Visual Working Memory: Disentangling the Effects of Meaningfulness on Item- and Binding Memory

Submitted by: Anh TO-LEFEBVRE

Internship supervisor: Dr. Philipp MUSFELD & Prof. Dr. Klaus OBERAUER

Internship advisor: Dr. Pascal ROULLET

Internship institution: Department of Psychology, Cognitive Psychology,

University of Zurich, Switzerland

Adresse: Binzmühlestrasse 14, Box 22 I CH-8050 Zürich

Phone: +41 44 635 74 70

E-Mail: k.oberauer@psychologie.uzh.ch

Remerciement

A special thanks to Prof. Klaus Oberauer, Ordinarius of the Department of Cognitive Psychology at the University of Zurich, for the opportunity to work on this project with you and Philipp, and for all the valuable and constructive advice.

I would like to express my gratitude to Dr. Philipp Musfeld, from the University of Zurich, who supervised me throughout my internship. Thank you for hours of discussion on Bayesian statistics and the M^3 model, for the exchange of ideas on the project, and for your patience in teaching me.

Je tiens à exprimer ma gratitude à Monsieur Devaud et Monsieur Roullet pour tout le soutien et l'attention qu'ils ont apportés aux étudiants en les encourageant à mener à bien leur premier stage de Master en Neurosciences.

Enfin, je voudrais remercier David Lefebvre pour m'avoir toujours encouragé à aller de l'avant.

Abstract

Working memory is the mechanism that allows us to actively maintain and momentarily manipulate information during complex cognitive activities, such as memorising a sentence or tracking several items in sequence. This capability, however, is not limited by the number of elements but by the ability to associate the item and its context, as suggested by the *Binding Hypothesis.* While semantic meaning is known to enhance item memory in visual working memory (VWM) tasks, its effect on binding and item memory remains unclear. Some evidence suggests that meaningfulness may enhance feature binding, while other studies observed the effect for item memory. To address such questions, this study investigates the effect of semantic meaning on both item memory and binding memory using a 6-alternative forced-choice (6-AFC) task with real-world objects, their scrambled counterparts, and artificial objects. We estimate separately memory parameters for item memory and binding memory by applying a Bayesian Hierarchical Framework to the *Memory Measurement Model* (M^3). The results of the study confirm the beneficial effect of semantic meaning on item memory: participants showed higher performance for real-world objects compared with scrambled and artificial objects. While there is evidence of the effect of meaningfulness on item memory, the evidence of an effect on binding memory remains inconclusive.

Keywords: Visual Working Memory, Binding Memory, Item Memory, Memory Measurement Model

Table of Contents

1. Introduction

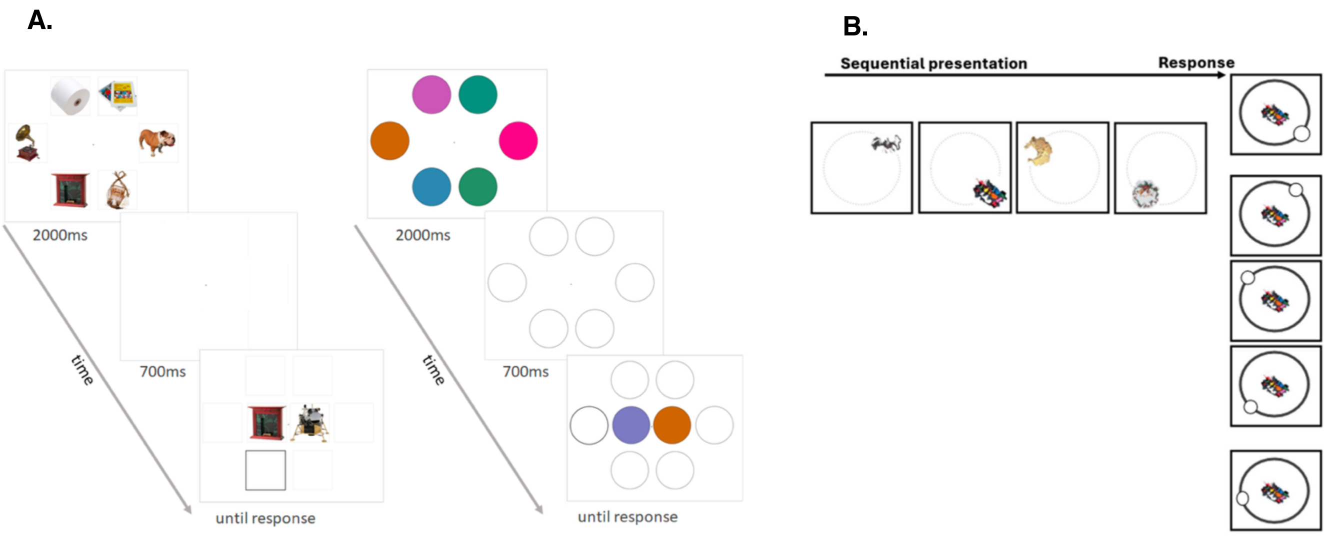
From retaining a sentence to keeping track of several items in sequence, working memory is the system that allows us to actively maintain and manipulate information temporarily while performing complex cognitive tasks. Evidence shows that such impairment in working memory is a characteristic of certain neurological and psychiatric syndromes, such as attention deficit, Alzheimer's disease, or Schizophrenia (Postle & Oberauer, 2024). According to Oberauer (2019), there are three principal functional features of working memory. Firstly, working memory provides a medium for forming and retaining temporary bindings between content elements and their context, which in turn allows structured representations to be constructed. Secondly, mechanisms must exist for manipulating these structures, such as an elective attentional mechanism or a mechanism for holding the procedural representation. Thirdly, we do not hold all the information, but only that which is relevant to guiding behaviour and which is consequently protected from retrograde and antegrade inferences and can be updated flexibly when necessary.

Working memory can be categorised into item memory and binding memory. Item memory refers to the retention of the items themselves, while binding memory involves the integration of these items with contextual information, such as their spatial location or their associations with other features (Oberauer, 2009). According to the *Binding Hypothesis* (Oberauer, 2019), the limited capacity of working memory is not simply due to the number of elements retained, but rather to the level of maintenance of these links between elements and their contexts. Through an experiment with a list of words in a specific location, it is demonstrated that as the set size increases, item memory remains generally robust, whereas binding memory becomes more error-prone, which confuses participants between the target item and another item.

The evidence of the binding mechanism is also observed in a neuropsychological study from Pertzov et al. (2013) which shows that during the visual working memory (VWM) test for both location and items, patients with limbic encephalitis tend to confuse the location of items that were shown in the encoding phase, while they do not have any problem in retrieving solely items. This suggests the importance of the medial temporal lobe, particularly the hippocampus, not for storing individual elements, but for binding elements or features together to create coherent representations.

Prior research has shown that semantic meaning enhances VWM performance. For instance, participants remember real-world objects better than low-level features like colour (Brady et al., 2016), or scrambled counterparts of such objects that are matched in visual complexity (Brady & Störmer, 2022; Thibeault et al., 2024). Colour can also be remembered better if it is part of a real-world object rather than a scrambled and meaningless object (Chung et al., 2023). Similarly, evidence shows that people remember meaningful stimuli better than visually similar but meaningless stimuli, such as real versus fake country flags (Conci et al., 2021) or identifiable versus unidentifiable Mooney faces (Asp et al., 2021), which highlights the role of semantic meaning in VWM.

Figure 1. (A): Experiment design by Brady and Störmer (2022) that arguably measured item memory, using a 2-AFC procedure with one target and one new item. **(B**) Experiment design from Sahar et al. (2024), which tested the effect of meaningfulness on the memory of item's location.



Although the effect of meaningfulness on VWM has been widely reported, and yet no research has examined its influence on both item memory and binding memory. Instead, existing designs have generally been found to test either item memory or binding memory, but not both. For instance, many studies employed recognition or 2-alternative-forced-choice (AFC) tasks (Brady et al., 2016; Brady & Störmer, 2022, 2023; Thibeault et al., 2024; Torres et al., 2024), as shown in Figure 1A. In this task, participants first see a list of stimuli during the encoding task and identify the stimulus that was presented earlier in the cued location by choosing between two response options during the retrieval task. At first glance, this task appears to test binding memory. However, one of the response options always contains a novel item that has never been presented before, which makes it more similar to a recognition task.

Based on the results of previous studies, it seems that the item memory was activated by the effect of semantic meaning.

Yet, other evidence suggests that meaning facilitates binding memory. Sahar et al. (2024) used a design in which different stimuli were presented at different locations, see Figure 1B. Rather than retrieving items from the specified location, participants had to remember the location of specific items. They focused on two types of response errors, such as choosing the location of other stimuli or guessing the location randomly. The result showed that swapping errors were reduced for real-world items compared to scrambled items. This suggests that meaning can reinforce the binding between an item and its location. Separating the contributions of meaning to item memory and binding memory is important for understanding their effect on VWM and is therefore a central focus of our research.

Additionally, the effect of meaningfulness has so far been examined by contrasting real-world objects with their scrambled counterparts. This allows for matching stimuli in visual complexity while suppressing the semantic meaning of the stimuli (Brady & Störmer, 2022; Chung et al., 2023, 2024; Thibeault et al., 2024). Nonetheless, the use of scrambled objects has certain drawbacks, such as creating unnatural object structures or irregular distributions of shapes, which can disrupt perceptual processing by activating solely low-level features or disjointed parts, rather than supporting recognition of a coherent whole. Consequently, the visual distinctiveness of scrambled objects is reduced. So, while scrambled objects represent a way of suppressing the object's meaning, they differ from real-world objects in other dimensions of visual processing. Nevertheless, no study so far has yet used novel objects – those that lack any conceptual meaning, while having a similar visual structure as real-world objects – to systematically examine the effect of semantic meaning.

In this study, we aim to replicate and extend previous findings on the meaningfulness benefit in VWM by disentangling the effects of semantic meaning on item versus binding memory. If semantic meaning does indeed facilitate such working memory processes, we expect to see an effect of meaningfulness not only on item memory but also on binding memory. On the other hand, if semantic meaning only improves item memory but not binding memory, this suggests that the effect might not be a benefit on working memory processes per se, but could reflect contributions from episodic long-term memory to working memory task performance (Bartsch & Oberauer, 2023; Oberauer, 2019). Additionally, we assess the benefit of semantic meaning by comparing memory performance for meaningful real-world objects to a) their scrambled counterparts and b) meaningless artificially created objects. If the effect does arise from the semantic meaning, we expect to see a memory benefit for real-world objects over both scrambled and artificial objects. If, however, the benefit is due to some form of holistic processing or visual distinctiveness, we expect to see a memory benefit for real-world objects only compared to scrambled objects, but not compared to artificial objects.

This paper is organised as follows: The methodology section describes the process of recruiting participants using the online platform *Prolific* and explains how the set of stimuli was defined. We use a 6-alternative forced-choice (6-AFC) task to dissociate binding and item memory. Details of how we used a Bayesian hierarchical framework with the *Memory Measurement Model* (M^3) to estimate memory parameters will also be described. The results section will present descriptive statistics and model-based estimates of item and binding memory parameters. This will be followed by a discussion that builds on the results in the context of our research questions and their theoretical implications on VWM. A short limitation observed during the studies and the suggested future directions will be provided.

2. Methodology

The present experiment is part of a large-scale project investigating the effects of semantic meaning on item memory and binding memory and was pre-registered on the Open Science Framework (OSF) before data collection to ensure transparency and openness of the research and data collection processes. Once the project is finished, the experiment codes, data analysis and data file, following GDPR, will also be uploaded to OSF. The experiment was programmed using *LabJS* (Henninger et al., 2022), an online study builder. The analyses were conducted using the R language and R packages such as *bmm* and *brms* (Bürkner, 2017; Frischkorn & Popov, 2025). We conducted this study in accordance with the guidelines proposed by the *Ethical Committee of the Faculty of Arts and Social Sciences* at the University of Zurich, and this experiment was carried out after the participants had given their informed consent. Before conducting the main experiment, we conducted a small pilot study of ten participants to test our experimental setup and to assess the difficulty of the task (See Appendix A)

2.1. Participants

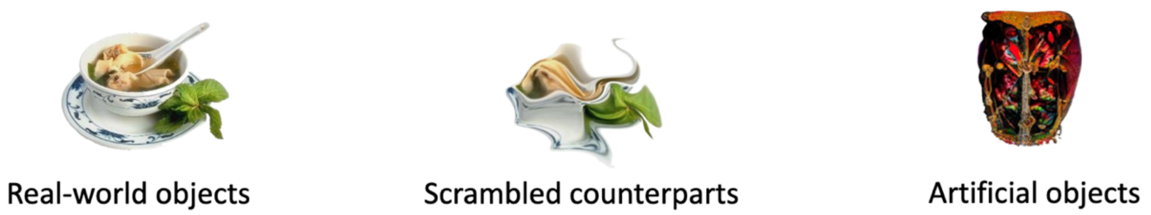
We collected initial data from 30 native English speakers between 18 and 35 years old using the online participation platform *Prolific*. Participants were compensated 9£/hour for their participation. Only individuals with normal or corrected to normal vision, no colour-blindness, and whose Prolific approval rate should be at least 98% took part in this experiment. The approval rate is based on the ratio of approval submissions to total submissions from previous studies. A higher rate indicates greater seriousness and reliability. After reading the instructions, participants were asked to answer a brief questionnaire about the task to ensure their full understanding.

2.2. Stimuli

To replicate the effect of meaningfulness on VWM, we used the same stimulus sets of real-world objects and their scrambled counterparts as in Brady & Störmer (2023). Additionally, we extended previous studies by including artificial objects using a stimulus set from Cooper et al. (2023). Because of identified duplicates in the stimulus set from Cooper et al. (2023), we reduced the stimulus size to 396 images based on the Structural Similarity Index. This index quantifies perceptual differences by assessing luminance, contrast, and structural similarity — attributes that are closely aligned with how the human eye processes visual information, such as sensitivity to edges and textures (Renieblas et al., 2017). We set the similarity threshold to 0.9 to identify duplicates and 0.8 to detect highly similar objects. Any objects above the threshold for duplicates were removed, while those exceeding the high similarity threshold were manually reviewed and removed if necessary.

To maximise comparability between stimulus sets, we also reduced the set of real-world and scrambled objects to 396. Images were excluded based on criteria such as low contrast, duplicated categories or potentially distressing content such as phobias. From the excluded objects, we took a set of 36 real-world objects for use in practice trials. In total, we used three different stimulus sets – real-world, scrambled, and artificial – each containing 396 images (see Figure 2). All images were standardised to a ratio of 300x300 pixels.

Figure 2. Example of stimuli



2.3. Design

The experiment used a within-subjects design with one three-level factor: real-world, scrambled and artificial stimuli. The design had one independent variable – type of object, which included three conditions: real-world, scrambled and artificial objects. We defined real-world objects as those that have both meaning and visual distinctiveness, scrambled objects as those that lack both characteristics, and artificial objects as those that have visual distinctiveness without semantic meaning. The dependent variable was the proportion of responses selected by participants in each category: target items that are identified in the correct location, within-list foils – items from the same trial but different locations and extra-list foils, which are new items that were never shown in the current trial.

The trial conditions were randomly intermixed throughout the experiment, with 50 trials per condition. In each trial, six unique objects of the same stimulus condition were presented at six different locations on the screen, arranged in an oval layout (see Figure 3). During the retrieval phase, one of these locations was cued, and participants had to identify the object that was originally presented at that location using a six-alternative forced choice task (6-AFC). The six response options contained one target item, two randomly selected within-list foils, and three randomly selected extra-list foils. Cued locations were counterbalanced across trials within the same condition to ensure equal probing frequency.

Each trial will require nine unique stimuli, with six for presentation and three for the response grid. Given that we only have 396 images in each stimulus set, we can only achieve 44 trials with unique stimuli. To reach 50 trials per condition, six additional trials were created by recycling images that had been used as extra-list response options at the beginning of the experiment to reach a total of 50 trials per condition. We argue that images used only in the extra-list response will not cause retrograde interference with the binding memory because they were not presented at any location. Trials with these repeated stimuli will all appear at the end of the experiment to minimise any potential confounding effects of stimulus repetition.

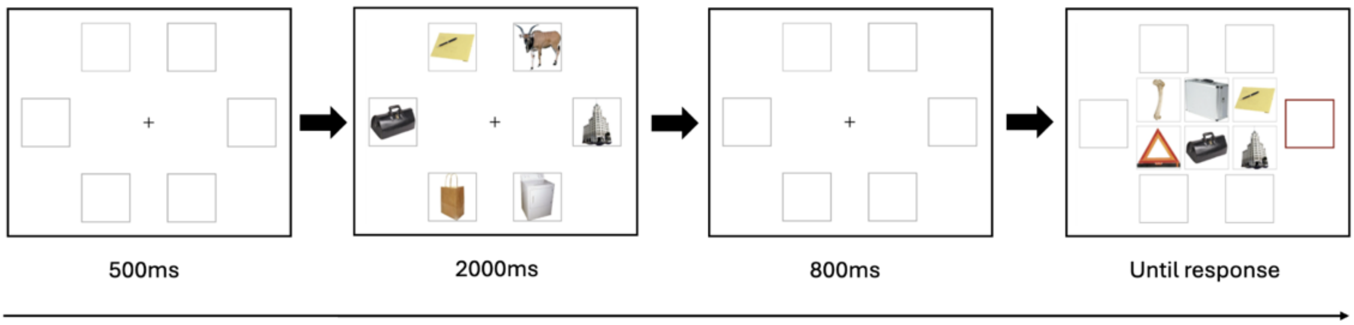
2.4. Procedure

Participants first completed three practice trials using real-world objects to familiarise themselves with the task. The main part of the experiment consisted of 150 trials in total, with 50 trials for each stimulus condition randomly intermixed. Each trial began with a fixation cross and six empty placeholders being presented on the screen for 500 ms. This was followed by the simultaneous presentation of six stimuli inside the placeholders for 2,000 ms and an 800-ms retention interval showing only the empty placeholders again (see Figure 3).

In the retrieval phase, one of the six locations was highlighted by a red frame, and participants identified the object and responded by clicking on one of the six options displayed in the centre of the screen. Immediate feedback was provided after each trial to encourage engagement with the task. After every 30 trials, participants could take a short break. Additionally, to encourage participants to make a consistent effort, a warning message was displayed if their accuracy fell below 20% over the last ten trials.

After completing all trials, participants were asked to provide some demographic information and indicate whether they had participated seriously in the experiment and whether they had used any aids to improve their performance in the task. Participants were then debriefed about the purpose of the study and asked to finish the experiment. The experiment lasted approximately 45 minutes.

Figure 3. Illustration of the experimental procedure. The participant was asked to recall six items and then to perform an immediate retrieval test, in which they had to remember the item at the indicated location.



For exclusion criteria, we decided to exclude participants who failed the instruction questionnaire twice in a row, left the experiment window more than three times, did not complete the experiment, admitted to using aid or did not participate seriously. Furthermore, we excluded participants whose performance was just above chance level. We defined this chance level as the 95th percentile of the binomial distribution and guessing probabilities of 1/6, as we have 6 response options and a total of 50 trials. This corresponds to 13 correct responses (26%) per condition.

2.5. Memory Measurement Model (M^3)

To dissociate the effects of meaningfulness on item vs. binding memory, data were analysed using the Memory Measurement Model (M^3), which is a computational measurement model framework for VWM developed by Oberauer & Lewandowski (2019). This model is developed to quantify continuous memory strength along two dimensions: memory for individual items and memory for temporary bindings between items and their context. Instead of giving a task on the recognition of location and asking whether they have seen this item in this location, the M^3 interprets participants' performance distribution by estimating memory strengths for both parameters, item memory (*a*) and binding memory (*c*). In our 6-AFC task, responses were grouped into three categories: (1) target items, which appeared in the correct location, (2) within-list foils, which were in the current trial but not in the cued location, and (3) extra-list foils, which were completely new and were not presented in the current trial. Consequently, each response category corresponds to a different level of activation.

According to M^3 estimation, the strength of each type of memory can be explained by an additive activation rule. As the extra-list foils were not presented in the trial, they should not elicit any activation beyond random background noise. This background noise parameter is called *(b)* and is fixed to 0. When participants made within-list errors, they mixed up the items presented during the trial, which indicates only activation of memory for the objects in the encoding phase, which is called *(a)* and provides a free parameter for measuring item memory. If participants selected target items, this indicates that they had correctly identified the location of the target. This process activates not only the item memory but also the memory of the location, which is referred to as *(c)* and is used to measure binding memory. In this way, each type of response can be classified according to the level of activation as follows:

$$A(target) = b + a + c$$

$$A(within - list) = b + a$$

$$A(extra - list) = b$$

2.6. Data Analysis

We performed all statistical analyses within a Bayesian hierarchical framework to compute M^3 as provided in the *bmm* package for R (Frischkorn & Popov, 2023), which conveniently estimates the parameters for item memory *(a)* and binding memory *(c)* separately for each stimulus condition. Both parameters were estimated with a log-link to ensure that the values are positive. We then apply Normal(0, 0.7) priors on both the binding and item memory parameters for all stimulus conditions, approximately resulting in a Normal(0, 1) prior on the differences between conditions. A *softmax* normalisation function was applied to transform the activation values into probabilities, with the response option with the highest activation resulting in the highest probability for being selected.

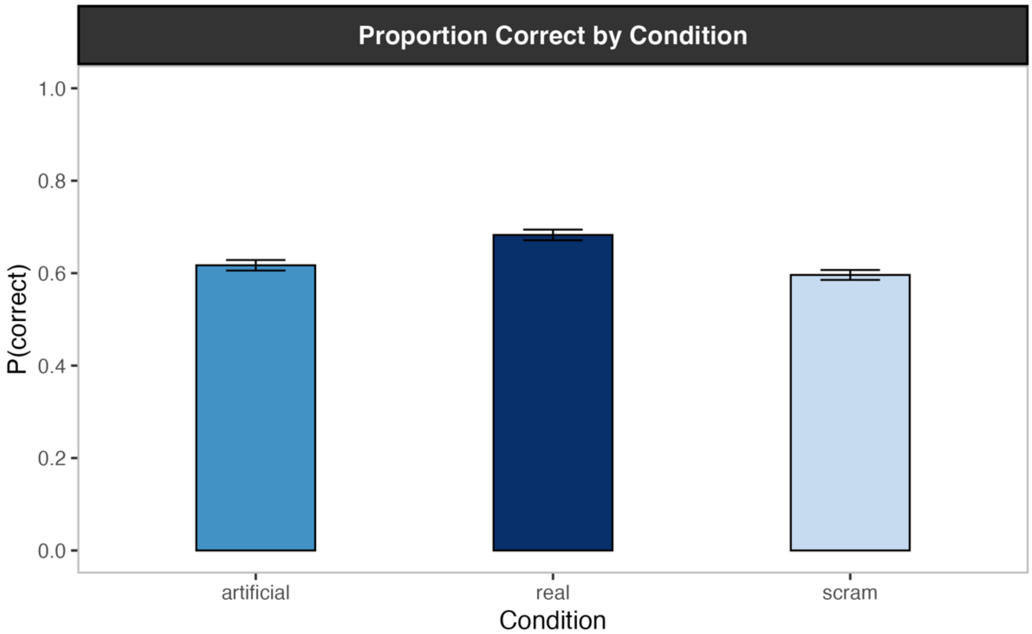
For our critical tests, we considered the difference in binding memory between real-world and scrambled objects, as well as between real-world and artificial objects. We calculated the posterior distributions for all pairwise differences to compare the two parameters *(a)* and *(c)* across stimulus conditions. Bayes Factors were computed using the *Savage-Dickey Density Ratio*, which compares the probability of each parameter value under the prior distribution to the probability of the same parameter value under the posterior distribution. We applied a Bayesian optional stopping procedure to determine the final size of our sample (Rouder, 2014; Schönbrodt et al., 2017), which means we initially collected a sample of N = 50 participants before conducting our initial analysis. If Bayes Factors for these comparisons did not meet our defined thresholds, $BF_{10} \geq 10$ for the presence or absence of a difference, we continued collecting data in batches of 10 participants. We decided not to collect data from more than 150 participants in total.

3. Results

3.1. Accuracy

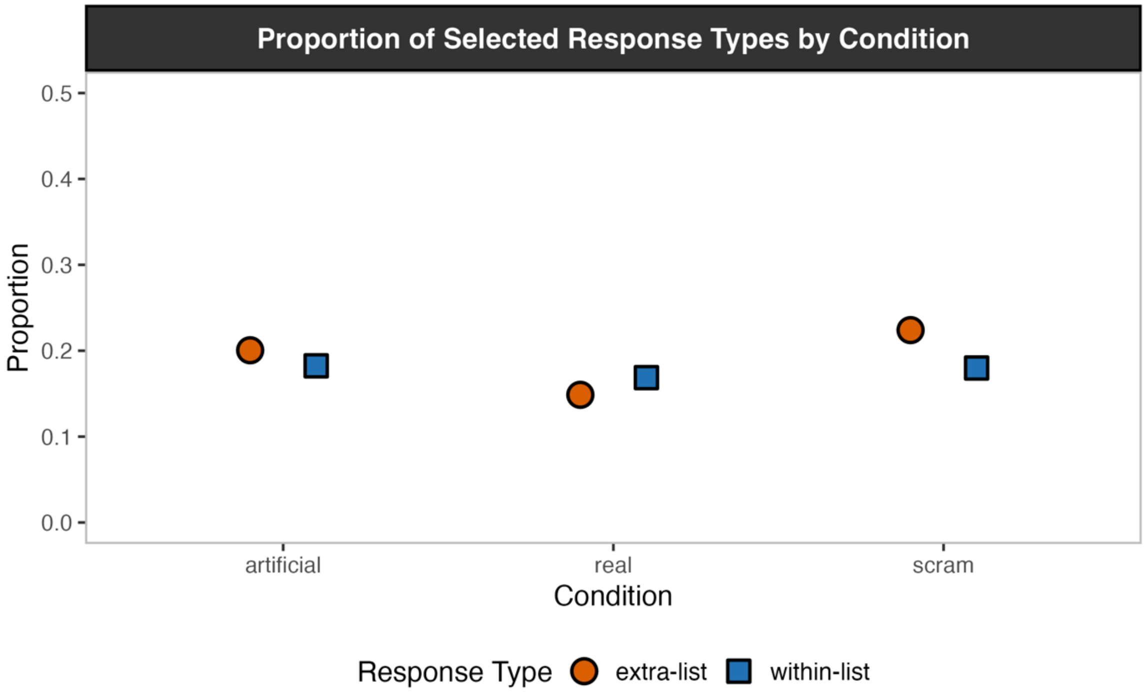
In total, we collected data from 150 participants. First, we examined the accuracy of participants' retrieval of the correct response in each of the three stimulus conditions by calculating the average proportion of correct responses in each condition, P(correct). As Figure 4 shows, overall performance was well above the 50% chance level, which indicates that participants performed better than would be expected by guessing. Their working memory performance was most accurate when the stimuli were real-world objects (M = 0.68, 95% CI [0.67, 0.69]). This was followed by the artificial condition, which had a slightly lower proportion of correct answers (M = 0.62, 95% CI [0.61, 0.63]). Scrambled objects had the lowest accuracy (M = 0.60, 95% CI [0.59, 0.61]. Among the conditions, participants' performance with the artificial objects was lower than with the real-world objects, which suggests that visual distinctiveness, shared by both types of objects, was not the main factor behind the differences in performance on the task. Regarding the accuracy of individual performances, we found that 19/150 participants (12.7%) achieved significantly high accuracy, above 0.9; we will address this issue in the Discussion section.

Figure 4. Analysis of average correct responses (N = 150). Proportion of correct responses across stimulus conditions (artificial, real-world and scrambled objects), with error bars representing 95% confidence intervals (CIs).



We then examined the types of errors made by the participants during the experiment by analysing the average proportion of responses that were either within-list or extra-list, in each of the three conditions. The results are shown in Figure 5. Overall, the proportion of each type of error was relatively similar and low at around 0.2 for each condition. The within-list error rate remained consistent across all conditions. Closer examination of the within-list and extra-list errors in each condition reveals different patterns, with higher within-list errors observed in real-world conditions and higher extra-list errors in scrambled and artificial conditions. This sheds light on the fact that item memory is more strongly activated when the objects are meaningful than when they are meaningless. However, given the relatively small and similar proportion of error types across all conditions, no conclusions should be drawn about the benefits of meaning for binding or item memory based on these results.

Figure 5. The proportion of response types (within-list and extra-list) selected across conditions (artificial, real-world, and scrambled objects) corresponds to the error types that participants made during the experiment when selecting another item from the encoding phase or choosing a new item (N = 150).



3.2. Fitting Memory Measurement Model (M^3)

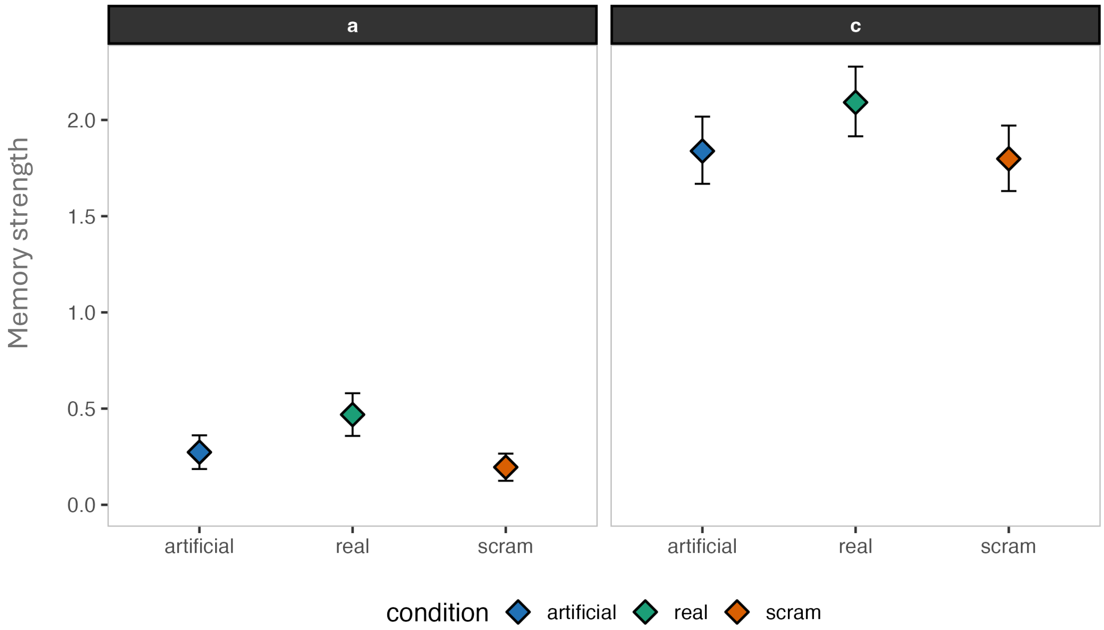
We applied Optional Stopping rule[1], which we started with the data collection of a sample size of 50 participants and continued to collect the data until the Bayes Factors reached the predefined threshold ($BF_{10} \geq 10$), or until a maximum predefined sample size of 150 participants was reached. We found that the early Bayes Factors remained below the predefined threshold, which led us to update our priors and continue collecting data. Eventually, we reached the maximum sample size of 150 participants.

We applied M^3 to estimate the memory strength for binding and item memory parameters separately, as shown in Figure 6. Overall, the estimates suggest that semantic meaning enhances both binding and item memory, as there is memory strength in both cases. As expected, real-world objects activate the highest memory strength for item memory (M = 0.46, 95% CI [0.35, 0.58]) and binding memory estimates (M = 2.09, 95% CI [1.92, 2.28]). Artificial objects activated item memory to a lesser extent (M = 0.27, 95% CI [0.18, 0.36]) and had a lower binding memory estimate (M = 1.83, 95% CI [1.67, 2.02]). Scrambled objects showed the

[1] *Optional stopping* is a robust Bayesian statistical practice that allows researchers to stop collecting data once the Bayes factor reached a predefined threshold indicating strong evidence for or against the hypothesis. This method enables flexible and efficient analysis while ensuring the validity of the results (Rouder, 2014)

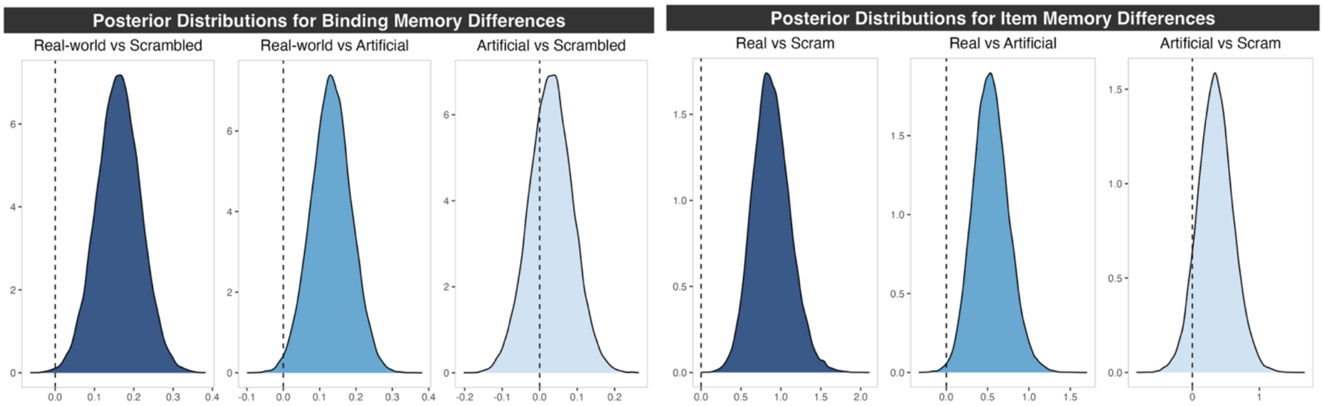
least activation across both types of memory, with estimates of M = 0.19, 95% CI [0.12, 0.27] and M = 1.79, 95% CI [1.63, 1.97], respectively.

Figure 6. Estimates of memory strength of the two parameters - item memory *(a)* and binding memory *(c)* - along with their posterior means and 95% credible intervals (CIs), which were exponentially transformed to ensure the positive value of parameters.



From the posterior distribution of differences in Figure 7, both binding and item memory show a consistent shift away from zero, i.e. the null hypothesis of no difference, in the pairwise comparison between the real-world and scrambled objects, and between the real-world and artificial objects. This suggests that both item and binding memory may benefit from semantic meaning. This does not apply to both binding and item memory when comparing the artificial and scrambled, as the posterior distribution is not shifted away from zero. Computing Bayes Factors revealed that in terms of item memory, a comparison between real-world and scrambled objects shows strong evidence for the effect of meaning (BF_{10} = 5571.41), whereas moderate support is also found when comparing real-world and artificial objects (BF_{10} = 7.97). However, this is not the case for binding memory. The difference between real and scrambled objects, as well as the difference between real and artificial objects, provides no evidence of a significant effect on binding memory, with BF_{10} = 0.95 and BF_{10} = 0.07, respectively.

Figure 7. Posterior distribution of the difference between three conditions for the item memory parameter *(a)* and the binding memory parameter *(b)*. The dotted line at zero represents the null hypothesis of no difference.



4. Discussion

Our present studies investigated how meaningfulness influences VWM performance. We attempted to first replicate the effect of meaningfulness of VWM, as demonstrated in previous studies using real-world and scrambled objects. We also extended the studies to artificial objects, which, despite lacking semantic meaning, can be perceived as ‘objects’ rather than visual blobs due to their visual distinctiveness. Our objective was to disentangle the effect of meaning on item and binding memory by estimating these parameters using M^3. We manipulated item and binding memory using 6-AFC tasks with three possible responses: target, within-list error and extra-list error.

The results show that meaningfulness enhances VWM performance. We found that real-world objects are remembered most accurately across the three conditions, and the estimates show the strongest memory for real-world objects compared to scrambled and artificial objects. Performance accuracy decreases for artificial objects, which are objects whose structure resembles that of an object, but which have no meaning, and the lowest accuracy is observed for scrambled objects. We have therefore successfully replicated the benefit of semantic meaning on VWM as demonstrated by previous studies (Brady & Störmer, 2022, 2023; Chung et al., 2024; Conci et al., 2021; Sahar et al., 2024; Thibeault et al., 2024; Torres et al., 2024). It suggests that the enhancement of VWM performance that we observed is primarily due to semantic processing rather than to low-level visual features or holistic object representations.

Regarding the types of errors in responses, within-list errors were slightly higher than for extra-list errors for real-world objects, whereas this tendency is reversed for artificial and scrambled objects. This seems to contradict the findings of Sahar et al. (2024) that meaningful objects can reduce confusion over the location of one item compared to another. The difference may result from the different experimental designs. Whereas in our studies we provided the location and asked participants to select the item presented there, in Sahar's studies, the item was provided, and participants were asked to select its location. Each design may influence the VWM differently and thus, activate different memory traces. While the results seem conflicting at first glance, they are nonetheless reasonable. As our studies did not yield conclusive evidence regarding binding memory, it would be necessary to replicate Sahar's studies using our stimulus set to verify whether the activation of binding memory varies accordingly. Future studies could then compare the differences in memory performance between these two designs.

So far, we know that memory performance improves when the stimuli are meaningful. The question is what drives this effect? Much research has emphasised the contribution of episodic long-term memory in facilitating the VWM capacity. For instance, Bartsch & Oberauer (2023) observed that VWM has the strongest influence on performance for tasks with smaller set sizes, whereas for larger set sizes, performance is mainly shaped by episodic long-term memory. It is possible that meaningfulness may be linked to some extent to long-term memory, which drives the enhancement of VWM tasks. Chung et al. (2024) assert that a meaningful object can activate several feature dimensions, such as high-level semantic knowledge, which in turn reinforces the memory trace when objects are meaningful. According to this point of view, the more semantically detailed the objects are, the better the VWM performance for that object. This is, however, not the case as demonstrated in another study from Kowialiewski et al. (2024), in which they compared the accuracy of performance of two different sets of words, one from six different categories and the other from six words in the same category. The results show that semantic similarity interference did not decrease performance, but rather enhanced VWM performance. Errors within the same category are also found to occur more frequently than errors arising from items in one category being confused with items in another. Further studies could explore the impact of semantic similarity between meaningful objects to better understand the specific mechanisms by which shared semantic features influence VWM performance.

Estimates from M^3 suggest strong evidence for the effect of semantic meaning in enhancing item memory, but inconclusive for binding memory. If there is indeed an effect on binding memory, this supports the *Binding Hypothesis* that the limitations of VWM stem not from limitations in item memory, but from the difficulty of maintaining contextual binding between items and locations (Oberauer, 2019). On the other hand, if meaningfulness does not activate binding memory, the benefit observed in the item memory may be driven by the activation of episodic long-term memory (Bartsch & Oberauer, 2023). Future research should focus on confirming the effect of meaningfulness on binding memory.

Upon data analysis, we found that 19 out of 150 participants achieved an accuracy rate of 90% in our experiment. As they did not admit to cheating at the end of the experiment, we have no evidence to accuse them of cheating. Nonetheless, we observed patterns in which some participants initially performed extremely poorly but then went on to achieve excellent results across all three conditions. Further analysis of the data also revealed a bimodal distribution. If these participants are high performers, our model may misestimate the parameters because these participants perform extremely well in all conditions, which consequently makes our model less sensitive. We argue that these participants are only interested in achieving high overall performance and do not favour any condition, so their data adds some noise to the data analysis. Additionally, we carried out a second analysis in which we excluded these high performers; the results did not alter any statistical conclusions (See Appendix B). For the following experiments, we decided to modify the exclusion criteria in preregistration and remove participants whose performance is above 90%.

Our results support the idea that the meaningfulness effect is driven by enhanced item memory, but it is inconclusive in binding memory. The next question is whether there will indeed be an effect on binding or not. The following experiment should focus on confirming the presence or absence of the effect of meaningfulness in VWM. This can be done by replicating previous studies, such as Sahar et al. (2024), to specifically assess how meaning affects binding memory, while controlling potential confounders such as semantic category or similarity. In addition, we estimated the parameters for M^3 independently. However, we did not account for the possible correlation between the variables. For instance, participants who performed well in one condition also tended to perform well in other conditions. In other words, high performance in one condition is generally associated with relatively high performance in all conditions. The differences between the statistical results for the covariates should therefore be examined in future studies.

References

Asp, I. E., Störmer, V. S., & Brady, T. F. (2021). Greater Visual Working Memory Capacity for Visually Matched Stimuli When They Are Perceived as Meaningful. *Journal of Cognitive Neuroscience*, *33*(5), 902–918. https://doi.org/10.1162/jocn_a_01693

Bartsch, L. M., & Oberauer, K. (2023). The contribution of episodic long-term memory to working memory for bindings. *Cognition*, *231*, 105330. https://doi.org/10.1016/j.cognition.2022.105330

Brady, T. F., & Störmer, V. S. (2022). The role of meaning in visual working memory: Real-world objects, but not simple features, benefit from deeper processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(7), 942–958. https://doi.org/10.1037/xlm0001014

Brady, T. F., & Störmer, V. S. (2023). Comparing memory capacity across stimuli requires maximally dissimilar foils: Using deep convolutional neural networks to understand visual working memory capacity for real-world objects. *Memory & Cognition*. https://doi.org/10.3758/s13421-023-01485-5

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences*, *113*(27), 7459–7464. https://doi.org/10.1073/pnas.1520027113

Bürkner, P.-C. (2017). brms: An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Chung, Y. H., Brady, T. F., & Störmer, V. S. (2024). Meaningfulness and Familiarity Expand Visual Working Memory Capacity. *Current Directions in Psychological Science, 33(*5), 275–282. https://doi.org/10.1177/09637214241262334

Chung, Y. H., Brady, T. F., & Störmer, V. S. (2023). No Fixed Limit for Storing Simple Visual Features: Realistic Objects Provide an Efficient Scaffold for Holding Features in Mind. *Psychological Science*, *34*(7), 784–793. https://doi.org/10.1177/09567976231171339

Chung, Y. H., Brady, T. F., & Störmer, V. S. (2024). Sequential encoding aids working memory for meaningful objects' identities but not for their colors. *Memory & Cognition*, *52*(8), 2119–2131. https://doi.org/10.3758/s13421-023-01486-4

Conci, M., Kreyenmeier, P., Kröll, L., Spiech, C., & Müller, H. J. (2021). The nationality benefit: Long-term memory associations enhance visual working memory for color-shape conjunctions. *Psychonomic Bulletin & Review*, *28*(6), 1982–1990. https://doi.org/10.3758/s13423-021-01957-2

Frischkorn, G. T., & Popov, V. (2025). A tutorial for estimating Bayesian hierarchical mixture models for visual working memory tasks: Introducing the Bayesian Measurement Modeling (bmm) package for R. *Behavior Research Methods*, *57*(5), 144. https://doi.org/10.3758/s13428-025-02643-0

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

Kowialiewski, B., Majerus, S., & Oberauer, K. (2024). Does semantic similarity affect immediate memory for order? Usually not, but sometimes it does. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *50*(1), 68–88. https://doi.org/10.1037/xlm0001279

Oberauer, K. (2009). Chapter 2 Design for a Working Memory. In *Psychology of Learning and Motivation* (Vol. 51, pp. 45–100). Elsevier. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K. (2019). Working Memory Capacity Limits Memory for Bindings. *Journal of Cognition*, *2*(1), 40. https://doi.org/10.5334/joc.86

Oberauer, K., & Lewandowsky, S. (2019). Simple measurement models for complex working-memory tasks. *Psychological Review*, *126*(6), 880–932. https://doi.org/10.1037/rev0000159

Postle, B. R., & Oberauer, K. (2024). Working Memory: Theoretical, Computational, and Neural Considerations. In M. J. Kahana & A. D. Wagner (Eds.), *The Oxford Handbook of Human Memory, Two Volume Pack* (1st ed., pp. 371–414). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.14

Renieblas, G. P., Nogués, A. T., González, A. M., Gómez-Leon, N., & Del Castillo, E. G. (2017). Structural similarity index family for image quality assessment in radiological images. *Journal of Medical Imaging*, *4*(3), 035501. https://doi.org/10.1117/1.JMI.4.3.035501

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Sahar, T., Gronau, N., & Makovski, T. (2024). Semantic meaning enhances feature-binding but not quantity or precision of locations in visual working memory. *Memory & Cognition*, *52*(8), 2107–2118. https://doi.org/10.3758/s13421-024-01611-x

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Thibeault, A. M. L., Stojanoski, B., & Emrich, S. M. (2024). Investigating the effects of perceptual complexity versus conceptual meaning on the object benefit in visual working memory. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-024-01158-z

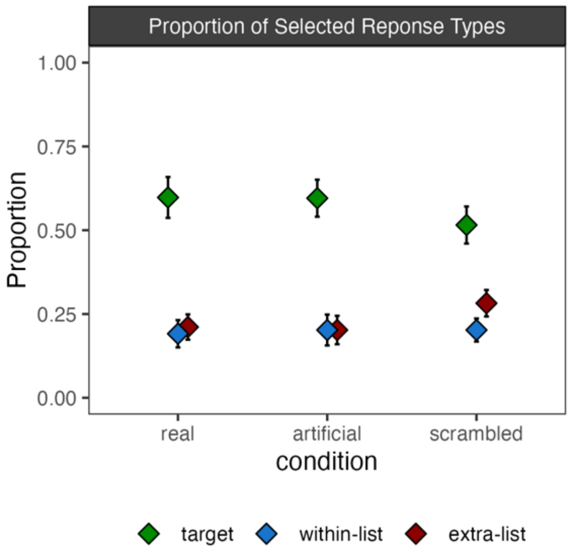
Torres, R. E., Duprey, M. S., Campbell, K. L., & Emrich, S. M. (2024). Not all objects are created equal: The object benefit in visual working memory is supported by greater recollection-like memory, but only for memorable objects. *Memory & Cognition*. https://doi.org/10.3758/s13421-024-01655-z

Appendices

Appendix A – Pilot studies

Before conducting our main experiment, we ran a pilot study with ten participants. One participant was removed due to admitting to using an aid during the experiment. The results show that there is a difference in the performance between real-world objects and scrambled objects. The performance of artificial objects was observed to be similar to that of real-world objects.

Figure 8. Visualisation of results from the pilot study (N = 9). Error bars reflect 95% within-subject confidence intervals.



Appendix B – Data excludes high-performance participants

We found that 19 out of 150 participants (12%) recruited for the experiment achieved an accuracy of 90%. Although this is a high level of accuracy, we have no evidence that they were cheating, so we included their data in the main analysis. However, we also ran an analysis excluding their data. The results did not change significantly and did not affect our conclusion based on Bayes Factors.

Figure 9. Estimates of item memory *(a)* and binding memory *(c)* after excluding 19 participants with extremely high performance (N = 131). Error bars reflect 95% within-subject confidence intervals.

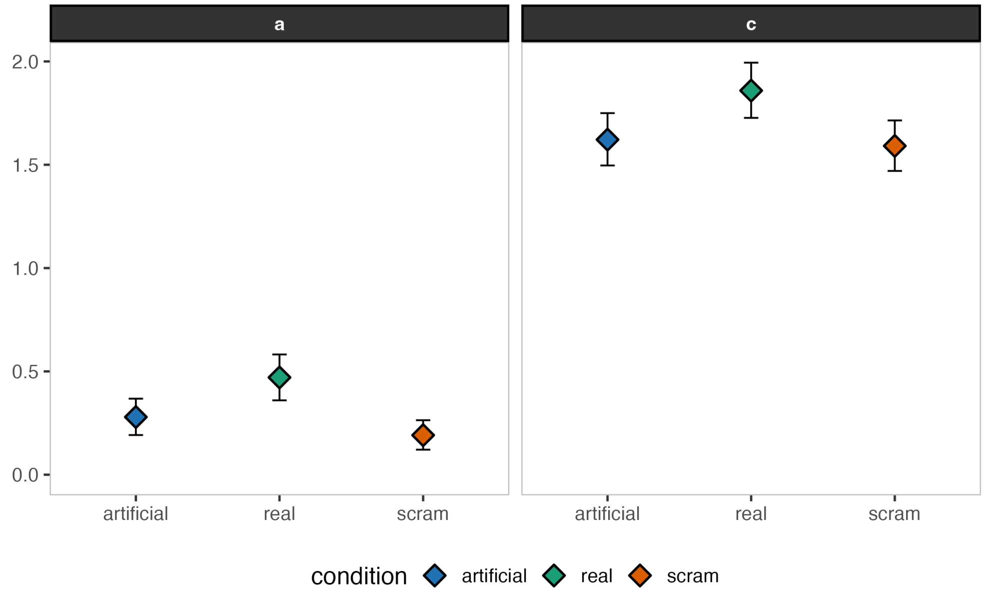


Table 1. Report of Bayes Factors in both item memory and binding memory across 3 pair-wise comparisons of conditions: artificial, real-world and scrambled objects.

Memory Type	Comparison	BF_{10} (Bayes Factor)
Item memory	(a) artificial vs. scram	0.79
	(a) real vs. artificial	6.09
	(a) real vs. scram	1540.35
Binding memory	(c) artificial vs. scram	0.06
	(c) real vs. artificial	1.31
	(c) real vs. scram	3.43