$), etc. The $2M$-task exists in two task variants, the $2M$-MIN and $2M$-MAX task. In the $2M$-MIN variant the participants' task is to select one non-studied item from the test set; in the $2M$-MAX variant the participants' task is to select one studied item from the test set. The $\text{Gumbel}_{\text{min}}$ uniquely predicts that in the $2M$-MIN task accuracy should be constant across M, whereas in the $2M$-MAX task accuracy should increase with M. In other words, increasing the number of studied and non-studied items together should not affect the accuracy for selecting a non-studied item whereas it should improve accuracy for selecting a studied item. Other continuous-resource models, such as those assuming Gaussian evidence distributions, predict that changes in M simultaneously affect performance for both the $2M$-MIN and $2M$-MAX task (not necessarily to the same degree).

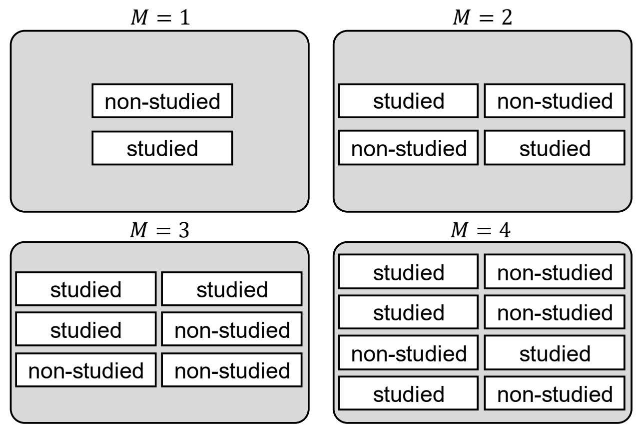

Figure 2: Illustration of the different test trials in the $2M$-task of Meyer-Grant et al. (in press). After studying a set of stimuli, each test trials presents participants with the same number of studied or non-studied stimuli. Depending on the task variant, they have to select one studied or one non-studied item.

Meyer-Grant et al. (in press) studied the $2M$-task in long-term memory. Participants first studied 100 words before working on forty test trials with varying M with test-type manipulated between-subjects. Participants either worked on the $2M$-MIN variant or the $2M$-MAX variant of the task for all forty trials. The corresponding results are shown in Figure 3. As can be seen, the prediction of the $\text{Gumbel}_{\text{min}}$ model, constant accuracy for the $2M$-MIN task and increasing accuracy for the $2M$-MAX task, is beautifully confirmed.

Independent evidence for the viability of the $\text{Gumbel}_{\text{min}}$ distribution in long-term recognition memory comes from a recent study by Dunn et al. (2025). They used a theoretically and technically sophisticated non-parametric approach to estimate the cumulative distribution function of the evidence distribution (based on methods developed in Dunn & Anderson, 2024). For two out of three experiments, the $\text{Gumbel}_{\text{min}}$ distribution provided a clearly better account of the estimated

[1] While we highlight the $\text{Gumbel}_{\text{min}}$ distribution here, the critical test actually applies to all members of the family of minimum-extreme generalised extreme-value distributions, $\text{Gumbel}_{\text{min}}$, $\text{Fréchet}_{\text{min}}$, and $\text{Weibull}_{\text{min}}$. Of these, the $\text{Gumbel}_{\text{min}}$ is the only Case V Thurstonian model (Thurstone, 1927). Therefore, to streamline exposition we focus on the $\text{Gumbel}_{\text{min}}$ in the text, but all results equally apply to the other members of the minimum-extreme generalised extreme-value family.

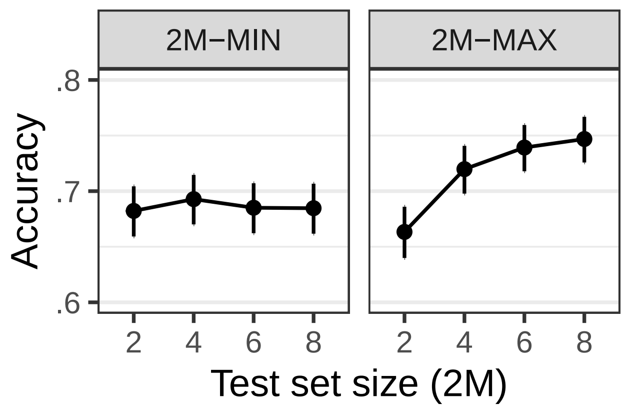

Figure 3: Results from $2M$-task in long-term recognition memory from Meyer-Grant et al. (in press), error bars show 95%-confidence intervals. In each test trial the same number of M studied and non-studied items are shown to participants and they have to either one non-studied (in the $2M$-MIN task) or one studied item ($2M$-MAX task). In line with predictions of the Gumbel$_{min}$ model, M does not affect accuracy in the $2M$-MIN task, but only in the $2M$-MAX task.

cumulative distribution function than the Gaussian distribution. Taken together, the results of Meyer-Grant et al. (in press) and Dunn et al. (2025) provide strong evidence that recognition decisions in long-term memory are best described by a signal-detection model assuming (equal-variance) Gumbel$_{min}$ distributions, which is illustrated in Figure 1.

The above summarised series of papers paints a clear picture of the shape of evidence distributions in long-term recognition memory. However, despite some evidence for the similarity of decisional processes between long-term and visual working memory (e.g., Williams et al., 2022), we cannot just assume that evidence distribution have the same shape in visual working memory. Such a conclusion requires new empirical evidence. A first step for providing such evidence was undertaken by He et al. (in press) who, building on the work of Kellen et al. (2021), applied the critical tests of random-scale representation to visual working memory. Experiment 1 of He et al. showed that recognition judgments in a simple working memory task testing only item memory satisfied the constraints of random-scale representation. Furthermore, Experiment 3 by He et al. showed that the same holds for a more complex working memory task requiring item-feature bindings, specifically, item-location bindings. Together, these results provide a strong empirical justification for using models with random-scale representation, such as the Gumbel$_{min}$ signal-detection model, in visual working memory. However, like Kellen et al. (2021), He et al.'s results provided limited insight into the shape of the underlying evidence distributions. This is the research gap we aim to address in the present study.

The Present Study

The aim of the present work is to apply the critical test for the Gumbel$_{min}$ signal-detection model developed by Meyer-Grant et al. (in press) to visual working memory. Participants performed a $2M$ visual working memory task in which they had to remember item-feature bindings. In each trial, participants were presented with a study display consisting of eight shapes in a unique, randomly chosen, colour. After the study display a test display with 2, 4, 6, or 8 items was shown (i.e., $M \in \{1, 2, 3, 4\}$). In each test display half of the presented shapes were shown in their original (i.e., studied) colour, the other half of the presented shapes were shown in a different, randomly chosen, colour. Participants were assigned to one of two between-subjects conditions. Half of the participants worked on the $2M$-MIN variant in which their task was to select one of the shapes shown in a new colour and the other half of participants worked on the $2M$-MAX variant in which their task was to select one of the shapes shown in its original colour. Our research question was whether changes in M only affected accuracy in the $2M$-MAX variant and left accuracy in the $2M$-MIN variant constant with M, which would indicate evidence for the Gumbel$_{min}$ signal-detection model. If in contrast we were to find that M affected accuracy for both task variants, this would provide evidence *against* the Gumbel$_{min}$ signal-detection model.

Method

Participants

A total of 194 young adults participated in the experiment. Of those, 94 participants were randomly assigned to the $2M$-MIN task variant and 100 to the $2M$-MAX task variant.

Participants were recruited from the online data collection platform Prolific. Inclusion criteria were being between 18–35 years old and having a normal or corrected-to-normal vision and a UK residency with English being their primary language, as the instructions for the experiment was in English. The experiment was programmed in `lab.js` (Henninger et al., 2022) and ran online. The experiment lasted 10-12 minutes, and participants were reimbursed with £9 per hour for their time. Participants completed an online informed consent form and were debriefed at the end. The experimental protocol was in accordance with the ethical guidelines of the Department of Experimental Psychology, University of Oxford.

Materials

As shapes we used 415 silhouette images of concrete nameable objects which were obtained via a web search for royalty-free clip art by Krasnoff and Souza (2024). Shapes were combined with one of eight colours we selected from the 360 colours of a colour wheel in the CIE L x a x b colour space, centred on L = 70, a = 20, and b = 38, with a radius of 60. The numbers of the eight colours we chose from this colour wheel were: 12, 61, 109, 163, 195, 231, 271, 309. This is the colour wheel most used in continuous-reproduction tests of visual WM, because it consists of colours that are

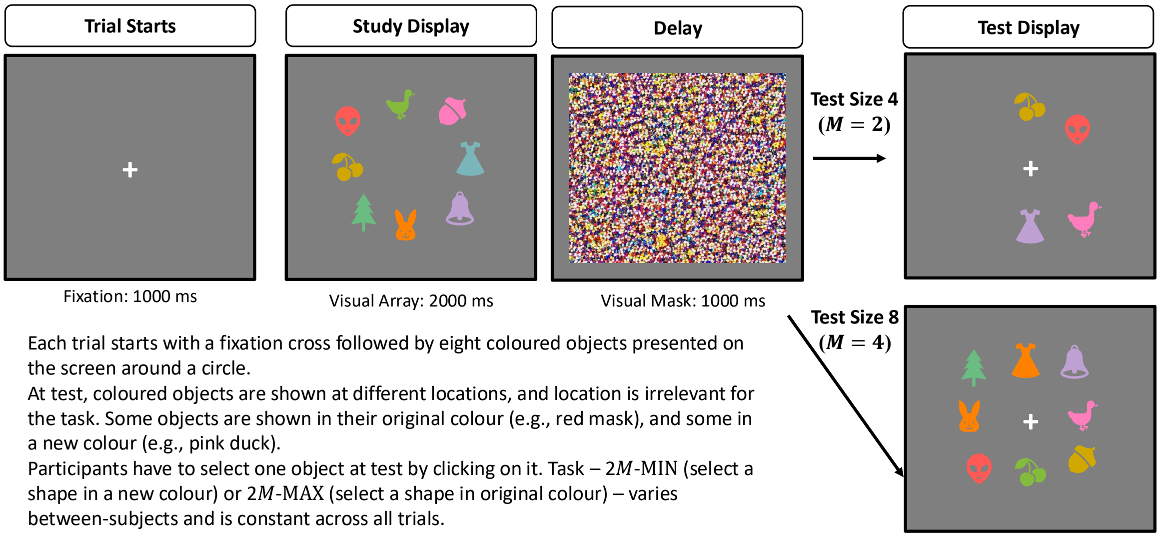

Figure 4: Overview of the experimental design and trial structure.

approximately equidistant in psychological similarity space, and approximately equally bright (for a critical discussion see Bae et al., 2014). The shapes (i.e., silhouette images) were presented uniformly in the chosen colour against a black background.

Procedure

Figure 4 gives an overview of the structure of each trial. Each trial began with a central fixation point displayed for 1000 ms. In the study display, eight coloured shapes (90 pixels[2] each) appeared simultaneously, arranged equidistantly around an invisible circle (radius = 150 pixels) for 2000 ms. A 1000 ms retention interval followed, during which a colourful square mask covered the full extent of the visual array. At test, participants saw either 2, 4, 6, or 8 shapes, all drawn from the study array but repositioned to new locations. Half of these test shapes retained their original colour, while the other half appeared in a different colour.

In the $2M$-MIN task variant, participants' goal was to select a shape from the test display shown in a new colour (i.e., a colour different from the one in the study display). In the $2M$-MAX task variant, participants' goal was to select a shape from the test display shown in its original colour (i.e., the colour in which this shape was shown in the study display). Participants responded by clicking directly on the object.

Shapes for each trial were sampled without replacement from the shape pool, with a new set appearing on each trial. The eight colours described above were used for all participants throughout the experiment. For the study display, colours were randomly assigned to objects without replacement, meaning each of the eight colours appeared exactly once and was paired with one random shape. At test, colours of the new objects were randomly assigned from the colour pool with replacement, after removing the original colour in which the item was studied. In other words, in the test sets each shape only appeared once, but the same colour could appear multiple times. The probability of a colour repeating in the test set increased with set size M (i.e., for $M = 1$ only 14% of trials included a repeated colour, for $M = 4$ over 99% of trials included a repeated colour).

To avoid that shapes could receive a memory boost through appearing at the same location at both study and test, the same shape could not appear at the same location at study and test. That is, each shape appeared at a different locations in the study and test display.

The experiment started with condition-specific instructions followed by three practice trials with a random test set size. After this, participants received a reminder of their instructions and worked on a total of 48 trials, 12 per test set size $2M$ with $M \in \{1, 2, 3, 4\}$. The trials were separated into four blocks, each containing 12 trials. There were five seconds breaks between blocks. Conditions were randomly assigned to each trial at each block. After working on all trials participants were debriefed and thanked for their participation.

Results

Figure 5 shows the distribution of by-participant mean accuracy (averaged across M) separately for each of the two task variants. Whereas for both task variants there were around ten participants with an accuracy at chance-level (i.e., .5), the vast

[2] Whereas all sizes are given in pixels, this does not reflect the actual size in pixels on the participants' screen. Instead, the pixel size given is for a screen of size 800 by 600 pixel. In the experiment, the full screen of the participants is used and the 800 by 600 pixels are upscaled to their actual screen size. In other words, the size in pixels is relative to a 800 by 600 screen, which was upscaled to match the participants' actual screen size. The experiment only ran in full screen mode and aborted if participants left full screen.

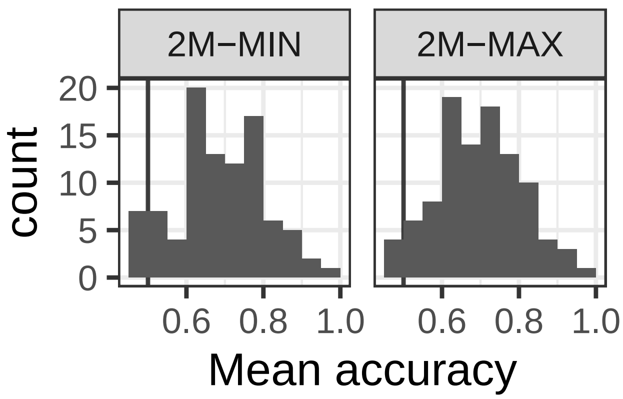

Figure 5: Distribution of by-participant average accuracy scores across both task conditions. The dark-grey vertical line at 0.5 indicates the chance-level performance.

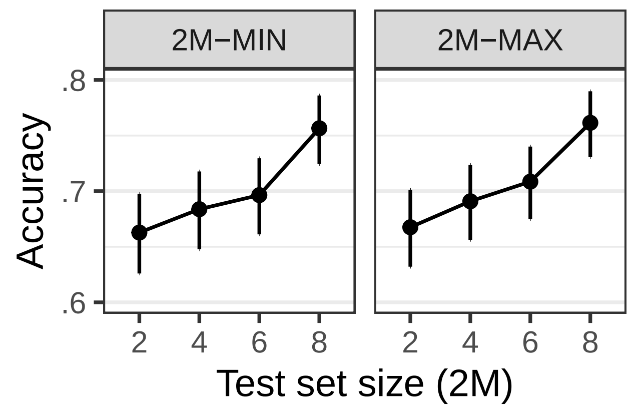

Figure 6: Accuracy across test set size $2M$ for each task variant. Black points show GLMM estimated marginal means with 95%-confidence intervals. Accuracy increases with M for both task variants, providing evidence against the Gumbel$_{\text{min}}$ model.

majority of participants performed the task well with most of the accuracies between .6 and .8. Importantly, for neither of the task variants were there individual participants with average accuracy clearly below chance indicating that none of the participants did the wrong task. Consequently, we included the data of all participants in the main analysis.

Figure 6 shows accuracy across task set size $2M$ separately for each task variant and a clear result: For both, $2M$-MIN and $2M$-MAX task, accuracy improves with M. Thus, this data appears to provide evidence *against* the Gumbel$_{\text{min}}$ model in visual working memory.

To further substantiate the pattern, we analysed participants' accuracy data with a generalised-linear mixed model (GLMM) assuming a binomial response distribution. For the main analysis we analysed both tasks together with fixed effects M, task variant, and their interaction as well as by-participant random intercepts and by-participant random slopes for M. As the maximal model with correlation among random term showed a singular fit, we report the results of a reduced model without correlations (Singmann & Kellen, 2019). The results of the maximal and null-correlation model were essentially the same (i.e., very minor quantitive differences and no differences in terms of statistical significance).

In line with the pattern seen in Figure 6, the GLMM showed a significant main effect of M, $\chi^2(3) = 50.02$, $p < .001$, and no significant task variant by M interaction, $\chi^2(3) = 0.09$, $p = .993$. Furthermore, there was no main effect of task variant, $\chi^2(1) = 0.18$, $p < .670$. As another robustness check, we also split the data by task variant and ran separate GLMMs for each task with a fixed effect of M. Both of these GLMMs also showed a clear effect of M, $\chi^2(3) \approx 24$, $p < .001$ (and this pattern was robust across random effect structures for the separate models).

One potential confound in our design was that for some, but not for all trials, at least one colour appeared multiple times among the test set. In the trials in which at least one colour appeared twice, participants could have used a heuristic and selected a shape that was shown in a repeated colour at random. If participants had followed this heuristic this would have resulted in a somewhat increasing accuracy for the $2M$-MIN task (0.50, 0.58, 0.60, and 0.60 for $M = 1$ to $M = 4$) and a somewhat decreasing accuracy for the $2M$-MAX task (0.50, 0.41, 0.40, and 0.39). Thus, this heuristic provides an alternative explanation for the results in the $2M$-MIN task.

To further evaluate the effect of this heuristic we looked at only those trials in which at least one colour appeared multiple times. We further split the data into two subsets, the trials where the participant chose a shape with a repeated colour and the trials where they chose a shape in a unique colour. The results of this way of looking at the data are shown in Figure 7 and are in line with the use of the heuristic (note that for $M = 1$, the choice is always for a repeated colour if only trials with repeated colours are included). In the $2M$-MIN task choosing a repeated colour leads to better performance that increases with M whereas it leads to worse performance with a potentially u-shaped pattern for the $2M$-MAX task. Interestingly, if we only look at the trials in which a unique colour is chosen we only see a clear increase of accuracy with M for the $2M$-MAX task, $\chi^2(2) = 12.23$, $p = .002$. For the $2M$-MIN task the main effect of M is no longer significant, $\chi^2(2) = 1.33$, $p = .514$, which would be in line with the Gumbel$_{\text{min}}$ model.

Taken together, the results provide a somewhat ambiguous picture. Whereas the pattern of all trials suggest that M affected accuracy for both task variants to an approximately equal degree, there is the potential that this increase for the $2M$-MIN task is driven at least partially by the use of a heuristic enabled by a possible experimental confound. On the other hand, across all trials accuracy is very similar across both task variants which means the use of the heuristic across conditions must have happened exactly in such a way to produce this

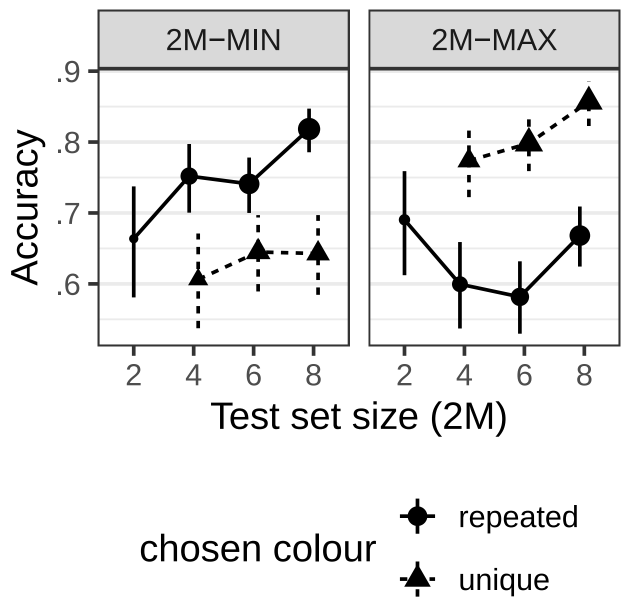

Figure 7: Accuracy across task set size $2M$ for only those trials in which at least one colour appeared twice, split by whether participants chose a shape in a unique or repeated colour. The size of each data point shows the number of observations per point which ranges from 161 to 730.

similarity across condition, an a priori unlikely coincidence. Thus, on balance the data seems to provide some evidence against the Gumbel$_{min}$ model in visual working memory.

Discussion

The present study aimed to test the extent to which memory representations are similar across visual working memory and long-term memory. To this end, we applied the critical test for the Gumbel$_{min}$ signal-detection model (Meyer-Grant et al., in press) to a $2M$ visual working memory task. After studying a set of eight shape-colour bindings, participants were presented with a test display containing M studied and M non-studied shape-colour bindings. Half of the participants had to select a non-studied shape-colour pair (in the $2M$-MIN task) and half of participants had to select a studied shape-colour pair (in the $2M$-MAX task). In the long-term memory domain this critical test had provided evidence for the Gumbel$_{min}$ model; M did not affect accuracy in the $2M$-MIN task but only in the $2M$-MAX task. Our results, while somewhat ambiguous, showed that overall accuracy increased with M for both task variants. Thus, our results provide some evidence against the Gumbel$_{min}$ model for visual-working memory.

The validity of our results relies on the assumption that at test all studied shape-colour bindings and all non-studied shape-colour bindings are exchangeable; that is, there are no features that allow participants to distinguish the different stimuli within one item class. One could argue that by having some colours repeated at test, this assumption was potentially violated. One way in which we have attempted to address this potential confound is by focussing only on trials in which at least one colours was repeated and looking at the pattern separately for those trials in which participants chose a repeated colour or in which they chose a unique colour. This reanalysis showed that, when only focussing on unique colours we found that accuracy seemed to be unaffected by M in the $2M$-MIN task. However, what this result cannot readily explain is why, when looking at the data across all trials, we see virtually no differences across conditions.

Our ultimate goal was to see whether the critical test for the Gumbel$_{min}$ signal detection model provides a way of distinguishing working memory from long-term memory – a distinction also known as short-term versus secondary memory, or primary versus secondary memory – a question that has sparked debates for decades. Some researchers question whether the distinction holds up at all (Nairne, 2002). Even among those who accept that there is a difference, disagreement persists: Where exactly does one form of memory end and the other begin? How separate are they really? And how do they interact? These questions are still discussed intensely (Cowan, 2019; Foster et al., 2024; Norris, 2017).

Our study builds on earlier critical tests providing evidence that models assuming a random-scale representation (Falmagne, 1978) must underlie recognition decisions in both long-term (Kellen et al., 2021) and visual working memory (He et al., in press). These critical tests did not provide evidence on the shape of the evidence distributions leaving the possibility for a difference between both memory domains open. For long-term memory Dunn et al. (2025) and Meyer-Grant et al. (in press) provided evidence that it is the Gumbel$_{min}$ model. Our results provide some suggestive evidence that in working memory it is not the Gumbel$_{min}$ model.

Because of the potential confound in our experimental design and the resulting ambiguity of our results the next step needs to be replicating our results in a design in which the potential confound is removed. One obvious way to do so would be to make sure that each colour at test is unique. However, such a design would introduce another potential confound and is therefore not recommended: In the $2M$-MIN task participants could use a “recall-to-reject” strategy if they remember only a single shape-colour binding. For example, if participants know that acorn was shown in pink, seeing a different object in pink would be enough to know that this is a new shape-colour binding if colour is unique. Therefore, a better way would be to have repeated colour at both study and test. Consider a design in which repeated colours could appear both at study and at test; such a design would not allow a recall-to-reject strategy. However, such a design also need to be set up in a way that it does not allow the use of heuristics as done in the current study. Future research should pursue the idea that working memory and long-term memory can be differentiated in their recognition decision by improving the study design.

Author Information

Both authors contributed equally to this work.

References

Bae, G.-Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), 7. https://doi.org/10.1167/14.4.7

Block, H. D., & Marschak, J. (1974). Random Orderings and Stochastic Theories of Responses (1960). In *Economic Information, Decision, and Prediction* (pp. 172–217). Springer Netherlands. https://doi.org/10.1007/978-94-010-9276-0_8

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin*, *145*(8), 822–847. https://doi.org/10.1037/bul0000199

Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. *Cognitive Psychology*, *85*, 30–42. https://doi.org/10.1016/j.cogpsych.2016.01.002

Dunn, J. C., & Anderson, L. M. (2024). The monotonic linear model: Testing for removable interactions. *Psychological Methods*. https://doi.org/10.1037/met0000626

Dunn, J. C., Anderson, L. M., & Stephens, R. (2025, November 17). *A shift representable signal detection model of recognition memory*. https://doi.org/10.31234/osf.io/6mv75_v1

Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, *18*(1), 52–72. https://doi.org/10.1016/0022-2496(78)90048-2

Foster, J. J., Vogel, E. K., & Awh, E. (2024, June 25). Working Memory as Persistent Neural Activity. In M. J. Kahana & A. D. Wagner (Eds.), *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.13

He, Y., Kellen, D., & Singmann, H. (in press). Examining the Signal-Detection Account of Visual Working Memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*. https://doi.org/10.31219/osf.io/w287y_v3

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). Lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573. https://doi.org/10.3758/s13428-019-01283-5

Kellen, D., Winiger, S., Dunn, J. C., & Singmann, H. (2021). Testing the foundations of signal detection theory in recognition memory. *Psychological Review*, *128*(6), 1022–1050. https://doi.org/10.1037/rev0000288

Krasnoff, J., & Souza, A. S. (2024). I remember it now, so I'll remember it later: Working memory strength guides predictions for long-term memory performance. *Memory & Cognition*, *52*(8), 1775–1797. https://doi.org/10.3758/s13421-023-01514-3

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

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Meyer-Grant, C. G., Kellen, D., Harding, S., & Singmann, H. (in press). Extreme-Value Signal Detection Theory for Recognition Memory: The Parametric Road Not Taken. *Psychological Review*. https://osf.io/qhrfj

Nairne, J. S. (2002). Remembering Over the Short-Term: The Case Against the Standard Model. *Annual Review of Psychology*, *53*, 53–81. https://doi.org/10.1146/annurev.psych.53.100901.135131

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, *143*(9), 992–1009. https://doi.org/10.1037/bul0000108

Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*(6), 1173–1203. https://doi.org/10.1037/a0033044

Robinson, M. M., Benjamin, A. S., & Irwin, D. E. (2020). Is there a K in capacity? Assessing the structure of visual short-term memory. *Cognitive Psychology*, *121*, 101305. https://doi.org/10.1016/j.cogpsych.2020.101305

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975–5979. https://doi.org/10.1073/pnas.0711295105

Singmann, H., & Kellen, D. (2019). An Introduction to Mixed Models for Experimental Psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4–31). Psychology Press.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273–286. https://doi.org/10.1037/h0070288

Williams, J. R., Robinson, M. M., Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2022). You cannot "count" how many items people remember in visual working memory: The importance of signal detection–based measures for understanding change detection performance. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(12), 1390–1409. https://doi.org/10.1037/xhp0001055

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Yellott, J. I. (1977). The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*(2), 109–144. https://doi.org/10.1016/0022-2496(77)90026-8

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