Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Highly similar and competing visual scenes lead to diminished object but not spatial detail in memory drawings

Elizabeth H. Hall, Wilma A. Bainbridge & Chris I. Baker

To cite this article: Elizabeth H. Hall, Wilma A. Bainbridge & Chris I. Baker (2022) Highly similar and competing visual scenes lead to diminished object but not spatial detail in memory drawings, Memory, 30:3, 279-292, DOI: 10.1080/09658211.2021.2010761

To link to this article: https://doi.org/10.1080/09658211.2021.2010761

View supplementary material

Published online: 16 Dec 2021.

Submit your article to this journal

Article views: 329

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=pmem20

MEMORY
2022, VOL. 30, NO. 3, 279–292
https://doi.org/10.1080/09658211.2021.2010761

Routledge
Taylor & Francis Group

Check for updates

Highly similar and competing visual scenes lead to diminished object but not spatial detail in memory drawings

Elizabeth H. Hall[a,b,c], Wilma A. Bainbridge[d] and Chris I. Baker[c]

[a]Department of Psychology, University of California Davis, Davis, CA, USA; [b]Center for Mind and Brain, University of California Davis, Davis, CA, USA; [c]Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD, USA; [d]Department of Psychology, University of Chicago, Chicago, IL, USA

ABSTRACT
Drawings of scenes made from memory can be highly detailed and spatially accurate, with little information not found in the observed stimuli. While prior work has focused on studying memory for distinct scenes, less is known about the specific detail recalled when episodes are highly similar and competing. Here, participants ($N = 30$) were asked to study and recall eight complex scene images using a drawing task. Importantly, four of these images were exemplars of different scene categories, while the other four images were from the same scene category. The resulting 213 drawings were judged by 1764 online scorers for a comprehensive set of measures, including scene and object diagnosticity, spatial information, and fixation and pen movement behaviour. We observed that competition in memory resulted in diminished object detail, with drawings and objects that were less diagnostic of their original image. However, repeated exemplars of a category did not result in differences in spatial memory accuracy, and there were no differences in fixations during study or pen movements during recall. These results reveal that while drawings for distinct categories of scenes can be highly detailed and accurate, drawings for scenes from repeated categories, creating competition in memory, show reduced object detail.

ARTICLE HISTORY
Received 21 June 2021
Accepted 20 November 2021

KEYWORDS
visual recall; scene perception; object recall; memory interference; categorical interference

In our daily lives, we often need to juggle memories for multiple similar scenes at the same time. For example, when buying a home, you may visit dozens of open houses and see dozens of kitchens, and you need to remember the features specific to each one to inform your purchase decision. Separating these scenes in memory can be an incredibly difficult task; while they may have unique visual features and layouts, they also likely share many objects, spatial qualities, and fall under the same umbrella semantic category (i.e., a kitchen). It is still largely debated the degree to which our memories for a scene are represented by broad semantic or gist information (e.g., Oliva & Torralba, 2006) versus detailed visual information (e.g., Hollingworth, 2005). Here, utilising a free recall drawing task combined with in-depth crowd-sourced scoring of these drawings, we assess the influence of categorical competition on the visual accuracy of scene memory. When we see multiple scenes from a given category, to what degree are we able to maintain isolated, accurate visual memories for a scene, or how much do we confuse these memories?

One perspective suggests that these competing memories should result in high memory interference. The Deese-Roediger-McDermott (DRM) paradigm is a methodology commonly used to elicit verbal false memories by testing participants on recognition of word lures that are semantically related to the study set (Roediger & McDermott, 1995). This paradigm can be adapted for images as well; showing participants images of semantically related objects to a studied scene (e.g., a broom that was not originally present in a cleaning scene) will also cause false memories for those objects (Miller & Gazzaniga, 1998; Baioui et al., 2012). Children have also been found to have high rates of false memories for semantically related items in an experiment using a similar paradigm (Otgaar et al., 2014). These same false memory effects occur not only just for semantically related items, but also when studying multiple exemplars of a given category, for both words (Hintzman, 1988) as well as object images (Strack & Bless, 1994; Koutstaal & Schacter, 1997; Seamon et al., 2000). This is thought to occur because of *associative activation*, in which semantically and categorically related information is co-activated during the

CONTACT Elizabeth H. Hall ehhall@ucdavis.edu Department of Psychology, University of California Davis, Davis, CA 95616, USA; Center for Mind and Brain, University of California Davis, Davis, CA 95618, USA; Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD 20814, USA
Supplemental data for this article can be accessed at https://doi.org/10.1080/09658211.2021.2010761.
© 2021 Informa UK Limited, trading as Taylor & Francis Group

activation of representations of studied items (Otgaar et al., 2019). Collectively, these results suggest that memory for specific details can be fallible, and that we are prone to confusing individual items, especially if they share the same category or semantic features.

However, an alternate perspective suggests that our memories for visual scenes are highly detailed and show little semantically-driven interference. In fact, when illustrative line drawings accompany the words in a DRM task, false memories sharply drop, attributed to the distinctive visual details introduced by the images (Israel & Schacter, 1997). Our ability to recognise specific scene images is also incredibly high, even when the memory system is overloaded with thousands of images across several hours (Standing, 1973; Konkle et al., 2010a). Importantly, even when participants are shown multiple images from the same category (as many as 64 exemplars per scene category), they only show a minimal decrease in recognition performance, with a 1.8% hit rate decrement for every doubling of same-category exemplars (Konkle et al., 2010b). Change blindness – our frequent inability to detect changes between two similar scenes – is often used as strong evidence that our memories for scenes are low-resolution (Simons & Levin, 1997). However, higher encoding times allow people to perceive more details within the scenes and detect these changes (Brady et al., 2009). Importantly, even during a rapid change blindness task, observers can detect changes in object location and within semantic category (Hollingworth & Henderson, 2002). This suggests that our memory representation for a scene is not merely a gist-based or semantic representation, as we are still sensitive to detailed visual changes, even after a 24-hour delay (Hollingworth, 2005).

A majority of these studies investigating the effects of semantic and categorical competition on visual false memories have focused on recognition rather than recall. It is thus unclear what specific features of an image drive false memories. An observer only needs to recognise (or not recognise) one specific visual detail to make a correct response on a recognition task. In contrast, a free recall task reveals which aspects of a memory are detailed and accurate, which are fuzzier, and which are clearly inaccurate. In addition, many of the studies showing an effect of semantic competition on false memories utilise low-information images, such as line drawings and objects (e.g., Strack & Bless, 1994; Koutstaal & Schacter, 1997; Miller & Gazzaniga, 1998; Seamon et al., 2000; Baioui et al., 2012), which may comprise few visual details and a sparse semantic representation. In contrast, real-world scenes capture the visual richness we encounter in our daily lives, with multiple levels of detail, including object identities, spatial relationships, and details within objects. This richness within scenes affords us a unique opportunity to investigate the specific source of memory errors, whether it be in specific types of objects or spatial detail. Thus, to understand the nature of memory interference, we argue that it is essential to examine free recall with complex images.

Recent work has employed *drawing* as a measure to investigate free recall of memory for scene images (Bainbridge et al., 2019). In contrast with verbal recall, drawing allows for individuals to visually produce representations of a memory, without translation into a verbal code. Verbal recall may also lose important visual information, such as precise spatial information or colour. Drawing as a method for visual recall has been able to reveal impressive levels of detail in memory for category-unique, real-world scene images; in a prior drawing recall study with 30 viewed scenes, observers drew on average 151.3 different objects across 12 scenes, with pixel-precise spatial accuracy, and only 1.83 falsely recalled objects (Bainbridge et al., 2019). This task has also been used to probe the relationship of semantic and visual information on drawing performance. For example, individuals with intact semantic representations but impaired visual recall and imagery (a condition termed *aphantasia*) showed impaired levels of object detail in this memory drawing task (Bainbridge et al., 2021a). This suggests that nuanced differences in memory for specific visual details can be captured by memory drawings. This task has also been used to show that disruptions of object-scene semantics can lead to diminished visual memory detail and loosened object-scene bindings (Bainbridge et al., 2021b).

Here, using a drawing free recall task, we examined the influence of maintaining highly similar, competing scenes in memory on recall performance for individual scenes. Participants studied real-world scene images, where some scenes were unique to a category, while others were all from the same category. The objective scoring of these drawings was then crowd-sourced to 1764 unique scorers online, who completed a series of tasks to quantify the overall drawings, their objects, object locations, and false memories. In the face of competition, we observed a significant influence of semantic competition on memory detail, with drawings and their objects becoming less detailed and less diagnostic. Surprisingly, we did not observe an influence of competition on spatial accuracy, false memories, or temporal fixation or drawing behaviour. Overall, these results show that categorical interference occurs during free recall of scenes, suggesting an influential role of categorical information on our visual memories.

Methods

In the current study, we quantified the amount of detail participants could recall from memory for scenes that were visually related, with a shared semantic label. First, we ran an online experiment to identify the typical objects within common scene categories and used these data to assemble a set of scene images that were distinct but also would cause competition within the same category. In the main in-lab experiment, participants studied eight of these scenes, (four category-unique, and four category-repeated), while their eye movements were tracked. After completing a brief distractor task, participants were asked to draw as many images as they could recall while their pen movements were tracked. Finally, they completed an old/new recognition task for the images. Participants' drawings were then uploaded to a series of five online scoring experiments, in which online scorers rated a drawing on (1) its diagnosticity for the original image, (2) which objects were present, (3) the diagnosticity of its objects, (4) the spatial locations of its objects, and (5) the presence of false additional information. We then examined how these measures differed between category-unique and category-repeated drawings.

Participants

Thirty-one experimentally-naive adults (8 male, 23 female; average 23.9 years of age, SD = 2.8) were recruited from the local Washington, DC area for participation in this experiment. Data from one participant were excluded due to calibration errors in the eye-tracking portion of the experiment, leading to the participant seeing images twice during the study phase. This left data from 30 participants available for analysis. This sample size of 30 was predetermined based on earlier drawing experiments that found this sample size was sufficient to identify group differences on other tasks (Bainbridge et al., 2019, 2021b). Participants were healthy native English speakers with normal or corrected-to-normal vision and were compensated for their participation, following the guidelines of the National Institutes of Health (NIH) Institutional Review Board (NCT00001360, 93M-0170). A separate group of 1396 online scorers were recruited from online crowdsourcing task platform Amazon Mechanical Turk (AMT). They acknowledged their participation following the guidelines of the NIH Office of Human Subjects Research Protections (OHSRP) and were also compensated for their participation.

Stimuli

Our goal was to select a set of scene images from highly familiar scene categories, where each exemplar is visually distinctive and unique, yet at the same time representative of its scene category. We thus chose scene stimuli that had a mix of both objects distinctive to an individual image as well as objects that were shared across images of a given category. Prior studies have shown that participants can recall on average about twelve category-unique scene images (Bainbridge et al., 2019). Therefore, we decided to use 8 images, so that participants would be likely to remember more details per image.

An initial AMT experiment was conducted to determine the optimal scene categories for our study. In this experiment, we queried what participants thought to be the most common objects in twelve different common scene categories: *bathroom, beach, campsite, city street, garden, highway, kitchen, living room, playground, park, pool*, and *office*. Online participants were given a scene category name and were instructed to list the 10 most common objects that could likely be found in the scene. They were asked not to list people, animals, sky, walls, floors, or ceiling, and to group plurals under a singularised label (i.e., list "tree" and not "trees"). In all, 72 participants were recruited for this experiment, with 20 participants for each scene category. Participants were able to respond for as many scene categories as they desired, and completed on average 3.33 trials each (SD = 3.32). The participants listed 44.75 unique objects (SD = 7.42) on average per scene category (for a complete list of named objects see the Supplementary Materials). From these data, we selected five scene categories: *kitchen*, *living room*, *office*, *city street*, and *playground*, optimised to have the most consistency across participants in frequently-named objects within scene category, and least overlap in named objects between categories. This way, memory interference would be high within categories, but low across categories.

For each scene category, we identified eight images (from publicly available images on Google Image Search) that each contained the four most-named objects for that scene category (Figure 1). For example, *fridge*, *sink*, *microwave*, and *oven* were the most frequent objects in a kitchen as named by the online participants, and all kitchen images in our stimulus set contain those four objects. This was to ensure that these images were representative of their scene category. It also allowed us to assess participants' memory for objects similar in visual features and semantic meaning within a scene category. Objects that existed in all exemplars of a scene category we refer to as *shared objects*. The images were also selected to have several unique visual features that would differentiate them as clearly different scenes, with unique details, layouts, and objects (for example, in the kitchens in Figure 1: the *stools*, the *door*, the *tree*, and the *tea kettle*). Objects that only existed in a single exemplar of a scene category we refer to as *distinct objects*. With these stimuli, our goal was to emulate the natural variation in these scene categories in the real world (e.g., remembering the kitchens of multiple open houses), through balancing these both shared and distinctive traits.

During the main experiment, each participant viewed eight images (Figure 1), where four images were selected to be the only image presented for that scene category (*category-unique*) and four images were different exemplars from the same scene category (*category-repeated*). Which categories were unique or repeated was pseudorandomly counterbalanced across participants, so each of the 20 images was ultimately seen by 12 participants, and were category-unique for some participants, and category-repeated for others. Importantly, this means that

Figure 1. *Stimulus Selection*. All stimuli in the main experiment, with a sample experimental design for one participant. This participant viewed all four playground images (*category-repeated condition*), and saw one image each of an office, living room, kitchen, and city street (*category-unique condition*). On the right are examples of objects that were shared across all images for a given scene category, and objects that were distinct to one image per scene category.

our analyses compared the same images to themselves, but when presented in these two conditions (category-unique and category-repeated). For each scene category, four previously unseen foil images with similar visual characteristics to the original images were used in the recognition phase of the experiment.

For the object-based analyses, we segmented the outlines of the objects in the images using LabelMe, an online object annotation tool (Russell, Torralba, Murphy, & Freeman, 2008). Annotations were created before any experiments were run, and objects were generally defined as isolated, nameable, visually distinct items. Each image contained on average 52.6 objects (SD = 19.7, min = 19, max = 97).

Experimental procedures

Each participant completed a study phase, distractor phase, recall phase, recognition phase, and review phase of the experiment in one session in the same order (Figure 2(A)).

The first phase was a study phase, in which participants viewed eight images (four category-unique, four category-repeated) for 10 seconds each while their eye movements were being recorded. Participants were instructed to study each image carefully, as they would later be tested on their memory. However, they were not informed how their memory would be tested, and had no awareness of a future drawing task and so could not use honed strategies to maximise drawing ability. Images were presented at their full resolution (1200 × 800 pixel) on a 24-inch screen at a width of approximately 28 degrees of visual angle. The experimental stimuli were presented using the SR Research Experiment Builder software (SR Research, 2010). Participants' head movements were limited by a chin and forehead rest. Eye movements were monitored with a head-mounted EyeLink 1000 Plus eye-tracking device sampling the right eye at 1000 Hz. Between each image, participants were required to fixate a cross displayed to the right of the image to avoid biasing eye movements to the centre of the image.

Participants were next asked to complete a digit span distractor phase which was used to prevent the participant from rehearsing labels for the images in working memory before the recall phase. For this task, participants were asked to memorise a series of numbers of 3–9 digits in length that were presented consecutively on the screen. For each set of digits, they were then asked to verbally repeat the series in order. Each digit was presented for 1 s with a 200 ms interstimulus interval. Participants completed 21 trials of the distractor task, for an average 6-minute delay (SD = 1.44 minutes) between study and recall phases of the experiment.

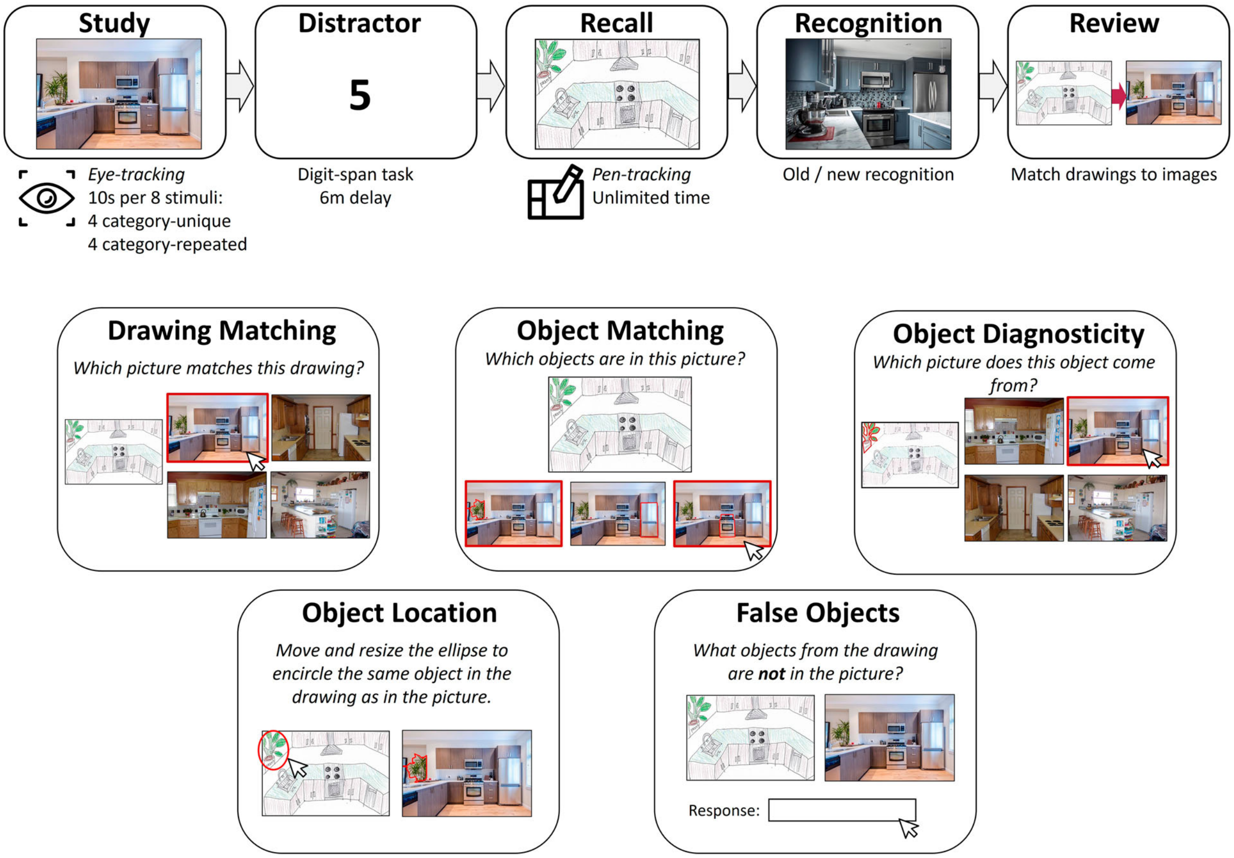

Figure 2. *Experimental Methodology*. (A) Participants engaged in five different phases for the experiment: (1) a study phase of the eight stimuli with eye-tracking, (2) a distractor digit-span phase, (3) a drawing free recall phase with pen-tracking, (4) an old/new recognition phase, and (5) a review phase where participants matched their drawings to the studied images. (B) Drawings made by in-lab participants were then scored online by crowd-sourced scorers. Five online scoring experiments were conducted: (1) Drawing Matching, where scorers matched each drawing to an image, (2) Object Matching, where scorers judged which objects were in the drawings, (3) Object Diagnosticity, where scorers matched each drawn object to an image, (4) Object Location, where scorers indicated the size and location of each drawn object, and (5) False Objects, where scorers described any objects in the drawings that were not in their corresponding images.

In the recall phase, participants were asked to draw as many of the 8 images as they could remember in as much detail as possible. They were told they could draw the images in any order, could take as long as they liked per drawing, and could label any details in their drawing that they felt were unclear. Participants drew each scene on a separate sheet of paper with a rectangular outline identical in size to the original images, and their pen strokes were tracked by a Wacom Paper Pro tablet, which recorded the pen strokes digitally in real time while the participants made inked drawings on paper. After completing the image in black ink, they were given coloured pencils to add any colour detail that they could remember from the scene but were asked not to include a colour unless they specifically recalled it from the image. These coloured pencils could not be tracked by the tablet. Participants took on average 28.3 minutes (SD = 7.08 minutes) to complete their drawings in the recall phase.

In the recognition phase, participants were presented with a series of images and asked to make a judgment indicating whether that image was old (from the study phase) or new. The images included the eight images seen in the study phase intermixed with eight category-matched foils, for 16 images total. For example, if a participant studied four images of playgrounds, and one image of each of the other scene categories (city street, kitchen, living room, office), then their recognition image set would include four foil playground images, and one foil image for each of the other categories. The recognition task was self-timed; each image remained on the screen until participants made a response.

Finally, for the review phase, participants were shown the images they saw during the study phase. They were asked to match each of their drawings made from memory with the corresponding image they saw during the study phase. These data were used to assess which image they were intending to capture with each drawing, creating a correspondence between drawings and images.

Online scoring procedures

The thirty in-lab participants drew 213 scenes from memory. Participants could choose to label any objects in their drawings that they felt were unclear and could add any colour detail that they could remember from the scene. Ninety-four percent of all scene drawings included colour, and 59% of drawings included at least one label. These drawings were scanned and uploaded to AMT, where scorers were asked to rate them on several properties. Drawings that were not matched to an original image by the participant during the review phase (14 out of 213 scenes, or 6% of drawings) were not scored, leaving 199 drawings for these analyses. Five different measures were collected for each drawing (Figure 2(B)). For all tasks, AMT scorers could participate in as many trials as they wanted. In total, 1764 unique AMT scorers participated across all online experiments.

Drawing Matching. This task was used to quantify how well a drawing matched its specific image, providing a rating of diagnosticity of that drawing. Scorers ($N = 210$) were shown a drawing and were asked to match it to one of the four images from the same scene category. If the drawing was of an image shown in the category-unique condition, scorers were shown the one image the participant viewed in the study phase along with the other three same-category images that the participant did not study. If the drawing was of an image in the category-repeated condition, the scorer was shown all four images that the participant saw during the study phase, with only one image being a match. Twenty-four unique scorers were asked to match each of the 199 drawings for a total of 4776 judgements. Scorers were asked to match five separate drawings in each trial and completed 4.59 (SD = 6.05) trials on average. A match was determined to be correct if the image that the majority of scorers selected as being the closest to the drawing was the same as the label given by the participant in the review phase. If the majority of scorers are unable to match a drawing done from memory to its original image, that drawing can be seen as a less accurate representation of the original image.

Object Matching. This task was used to identify which objects were in a given drawing. Scorers ($N = 368$) were shown a drawing and a series of five objects from its corresponding image (determined by the participant review). The five objects in the image were indicated by separate images with each object outlined in red, using the outline generated by the LabelMe annotations. Online scorers were asked to select which of those five objects (if any) were in the drawing. Scores were collected from five scorers per object, and an object was determined to be in a drawing if at least three of the five scorers identified it. This task was then conducted for all objects within each image.

Object Diagnosticity. This task was used to identify the sources of the objects in each of the drawings. Scorers ($N = 645$) were shown a drawing with an object outlined in red and were asked which one of the four images from the same scene category it came from. The scorers were warned that sometimes the overall drawing will look the same as one image, but the object will look like an object in another image, and in that case, they should pick the image where they found the object. They were also given the option to say that the object came from none of the shown images. Scorers were asked to rate five different objects per trial, and each object was also rated by five scorers. Objects were determined to be correctly matched to their studied image if a majority of scorers selected the object as being from the correct image (determined by the participant's label in the review phase). This task allows us to quantify the diagnosticity or specificity of objects; are they clearly identifiable as belonging to a given image, or are they ambiguous?

Object Location. This task was used to assess spatial accuracy of the memory drawings. Scorers ($N = 496$) were asked to indicate the size and location of each object present in a drawing. Scorers were shown a drawing and its corresponding image with an object outlined in red. They were asked to place and resize an ellipse around where that object was located in the drawing. Five scorers were asked to complete this task for each object, and the final object location and size was calculated as the median centroid and radii of the ellipse across scorers.

False Objects. This task was used to assess false intrusions in the drawings beyond the objects present in the original images. Scorers ($N = 139$) viewed each drawing and its corresponding image and were asked to write down the names of any objects in the drawing that were not in the photograph. Nine AMT scorers rated each drawing, and any objects listed by at least five scorers were counted as false alarms.

Fixation and pen-tracking analyses

Eye movements were monitored with a head-mounted EyeLink 1000 Plus eye-tracking device during the study phase of the experiment. One participant's eye movements were recorded but not saved due to an experimenter error, resulting in $N = 29$ for these analyses. We defined the time of first fixation for each object as the time in seconds after trial started that each object was fixated for the first time. The total fixation duration for each object was computed as the total time that an object was fixated across all fixations to the object in a trial.

To capture similar metrics during recall, we also tracked the start time and total time it took participants to draw each object from memory based on the pen strokes captured by the Wacom Paper Pro tablet. A naive scorer watched the video of pen strokes created by the tablet and recorded the start and end time of pen strokes for each object. Total amount of time spent drawing the object was calculated as the difference between the end time and the start time.

Results

Utilising the participant drawings, online scoring experiments, fixation data, and pen-tracking data, we conducted an in-depth comparison of the level of detail present in memory for category-unique versus category-repeated images along six different metrics. First, we compared overall recall and recognition performance for the entire image. Second, online scorers judged the degree to which drawings were diagnostic of their original images. Third, we examined object recall performance by calculating the proportion of objects recalled and their diagnosticity. Fourth, we examined the propensity for false memories in these two stimulus conditions. Fifth, we examined the accuracy of spatial location and size memory for objects. Finally, we compared fixation and pen-tracking duration and speed in the two conditions. To preview our findings, we found no significant difference in the overall rate of recall between category-repeated and category-unique exemplars, nor the number of false memories. However, we found a significant decrease in the diagnosticity and the overall object detail in drawings made from competing category exemplars. We found no difference in spatial accuracy in drawings for scenes from unique versus repeated categories, suggesting that object feature information, rather than spatial information, may be more susceptible to memory interference when tasked to recall multiple competing exemplars.

Category-repeated exemplars are less likely to be recognised, but not less likely to be recalled

Participants' recall performance was judged based on the number of scenes they successfully drew from memory. On average, participants recalled 6.63 scenes (SD = 1.06, min = 5, max = 8) out of 8 studied, for an average hit rate of 0.83 (SD = 0.18). Participants on average recalled 80.8% (SD = 19.3%) of the images studied in the category-unique condition, and 84.2% (SD = 16.7%) of the images studied in the category-repeated condition, although this difference in performance was not significant (non-parametric Wilcoxon signed-rank test, WSRT: Z = 0.73, $p = .446$). Across participants, there was also no significant difference in the proportion of category-unique images versus category-repeated images that were successfully recalled (Chi-squared Test: $\chi^2 = 0.46$, $p = .497$). Finally, there was no significant difference for whether a given image was recalled more frequently across participants when it was category-repeated than when it was category-unique (Z = 1.83, $p = .095$). Overall, there were no clear differences in recall performance between conditions.

Participants' memory for the scenes was also tested with an old/new recognition task. Participants had high recognition performance for images studied in both conditions, recognising 96.7% (SD = 8.6%) of the category-unique images, while recognising 90% (SD = 15.5%) of the category-repeated images on average. Category-unique images were recognised at a significantly higher rate than category-repeated images (WRST: Z = 2.20, $p = .031$). Participants had a lower rate of false alarms for category-unique images (M = 5.8%, SD = 15.7%) than category-repeated images (M = 8.3%, SD = 15.2%), although this difference was not significant (WRST: Z = 0.77, $p = .463$). We also examined the Lure Discrimination index (the corrected recognition rate minus the hit rate), and found that category-unique images also caused more successful correct rejections than category-repeated images (Unique M: 90.8%, SD: 18%, Repeated M: 81.7%, SD: 22.7%; WRST Z = 2.02, $p = .036$). Looking at recognition performance at an image level also revealed a significantly higher recognition rate when a given image was category-unique versus category-repeated (WRST: Z = 2.80, $p = 3.50 \times 10^{-3}$). A Chi-squared test also showed a significantly higher recognition rate for category-unique versus category-repeated items across participants ($\chi^2 = 4.29$, $p = .038$). Collectively, these data show that participants were better at recognising category-unique images than category-repeated images.

These results could suggest an interaction in which unique, distinctive stimuli may be easier to recognise, but not necessarily easier to recall. A two-factor repeated measures ANOVA (memory type: recall/recognition × condition: repeated/unique) comparing logit-transformed memory performance for a given image observed a significant effect of memory type ($F(1,76) = 12.54$, $p = 6.84 \times 10^{-4}$, $\eta_p^2 = 0.142$), no significant effect of condition ($F(1,76) = 2.05$, $p = .156$), and a significant interaction effect ($F(1,76) = 6.00$, $p = .017$). These results suggest an interaction of these numerically opposite effects for recall and recognition, where an image is better recognised when it is category-unique, but recalled similarly to category-repeated images.

Previously, we have failed to observe a correlation in performance between recognition and recall; images that are successfully recognised are not necessarily those that are successfully recalled (Bainbridge et al., 2019). We tested this same question in the current study. For individual images, we found no significant correlation between the percent of participants who recalled that image and the percent who recognised it ($\rho = -0.11$, $p = .491$), suggesting a dissociation between recall and recognition at the image level.

Overall, these results show that participants had relatively good memory for both types of images (category-repeated and category-unique). While participants were better at recognising unique images, they showed a slight tendency to recall more repeated images. A next key question, then, is whether these drawings of category-repeated images show any differences in their details when compared to those of category-unique images?

Drawings of category-repeated scenes are less representative of the original image

The prior analyses relied on participants to judge which images they were recalling with a given drawing. However, these judgments could be subjective, malleable, and inflated by a desire to report high memory performance. Therefore, we assessed how well independent scorers could match a memory drawing to the correct image. This quantifies how diagnostic a given drawing is of an image, as a less specific drawing will be matched to other images of the same category.

To assess this diagnosticity, online scorers were shown a drawing and were asked to match it to one of four images from the scene category (Figures 3 and 4). Category-repeated drawings were successfully matched to their original images 64.9% of the time (SD = 11.0%). Notably, participants' drawings made from the category-unique condition were correctly matched at a significantly higher rate (M = 71.2%, SD = 8.7%) than those of category-repeated exemplar images (WRST: Z = 2.48, $p = .014$). These results suggest that a drawing of an exemplar is significantly more diagnostic when it is drawn as the only exemplar in its category (unique) than when it was one of multiple exemplars studied (repeated).

Category-repeated drawings have fewer objects and less diagnostic objects than category-unique drawings

We next looked at the amount of object detail included in the drawings by condition. Each image contained on average 52.6 segmented objects (SD = 19.7, min = 19, max = 97). Online scorers judged the drawings for which objects were included from their original images. Drawings of the category-unique images contained a significantly higher proportion of objects compared to drawings made from category-repeated images (Unique: M = 32.0%, SD = 8.2%; Repeated: M = 27.7%, SD = 7.2%; WRST: Z = 2.54, $p = .011$). This indicates that participants recalled more objects for category-unique images than category-repeated images (Figure 4).

We also looked to see whether recall performance differed for distinctive objects (which only existed in a single image) versus shared objects (which existed in all images of a category). We conducted a 2-way repeated-measures ANOVA comparing the logit-transformed proportion of objects recalled for shared versus distinct objects, within category-unique versus category-repeated drawings. We observed that shared objects were significantly better recalled than distinctive objects ($F(1,116) = 8.24$, $p = .008$; $\eta_p^2 = 0.22$) and more of both objects were drawn in category-unique scenes than category-repeated scenes ($F(1,116) = 4.17$, $p = .050$, $\eta_p^2 = 0.13$), although there was no significant interaction ($F(1,116) = 1.69$, $p = .204$). These results suggest that shared objects for a scene category were more commonly recalled, regardless of how many exemplars of that scene category were studied.

We also tested the diagnosticity of the objects drawn in both conditions – how specific were drawn objects' appearances to the image they were drawn from? Objects from category-unique drawings were correctly matched by online scorers to their original image 65.7% of the time (SD = 17.7%) and were matched at a significantly higher rate than objects from category-repeated drawings (M = 51.9%, SD = 25.8%; WRST: Z = 2.88, $p = 2.70 \times 10^{-3}$). Thus, with more exemplars from a category, objects were drawn with less diagnostic detail or were pulled from other images (Figure 4). A 2-way repeated-measures ANOVA of logit-transformed matching success for shared and distinct objects revealed no effect of diagnosticity related to object type ($F(1,101) = 1.14$, $p = .289$), nor was there an interaction with scene condition ($F(1,101) = 0.09$, $p = .769$). Thus, this diminished object diagnosticity occurred across all object types.

Object intrusions and false memories in category-repeated and category-unique drawings

Given the diminished object detail in category-repeated drawings, we also investigated whether there was a difference in false memories for objects by condition (Figure 5). Online scorers judged for each drawing which objects were present in the drawing that were not in the original image. Participants on average drew 0.83 additional objects (SD = 1.09) across their category-unique drawings, and 1.17 additional objects (SD = 1.53) in their category-repeated drawings. While these false objects were numerically higher for category-repeated drawings, this difference was not significant (WSRT: Z = 1.05, $p = .297$), and the overall propensity to draw false additional objects was very low. Thus, while objects were less diagnostic in category-repeated images, there was not a sharp increase in falsely remembered information.

Participants also had false memories of entire scenes (Figure 5). Across the two conditions, participants drew 199 drawings that they were able to match to an original image. However, participants also drew an additional 14 drawings that could not be matched back to an original image in the review phase. Participants had on average 0.47 false memories of an entire scene (SD = 0.69, MIN = 0, MAX = 2). During the review phase, participants identified eight of these drawings as repeats of a single scene that they falsely remembered as being multiple scenes. The remaining six were not related to any of the studied images, including scenes such as a porch, swimming pool, and bedroom. Six of the repeated drawings were from the repeated category, while two were from a unique category.

Equivalent spatial accuracy in memory for both conditions

In addition to the visual detail for objects, we also assessed the accuracy of the size and location of drawn objects, by asking online scorers to locate objects by positioning and

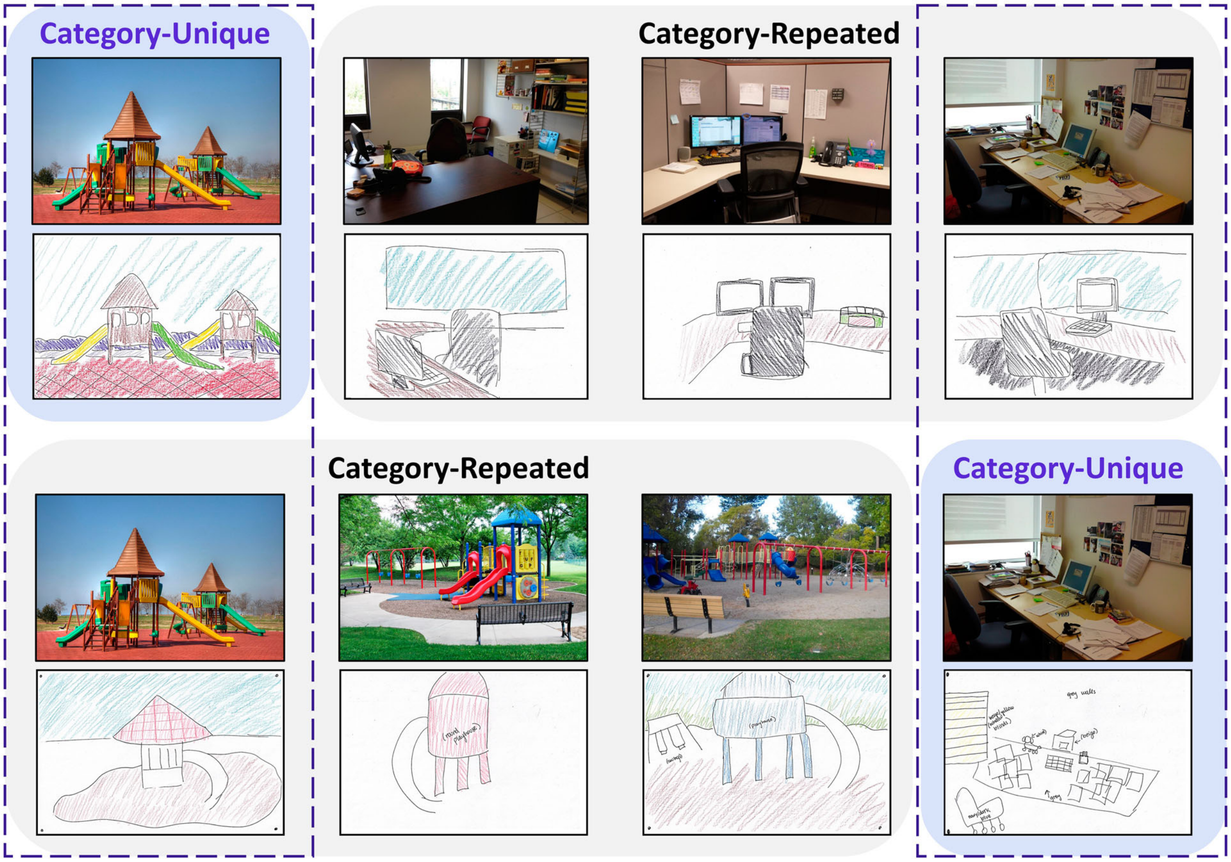

Figure 3. *Example drawings.* Example drawings from two participants (one participant per row) with counter-balanced stimulus conditions. For the top participant, a playground image was category-unique, while office images were category-repeated. Conversely, for the bottom participant, an office image was category-unique, while playground images were category-repeated. The category-unique drawings tend to include more details and image-diagnostic information than their category-repeated counterparts. All drawings can be seen in an online repository on the Open Science Framework (https://osf.io/syvjr/.).

resizing an ellipse. Overall there was relatively low error in X and Y location for drawn objects for both conditions (X-direction: Unique M = 17.4% of image width, Repeated M = 16.2%, Y-direction: Unique M = 12.5% of image height, Repeated M = 14.8%), and we found the two conditions did not differ significantly (X-direction: WRST Z = 0.77, p = .443; Y-direction: WRST Z = 0.94, p = .347). We also found low errors of object size in both conditions (Width: Unique M = 6.2%, Repeated M = 6.1%; Height: Unique M = 2.7%, Repeated M = 3.0%), and their difference in accuracy was not significant for either measure of width (WRST Z = 0.487, p = .627) or height (WRST Z = 0.55, p = .581). These results indicate that spatial accuracy was relatively high in memory drawings and did not differ by condition.

Eye movement differences during the study phase

Given the differences in recalled object detail between the category-unique images and the category-repeated images, we were interested to see whether there were differences in viewing behaviour on these images. We also examined fixations on distinct objects (those that only existed in one scene exemplar) versus shared objects (those that existed in all exemplars of a category) in the scenes.

First, we examined whether there was a significant difference in participant's eye movement behaviour based on whether they were studying a unique or repeated exemplar image. For these analyses, we excluded the first exemplar seen in the category-repeated condition for each participant to account for the fact that this first image in the category would be novel to the participant and not a repetition. During the 10s study period for each image, participants made on average 40.7 (SD = 18.3) fixations on category-unique images, and 39.3 (SD = 9.4) fixations on category-repeated images. We fit the data to a linear mixed effects (LME) model with number of fixations as the dependent variable, condition (category-unique vs. category-repeated) as the independent variable, and with nested random intercepts for image within subject. We found no significant effect of condition on number of fixations to a scene (b = 1.40, p = .325). We also fit a LME model with fixation duration as the outcome variable, and the same dependent and random effects as above, and found no significant effect of condition (Unique: M = 271.65 ms, SD = 155.76, Repeated: M =

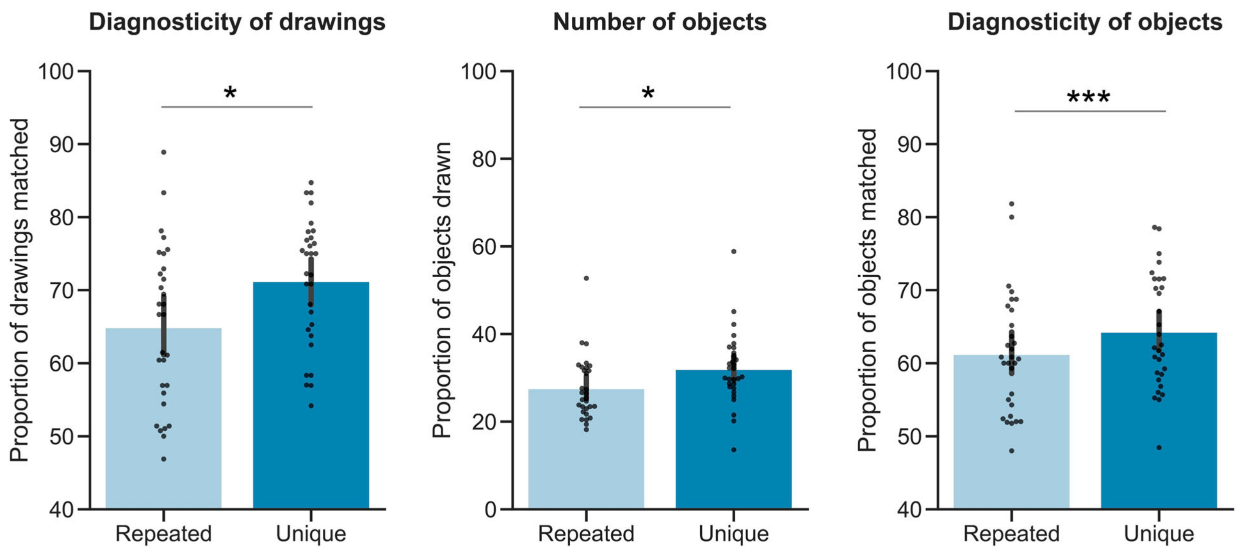

Figure 4. *Measures of image and object memory*. Each dot indicates each of the participants, and error bars indicate standard error of the mean. (Left) Category-unique drawings were matched by independent scorers to the original images at a significantly higher rate than category-repeated exemplars ($p = .014$). (Centre) Participants recalled more objects in category-unique drawings than category-repeated drawings ($p = .011$). (Right) Objects from category-unique drawings were correctly matched by online scorers to the original images at a significantly higher rate than objects from category-repeated exemplars ($p = 2.70 \times 10^{-3}$).

264.99, SD = 169.2; b = 5.36, $p = .158$). We also looked at the overall percent of the scene that was foveated by condition. Areas of the scene that were foveated were calculated by placing a uniform filter with a 1 degree of visual angle radius centred on each fixation. For each scene, we applied this filter to all fixations for each participant, and then compared the overall percent of the scene that was foveated by condition. On average, participants fixated 23.2% (SD = 0.05) of the scene in the category-unique condition, and 22.9% (SD = 0.05%) of the scene in the category-repeated condition. A LME with percent foveated as the outcome variable, and the same fixed and random effects as above, found no significant fixed effect of condition (b = 4×10^{-3}, $p = .469$). These results indicate no differences in overall viewing behaviour when studying a single image from a scene category, versus multiple images from that category.

Next, we investigated whether fixation behaviour differed based on whether objects were shared across scenes, or distinctive to a single scene. For these analyses,

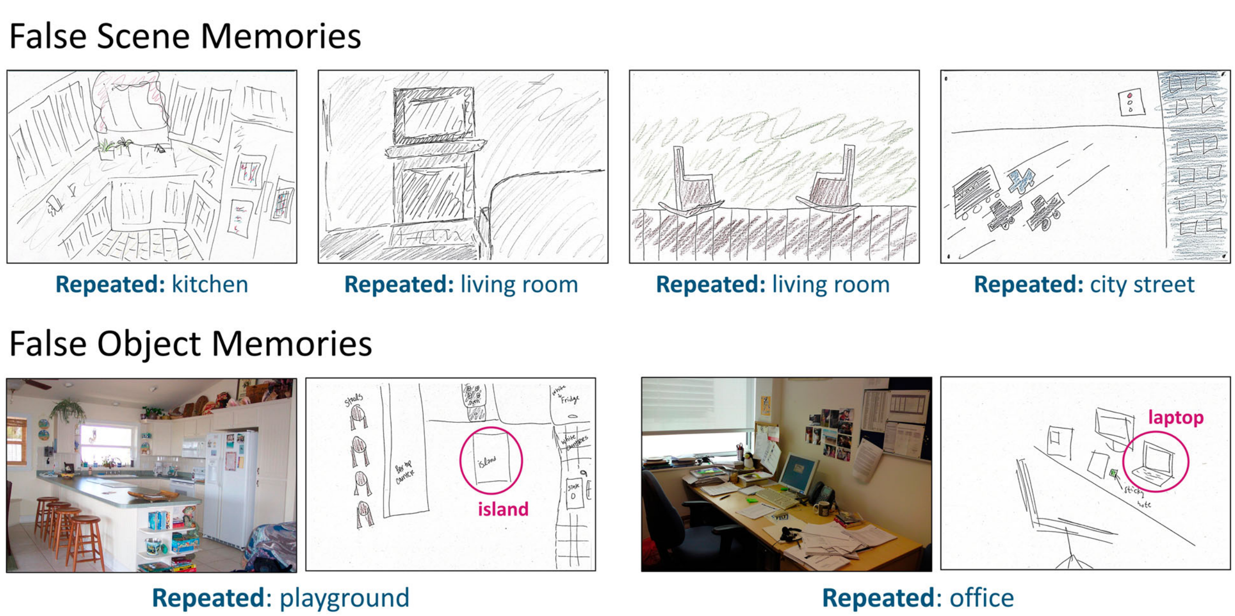

Figure 5. *Examples of false memories*. (Top) Examples of entire scenes that were drawn as false memories, with the repeated category for that participant noted below each drawing. (Bottom) Examples of objects that were drawn as false memories in an otherwise correct drawing, with the repeated category for that participant noted below.

we again removed the first category-repeated trial because during the encoding period the first exemplar would be experienced as the first for a category. A two-factor within-subject repeated-measures ANOVA on average fixation durations by object type in a given drawing found no difference in fixation behaviour between distinct and shared objects ($F(1,112) = 1.72$, $p = .200$). There was also no significant difference in fixation duration by scene condition ($F(1,112) = 0.56$, $p = .459$), nor was there a significant interaction ($F(1,112) = 0.26$, $p = .612$). Thus, participants' fixation behaviour was not modulated by the number of exemplars of that scene type. A repeated-measures ANOVA on the time to first fixation also showed no differences by object condition ($F(1,112) = 0.98$, $p = .324$), nor scene condition ($F(1,112) = 0.0003$, $p = .986$), nor an interaction ($F(1,112) = 0.07$, $p = .798$). These results suggest participants do not necessarily fixate distinctive versus shared objects earlier when exploring a scene.

Temporal order of object recall in scenes

During the recall phase of the experiment, we recorded real-time pen movement patterns as participants drew the scenes from memory. Nine participants did not draw objects in some conditions (eight did not draw distinct objects in category-repeated images, one did not draw distinct objects in category-unique images), so they were removed from these within-subject analyses. Of the participants who drew objects from all conditions, we looked at how long they took to draw each object, and the order in which they drew objects. A two-factor within-subject repeated-measures ANOVA found no significant difference in time spent drawing distinct versus shared objects ($F(1,80) = 2.56$, $p = .125$). There was also no difference in drawing time by scene condition ($F(1,80) = 0.38$, $p = .546$), nor was there a significant interaction ($F(1,80) = 1.09$, $p = .310$). This finding suggests that amount of time spent drawing detail was unaffected by the number of exemplars shown per category. A two-factor within-subject repeated-measures ANOVA similarly found no significant differences in the time of first pen stroke for distinct versus shared objects ($F(1, 80) = 2.89$, $p = .105$), nor was there a difference by scene condition ($F(1,80) = 0.81$, $p = .379$) nor an interaction ($F(1, 80) = 2.68$, $p = .118$).

Finally, we examined whether there were correlations between fixation behaviour and pen stroke behaviour. For each participant, we took a Spearman rank correlation between their fixation durations and drawing durations for objects. When testing these rank correlations (after a Fisher Z-transformation) versus a null correlation of 0, we observed no significant evidence for a correlation between fixation and drawing duration across participants (Mean $\rho = 0.08$, $t(28) = 1.07$, $p = .294$). We also saw no significant correlation of fixation order and drawing order (M $\rho = 0.07$, $t(28) = 1.04$, $p = .306$).

Discussion

Using the lens of drawing, this study provides an in-depth investigation of how competing scenes influence the detail in a visual memory representation. When participants were presented with multiple images from a scene category, they were worse at recognising these images amongst same-category foils, although they were slightly more likely to recall these images. This increased recall may reflect a greater memory for the category as a whole as these recollections of category-repeated images contained decreased levels of detail, with fewer objects that were also less specific to the image. Participants also showed a numerically higher tendency to draw more false objects and entire false images from the repeated image category, but false alarms were fairly rare and these differences were non-significant. In spite of the diminished details in memory, some aspects of the memory remained intact regardless of the number of exemplars shown per category. Specifically, spatial memory for objects was equally accurate regardless of number of scenes studied from a category. Further, fixation patterns during study and pen movements during recall were similar between scene conditions, indicating that divergent memory representations did not result from differences in study behaviour, or different strategies during recall.

These findings suggest that competition in memory does result in diminished detail for specific memories. Further, these results suggest that categorical information may play an important role in scaffolding accurate visual memories; even for these visually distinct exemplars, they suffer interference from same-category items. This runs counter to work suggesting that our memories for individual scenes are largely isolated, visual representations that are relatively unaffected by categorical interference (e.g., Konkle et al., 2010a). Our results instead align with classic findings demonstrating an influence of semantic and categorical relatedness on increased false memories (Hintzman, 1988; Strack & Bless, 1994; Roediger & McDermott, 1995; Koutstaal & Schacter, 1997; Seamon et al., 2000).

This distinction may be due in part to the difference in recognition versus recall memory. In line with our finding that category-unique scenes were better recognised by participants and easier to correctly reject, earlier work has found that the more conceptually distinct an image is, the easier it is to recognise (Konkle et al., 2010b). Similarly, while recall memory may benefit from grouping of scenes under a similar semantic label, helping participants to remember that they had seen four scenes from a particular category, this grouping of memories may make it harder to hold onto a cohesive representation for a single image and its specific details. This may have led participants to have less confidence in what specific scenes they had seen, leading to a diminished ability to correctly recognise images from the category-repeated category. If grouping in recall memory led to a diminished ability to encode these images as isolated collections of objects, we would expect drawings for category-repeated scenes to become a combined or averaged representation of the category as a whole, with less detail specific to any single image. Our subsequent analyses on the recall drawings found that category-repeated drawings were less diagnostic, containing fewer, and less distinguishable objects, suggesting that these recall memories were less precise than those from category-unique images.

Surprisingly, we did not observe any changes in spatial accuracy in memory between conditions. Our original hypothesis was that memories for scenes would converge to an average (either of the scene category, or of the studied set), and so shared objects would migrate to similar locations. However, this did not appear to be the case – spatial memory remained high and unaffected by repeated exemplars from a category. This echoes prior work using a similar drawing task investigating *aphantasics* individuals with no visual imagery – which found that while aphantasic participants showed a decrease in object detail in memory, they showed no impairment in ability to recall spatial information (Bainbridge et al., 2021a). Their decreased object detail memory was attributed to their lack of precise visual mnemonic representations; however, they were able to draw some information by combining semantic representations (e.g., a mental list of objects) with spatial information. In the current study, we observe a complementary effect – when semantic/categorical information is disrupted, there is still a decrease in visual detail in the memory. However, across both studies, spatial memory remained intact. In another study, scenes with inconsistent object-scene semantics (e.g., a mailbox in a bedroom) also showed diminished object detail and loosened binding of the object to its scene but showed no disruptions in spatial memory (Bainbridge et al., 2021b). The results from this paradigm are also in line with work finding two dissociable and parallel networks in human hippocampus responsible for pattern separation of object and spatial memories (Reagh & Yassa, 2014; Reagh et al., 2016). The combination of these results supports the notion of separate processes governing object versus spatial information (e.g., Farah & Hammond, 1988; Ungerleider & Haxby, 1994; Kravitz et al., 2011; Staresina et al., 2011).

In order to understand what aspects of a scene showed the greatest loss, we also looked at effects for two different types of objects – objects shared across images of a scene category, and objects that were distinct to a specific image. We expected these two object types to show large differences by scene condition. For example, repeated viewing of shared objects across scenes may result in increased recall but decreased diagnosticity of shared objects versus distinct objects in the category-repeated condition. However, we did not observe significant differences between distinct and shared objects in terms of diagnosticity, fixation behaviour or drawing behaviour. We did observe that shared objects were recalled more often than distinct objects, but this effect was not modulated by number of exemplars in a category. Thus, against our expectations, we observed no interactions for any measure between object type and scene condition. Increasing exemplars from a category did not differentially affect these two types of objects. Instead, the patterns we have observed in our study were present uniformly across all objects.

The current study provides a rich characterisation of how visual memories change with increased competition. However, there are some limitations of the current study that should be addressed in future work. Here we broadly examined the interference resulting from multiple category exemplars, following the tradition of studies in the visual object domain (Strack & Bless, 1994; Koutstaal & Schacter, 1997; Seamon et al., 2000). However, many different features define the members of a scene category; they are semantically similar, contain similar objects, have similar visual details, share a broad visual gist (Oliva & Torralba, 2006), and share similar spatial properties. In this study, we are unable to distinguish whether the source of interference is purely semantic, purely visual, or a mix of the two.

In prior work, we have found that both diminished visual information but intact semantics (Bainbridge et al., 2021a), as well as disrupted semantics with unchanged visual information (Bainbridge et al., 2021b) result in lowered object memory. This suggests that both types of information play a key role in forming accurate, detailed memories. However, when looking at real-world visual images, it may be difficult to disentangle the contributions of these two information types, given that these visual and semantic features are often intertwined; objects for specific spaces are often designed similarly. For example, dining room furniture may tend to be rectilinear and darker in colour, while bathroom appliances may tend to be curved and white. It is also possible that some exemplars in our study were more similar to each other than others, either visually, through properties like the number of objects, number of distinct objects, or object aesthetics, or semantically. These levels of exemplar similarity could vary between scene categories, and could have led to better memory for one scene category over another. While this would not have impacted the main results of our study, which primarily looked at the differences in memory when a scene was studied in either the category-repeated or category-unique condition, more work is needed to understand which object and scene properties can give rise to false memories when scenes look highly similar. A future study could use highly quantified or constructed stimuli, such as computer-generated scenes where semantic similarity and/or visual similarity across exemplars is varied in an intentional way. Indeed, recent work has utilised convolutional neural networks to calculate similarity across scenes and identify images from a large scene space that can create false memories in a scene image DRM paradigm (Děchtěrenko et al., 2021). Additionally, while we did not directly measure whether participants were able to correctly recall the colour of objects in their drawings, false colour memories have been reported in other studies (Lupyan, 2015; Witzel, 2016; Valenti & Firestone, 2019) and the accuracy of colour recall memory under competition is another interesting question that could be explored in future work.

In sum, these results show that competing memories result in a diminished level of detail within a memory. This diminished detail specifically manifests as lowered object detail and diagnosticity, although spatial information remains intact. Thus, when searching for that new home – it's worth relying on tools beyond just one's memory to keep those kitchens straight.

Acknowledgements

We thank Anna Corriveau for her help digitising the drawings from the study, Adam Dickter for his help with the eye tracker system, and Zoe Loh for scoring the pen-tracking videos. All drawings are made publicly available on the Open Science Framework (https://osf.io/syvjr/.).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the Intramural Research Program of the National Institutes of Health (ZIA-MH-002909), under National Institute of Mental Health Clinical Study Protocol 93-M-1070 (NCT00001360).

References

Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2019). Drawings of real-world scenes during free recall revealed detailed object and spatial information in memory. *Nature Communications*, *10*(1), 5. https://doi.org/10.1038/s41467-018-07830-6

Bainbridge, W. A., Kwok, W. Y., & Baker, C. I. (2021b). Disrupted object-scene semantics boost scene recall but diminish object recall in drawings from memory. *Memory & Cognition*, *49*, 1568–1582. https://doi.org/10.3758/s13421-021-01180-3.

Bainbridge, W. A., Pounder, Z., Eardley, A. F., & Baker, C. I. (2021a). Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. *Cortex*, *135*, 159–172. https://doi.org/10.1016/j.cortex.2020.11.014.

Baioui, A., Ambach, W., Walter, B., & Vaitl, D. (2012). Psychophysiology of false memories in a Deese-Roediger-McDermott paradigm with visual scenes. *PloS ONE*, *7*(1), e30416. https://doi.org/10.1371/journal.pone.0030416

Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative & Integrative Biology*, *2*(1), 1–3. https://doi.org/10.4161/cib.2.1.7297

Děchtěrenko, F., Lukavský, J., & Štipl, J. (2021). False memories for scenes using the DRM paradigm. *Vision Research*, *178*, 48–59. https://doi.org/10.1016/j.visres.2020.09.009

Farah, M. J., & Hammond, K. M. (1988). Visual and spatial mental imagery: Dissociable systems of representations. *Cognitive Psychology*, *20*(4), 439–462. https://doi.org/10.1016/0010-0285(88)90012-6

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 396–411. https://doi.org/10.1037/0278-7393.31.3.396

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113–136. https://doi.org/10.1037/0096-1523.28.1.113

Israel, L., & Schacter, D. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*(4), 577–581. https://doi.org/10.3758/BF03214352

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558. https://doi.org/10.1037/a0019165

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in old and younger adults. *Journal of Memory and Language*, *37*(4), 555–583. https://doi.org/10.1006/jmla.1997.2529

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230. https://doi.org/10.1038/nrn3008

Lupyan, G. (2015). Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*, *161*, 117–130. https://doi.org/10.1016/j.actpsy.2015.08.006

Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for scenes. *Neuropsychologia*, *36*(6), 513–520. https://doi.org/10.1016/S0028-3932(97)00148-6

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155(B)*, 23–36. https://doi.org/10.1016/S0079-6123(06)55002-2

Otgaar, H., Howe, M. L., Muris, P., & Merckelbach, H. (2019). Associative activation as a mechanism underlying false memory formation. *Clinical Psychological Science*, *7*(2), 191–195. https://doi.org/10.1177/2167702618807189

Otgaar, H., Howe, M. L., Peters, M., Smeets, T., & Moritz, S. (2014). The production of spontaneous false memories across childhood. *Journal of Experimental Child Psychology*, *121*, 28–41. https://doi.org/10.1016/j.jecp.2013.11.019

Reagh, Z. M., Ho, H. D., Leal, S. L., Noche, J. A., Chun, A., Murray, E. A., & Yassa, M. A. (2016). Greater loss of object than spatial mnemonic discrimination in aged adults. *Hippocampus*, *26*(4), 417–422. https://doi.org/10.1002/hipo.22562

Reagh, Z. M., & Yassa, M. A. (2014). Object and spatial mnemonic interference differentially engage lateral and medial entorhinal cortex in humans. *Proceedings of the National Academy of Sciences*, *111*(40), E4264–E4273. https://doi.org/10.1073/pnas.1411250111

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Russell, B C, Torralba, A., Murphy, K. P. (2008). LabelMe: A database and web-based tool for image annotation. *International journal of computer vision*, *77*(1-3), 157–173.

Seamon, J. G., Luo, C. R., Schlegel, S. E., Greene, S. E., & Goldenberg, A. B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory and Language*, *42*(1), 120–146. https://doi.org/10.1006/jmla.1999.2676

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267. https://doi.org/10.1016/S1364-6613(97)01080-2

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. https://doi.org/10.1080/14640747308400340

Staresina, B. P., Duncan, K. D., & Davachi, L. (2011). Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. *Journal of Neuroscience*, *31*(24), 8739–8747. https://doi.org/10.1523/JNEUROSCI.4978-10.2011

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language*, *33*(2), 203–217. https://doi.org/10.1006/jmla.1994.1010

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, *4*(2), 157–165. https://doi.org/10.1016/0959-4388(94)90066-3

Valenti, J. J., & Firestone, C. (2019). Finding the "odd one out": memory color effects and the logic of appearance. *Cognition*, *191*, 103934. https://doi.org/10.1016/j.cognition.2019.04.003

Witzel, C. (2016). An easy way to show memory color effects. *i-Perception*, *7*(5). https://doi.org/10.1177/2041669516663751