Cognition 217 (2021) 104891

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Semantic influence on visual working memory of object identity and location

Ruoyang Hu, Robert A. Jacobs *

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, United States

A R T I C L E I N F O

Keywords:
Visual perception
Visual working memory
Semantic influence
Categorization

A B S T R A C T

Does semantic information—in particular, regularities in category membership across objects—influence visual working memory (VWM) processing? We predict that the answer is “yes”. Four experiments evaluating this prediction are reported. Experimental stimuli were images of real-world objects arranged in either one or two spatial clusters. On *coherent* trials, all objects belonging to a cluster also belonged to the same category. On *incoherent* trials, at least one cluster contained objects from different categories. Experiments using a change-detection paradigm (Experiments 1–3) and an experiment in which participants recalled the locations of objects in a scene (Experiment 4) yielded the same result: participants showed better memory performance on coherent trials than on incoherent trials. Taken as a whole, these experiments provide the best (perhaps only) data to date demonstrating that statistical regularities in semantic category membership improve VWM performance. Because a conventional perspective in cognitive science regards VWM as being sensitive solely to bottom-up visual properties of objects (e.g., shape, color, orientation), our results indicate that cognitive science may need to modify its conceptualization of VWM so that it is closer to “conceptual short-term memory”, a short-term memory store representing current stimuli and their associated concepts (Potter, 1993, 2012).

1. Introduction

Visual working memory (VWM) is typically thought of as a limited-capacity store that represents the visual properties of objects in a scene, such as objects' shapes, colors, orientations, and locations. This information would be useful, for example, to plan actions, including eye movements and hand reaches. To build its representations, it is obvious that VWM must use bottom-up information derived from the pattern of light that falls on an observer's retina. What is less obvious, however, is whether VWM makes use of top-down information such as semantic information.

Consider a scene in which some objects belong to the same category. For example, the scene might contain a coat and a hat, both articles of clothing. Should VWM make use of this shared category structure when building its representations?

Our prediction is that the answer is “yes”, and it is motivated as follows. VWM has limited capacity, and thus to be efficient it should build compressed representations that allow it to minimize the number of memory resources used per to-be-remembered item (Bates & Jacobs, 2020; Brady, Konkle, & Alvarez, 2009; Mathy & Feldman, 2012; Sims, Jacobs, & Knill, 2012; Yoo, Klyszejko, Curtis, & Ma, 2018). However, compressed codes can only be achieved by taking advantage of statistical regularities of items. Such regularities exist at several levels of abstraction, including at the level of groupings of items. Information regarding group regularities is often referred to as “summary statistic” or “gist” information, and there is now ample data indicating that VWM stores summary-statistic information (Brady & Alvarez, 2011, 2015). Critically, summary-statistic information is essential to the design of compact codes. An implication of this statement is that VWM should indeed make use of shared category structure. For example, it can build compact codes representing the identities of a coat and a hat by making use of the fact that these objects are both articles of clothing.

However, there are reasons to believe that our prediction might be incorrect, meaning that the answer is “no”. For instance, compressed codes based on shared category structures might lead to memory errors. If an observer is more likely to confuse objects belonging to the same category (Roediger & McDermott, 1995), then the observer might misremember a blue coat and a red hat as a red coat and a blue hat (transposing which color goes with which article of clothing), or even as a blue coat and a red glove (introducing a new article of clothing that

* Corresponding author.
E-mail addresses: rhu13@ur.rochester.edu (R. Hu), rjacobs@ur.rochester.edu (R.A. Jacobs).

https://doi.org/10.1016/j.cognition.2021.104891
Received 17 September 2020; Received in revised form 23 August 2021; Accepted 24 August 2021

was not in the original scene).

To date, there have been few studies of the influence of semantic information on VWM. Brady, Störmer, and Alvarez (2016) reported that long encoding times aided VWM performance for real-world objects but not for simple stimuli. Asp, Störmer, and Brady (2021) found improved VWM performance when visually ambiguous stimuli were perceived as meaningful. Conci, Kreyenmeier, Kröll, Spiech, and Müller (2021) demonstrated better VWM performance when color-shape combinations formed meaningful stimuli (e.g., real flags of European countries) than when they formed meaningless stimuli ("fake" flags). Liu et al. (2021) interpreted neural (intracranial electroencephalography; iEEG) data as revealing evidence for visual and semantic influences on VWM. In contrast to these studies, however, Luu and Stocker (2021) argued that top-down category information does not influence VWM, though it does influence subsequent sensory recall and inference processes.

To help clarify the currently sparse and conflicting scientific literature, we report the results from four experiments manipulating the category structure of objects in a scene. Stimuli were images of real-world objects arranged in either one or two spatial clusters. On *coherent* trials, all objects belonging to a cluster also belonged to the same category. On *incoherent* trials, at least one cluster contained objects from different categories. Experiments 1–3 used a change-detection paradigm, a common experimental paradigm for studying VWM. These experiments revealed that participants showed better performance on coherent trials than on incoherent trials, indicating a semantic category influence on VWM representations or operations. Experiment 4 asked participants to recall the locations of objects in a scene. It also found that participants performed better on semantically coherent trials. Taken as a whole, these experiments provide the best (perhaps only) data to date demonstrating that statistical regularities in semantic category membership improve VWM performance.

2. Experiment 1

Experiment 1 used a change-detection paradigm. This paradigm has become perhaps the most common experimental procedure in the scientific literature over the past few decades for probing VWM. Indeed, the literature now contains scores of articles in which the results of change-detection experiments are interpreted as revealing VWM properties.

2.1. Participants

The study was approved by the Research Subjects Review Board at the University of Rochester. The experiment was conducted over the world wide web via the Amazon Mechanical Turk (MTurk) crowdsourcing marketplace. Interfacing with MTurk was facilitated through the use of the psiTurk programming platform (Gureckis et al., 2016). psiTurk was configured so that only individuals based in the United States could participate. Participants stated that they were at least 18 years old. Fifty participants took part in the experiment. It took approximately 10-15 min to complete the experiment, and participants received $2.00 for their participation.

2.2. Stimuli

Stimuli were comprised from 32 images of real-world objects selected from an image collection developed by Brady, Konkle, Alvarez, and Oliva (2008). Objects were eight exemplars from each of four categories (food, animal, furniture, clothing). For instance, "apple", "bread", "hamburger", "cake", "sundae", "avocado", "roast chicken" and "cappuccino" were the eight exemplars from the "food" category.

For each participant, stimuli were presented in a window whose size was fixed at 1024×768 pixels. Stimuli were rendered in a display box of size 800×600 pixels. An image of an individual exemplar was 50×50 pixels.

A stimulus showed a fixation mark at the middle of the display box, along with four evenly-spaced, distinct objects located on an imaginary circle (radius = 150 pixels) centered at the fixation mark. Although objects were always 90° apart, the placement of objects on the circle was chosen at random on each trial.

The experiment contained two trial types. On a *coherent* trial, all four objects were randomly selected from the same category. On an *incoherent* trial, two objects belonged to one category and the remaining two objects belonged to a different category. For instance, a stimulus on an incoherent trial may have contained two "furniture" objects and two "animal" objects.

2.3. Procedure

The experimental procedure is illustrated in Fig. 1. On a trial, a stimulus was shown for 1000 ms, then a blank screen was displayed for 1000 ms, and finally, another stimulus was shown. The second stimulus remained on the screen until a participant made a response.

The experimental task was a change-detection task. On a *same* trial, the second stimulus was identical to the first stimulus. On a *change* trial, the second stimulus contained the same objects as the first stimulus. From this set of objects, two objects belonging to the same category were randomly selected, and the second stimulus was created from the first stimulus by swapping their locations. If a participant judged the second stimulus to be identical to the first stimulus, the participant responded by pressing the "s" (same) key. Otherwise the participant pressed the "d" (different) key.

Each participant performed 120 trials: 8 practice trials and 112 experimental trials. Practice trials were excluded from data analysis. On the experimental trials, there were equal numbers of coherent-same, coherent-change, incoherent-same, and incoherent-change trials. The order of trials was randomized. In practice trials, a participant received feedback (correct or incorrect) on each trial after entering a response. In experimental trials, a participant received feedback at the end of each block of eight trials. This feedback displayed the number of correct responses during that block (e.g., "5 of 8 responses correct").

2.4. Results

We focused our analyses on data from participants with above-chance performance. A (one-sided) binomial test was used to identify these participants. Of the 50 participants, 37 (74 percent) performed better than chance.

For each participant, we compared their percent correct on coherent trials with their percent correct on incoherent trials. On average, participants performed better in the coherent condition (based on a one-

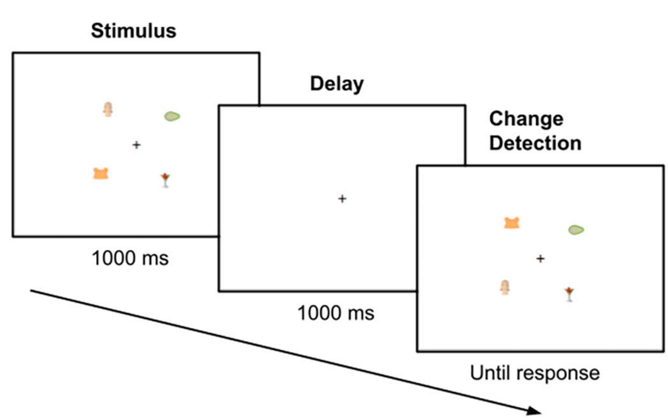

Fig. 1. Experimental procedure for Experiment 1. This figure illustrates an incoherent-change trial (i.e., a change trial in which two objects belonged to one category and the remaining two objects belonged to a different category).

sided t-test: $t = 1.874$ $p = 0.0345$). (A one-sided test is justified by our prediction that category coherence aids VWM performance.) The results are shown in Fig. 2 where each dot illustrates a participant's performance. Overall, these results indicate that participants found coherent trials to be easier than incoherent trials, thereby suggesting that semantic information influences VWM performance.

3. Experiment 2

Experiment 2 was identical to Experiment 1 with the following exception. In Experiment 1, a change trial was created by randomly choosing two objects from the same category in a trial's first stimulus, and then swapping their positions to create the second stimulus. In contrast, a change trial in Experiment 2 was created by randomly selecting an object in a trial's first stimulus, and then replacing this object with another object from the same category to create the second stimulus (with the constraint that the four objects in the second stimulus must be distinct). For instance, suppose the four objects in the first stimulus were “apple”, “bread”, “hamburger” and “cake”. And suppose that one of the objects, an “apple”, was randomly selected. To create the second stimulus, the “apple” might be replaced by a distinct “food” object, such as an “avocado”, to create the second stimulus. In addition, 52 participants took part in Experiment 2.

3.1. Results

To focus our analyses on data from participants with above-chance performance, a (one-sided) binomial test was used to identify these participants. Of the 52 participants, 41 (79 percent) performed better than chance.

Based on data from participants with above-change performance, the average performance was not significantly different in the coherent and incoherent conditions. This outcome, of course, is contrary to our prediction that category coherence aids VWM performance. Interestingly, however, a different outcome emerged when we conducted finer-scale analyses that separately considered change trials and same, or no-change, trials.

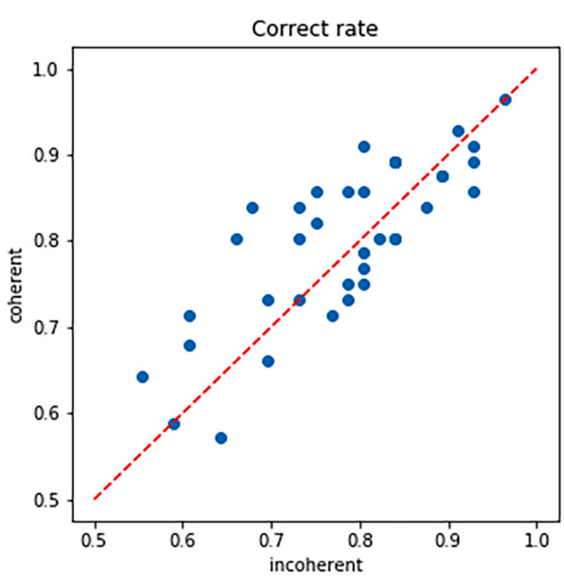

Fig. 2. Each dot illustrates a participant's performance on Experiment 1. The horizontal axis shows the percent correct of a participant on incoherent trials (out of 56 trials) and the vertical axis shows the percent correct on coherent trials (again, out of 56 trials). The red-dashed diagonal line indicates where the percentages are equal. (Although 37 participants had above-chance performance, not all dots are visible in this graph due to dot overlap.)

Starting with change trials, we found that participants performed significantly better (one-sided t-test: $p = 0.035$) on coherent trials (when all objects in a stimulus belonged to the same semantic category) than on incoherent trials (when two objects belonged to one category and the remaining two objects belonged to a different category). On no-change trials, participant performance did not significantly differ in coherent versus incoherent conditions. These results are shown in Table 1. For comparison purposes, this table also displays the analogous results from Experiment 1.

It seems that participants typically found no-change trials to be relatively easy, responding correctly on more than 90% of these trials regardless of whether trials were coherent or incoherent with respect to category membership. In contrast, participants found it challenging to detect changes present on change trials. Critically for our purposes, participants performed better on these trials when all objects in a trial's stimuli belonged to the same category. Similar to Experiment 1, this result indicates that semantic regularities—category coherence in this case—aids VWM performance.

4. Experiment 3

Experiment 3 was a variation of Experiment 1 with a more complicated spatial structure.

4.1. Participants

Fifty participants took part in Experiment 3. It took approximately 15–20 min to complete the experiment, and participants received $3.00 for their participation.

4.2. Stimuli

Experiment 3 used the same source of stimuli as Experiments 1–2 except that each category contained only four exemplars. For instance, “apple”, “bread”, “hamburger”, and “cake” were the exemplars from the “food” category.

For each participant, stimuli were presented in a window whose size was fixed at 1024×768 pixels. Stimuli were rendered in a display box of size 800×500 pixels. An image of an individual exemplar was 50×50 pixels.

An individual stimulus showed six distinct objects, spatially organized into two clusters with three objects per cluster. A cluster center

Table 1
Average participant performance (percent correct) in Experiments 1 (top) and 2 (bottom). Columns show performance on change versus no-change trials (performance on all trials is also provided). Rows show performance on category coherent versus incoherent trials. p values (based on one-sided t-tests) indicating the statistical significances of differences in performance on coherent versus incoherent conditions are also shown.

Experiment 1	Change	No-change	All trials
Coherent	0.68	0.923	0.801
Incoherent	0.625	0.938	0.781
p	**0.00302**	0.934	**0.0345**
Experiment 2	**Change**	**No-change**	**All trials**
Coherent	0.598	0.922	0.76
Incoherent	0.569	0.93	0.75
p	**0.0354**	0.788	0.132

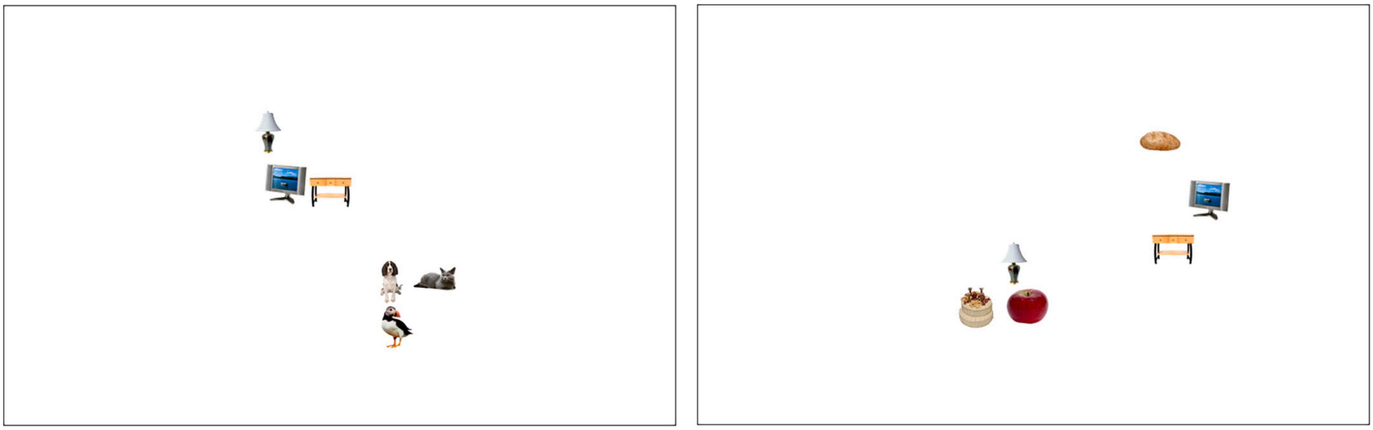

Fig. 3. Example stimuli of Experiment 3 from the coherent (left) and incoherent (right) trial types.

was randomly selected such that clusters were within the display box (each cluster had a radius of approximately 150 pixels; the distance between two cluster centers was, on average, approximately 250 pixels). Within a cluster, the locations of individual objects were sampled from a normal distribution whose mean was the cluster center. Locations were resampled if objects overlapped.

The experiment contained two trial types (see Fig. 3). In a *coherent* trial, each cluster contained objects from the same category. For example, one cluster may have contained “furniture” objects whereas the other contained “animal” objects. In an *incoherent* trial, objects also came from two categories. However, each cluster contained objects from two different categories. For instance, a cluster may have contained one “furniture” and two “animal” objects.

4.3. Procedure

The experimental procedure is shown in Fig. 4. On a trial, a stimulus was shown for 3000 ms, then a blank screen was displayed for 1000 ms, and finally, another stimulus was shown. The second stimulus remained on the screen until a participant made a response.

The experimental task of Experiment 3 was also a change-detection task. On a *same* trial, the second stimulus was identical to the first stimulus. On a *change* trial, the second stimulus contained the same objects as the first stimulus. However, one of the spatial clusters in the first stimulus was randomly selected, and the second stimulus was created by swapping the locations of two objects from the selected cluster belonging to the same category.

Critically, one object location was cued in the second stimulus using a box around that location. On a same trial, the cued location was centered at a randomly chosen object. On a change trial, the cued location was centered at one of the objects whose locations was swapped. A participant's task was to judge whether the object at the cued location was the same in the first and second stimuli or was different. The participant responded by pressing either the “s” (same) or “d” (different) keys.

Each participant performed 80 trials: 16 practice trials and 64 experimental trials. Practice trials were excluded from data analysis. On the experimental trials, there were equal numbers of coherent-same, coherent-change, incoherent-same, and incoherent-change trials. The order of trials was randomized. In practice trials, a participant received feedback (correct or incorrect) on each trial after entering a response. In experimental trials, a participant received feedback at the end of each block of eight trials. This feedback displayed the number of correct responses during that block (e.g., “5 of 8 responses correct”).

In general, the experimental task was difficult. Compared with Experiment 1, which was also a change-detection task in which object locations were swapped, Experiment 3 included more objects and a more complex spatial structure. In addition, the high similarity of first

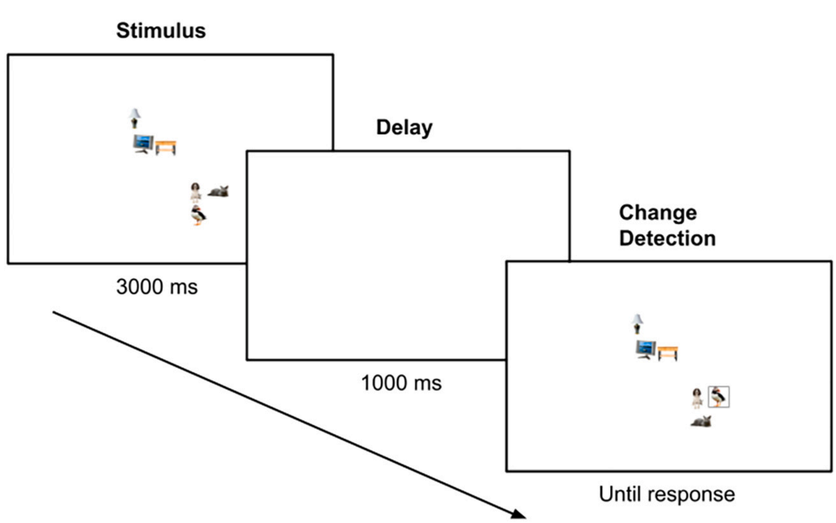

Fig. 4. Experimental procedure for Experiment 3. This figure illustrates a coherent-change trial (i.e., a change trial in which stimuli showed clusters of objects from the same category).

and second stimuli on change trials contributed to task difficulty. These stimuli used the same objects and the same locations. Of the six objects used on a change trial, four objects were at the same location. Of the two objects at new locations, these objects belonged to the same category. Moreover, they were not assigned novel locations, but rather their locations were swapped. For these reasons, good performance on the task required a high level of attention and focus.

4.4. Results

A (one-sided) binomial test was used to identify participants whose performance was greater than chance. Evidently, several participants found the task to be extremely challenging. Of the 50 participants, only 31 (62 percent) performed better than chance.

Using solely the data from participants with above-chance performance, we compared each participant's accuracy on the coherent and incoherent trials. On average, participants performed better in the coherent condition (based on a one-sided *t*-test: $t = 2.179$, $p = 0.0186$). The results are illustrated in Fig. 5 where each dot plots a participant's performance. In this graph, the horizontal axis shows a participant's accuracy (percent correct) on incoherent trials (out of 32 trials) and the vertical axis shows the accuracy on coherent trials (again, out of 32 trials).

Although Experiment 3 was difficult for many participants, its data revealed the same pattern as Experiments 1–2, namely that category regularities aided VWM performance. That is, VWM performance was better when stimuli contained clusters of objects belonging to a common semantic category.

5. Experiment 4

Methodologically, Experiment 4's stimuli and procedure were similar to those of Lew and Vul (2015). Participants viewed stimuli showing two clusters of objects, and then attempted to recall objects' locations. On *coherent* trials, objects belonging to the same cluster also belonged to the same category (e.g., animals). On *incoherent* trials, objects belonging to the same cluster belonged to different categories (e.g., animals, food). If category regularities aid VWM performance (as in Experiments 1–3), then we predict that participants will show better performance on coherent trials.

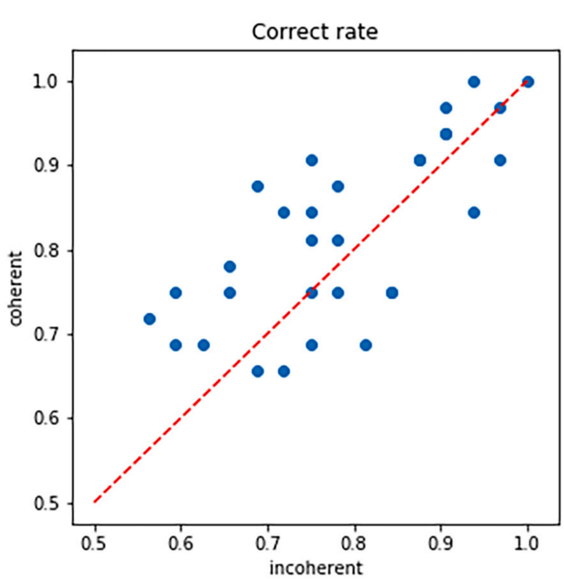

Fig. 5. Each dot plots a participant's performance on Experiment 3. The horizontal axis shows a participant's percent correct on incoherent trials (out of 32 trials) and the vertical axis shows the correct rate on coherent trials (again, out of 32 trials).

Experiment 4 used three stimulus display durations. As noted above, Brady et al. (2016) found that long encoding times aided VWM performance for real-world objects. If the influence of category coherence interacts with encoding times, then participants' responses should vary with stimulus duration.

5.1. Participants

Two hundred twenty-two participants (74 in each of three experimental conditions) took part in Experiment 4. It took approximately 20-30 min to complete the experiment, and participants received either \$2.25, \$2.50, or \$3.25 for their participation depending on the experimental condition.

5.2. Stimuli

Experiment 4 used similar stimuli as Experiment 3, with the exception that a stimulus in Experiment 4 showed eight distinct objects arranged in two clusters, four objects per cluster. The distance between two cluster centers was approximately 200 pixels. On an incoherent trial, each cluster contained two objects from one category and two objects from another category.

Three stimulus display duration conditions were used in Experiment 4: one used a stimulus duration of 4 s, another used a duration of 6 s, and the final condition used a duration of 9 s. Each participant participated in only one condition.

5.3. Procedure

Each participant performed 40 trials: four practice trials (excluded from data analysis) and 36 experimental trials. Half of the experimental trials were incoherent and half were coherent. These trial types were randomly intermixed. For each trial, two object categories were randomly selected.

On a trial, a stimulus appeared on the screen for either 4, 6, or 9 s, depending on the experimental condition. A participant attempted to memorize the objects and their locations. The stimulus then disappeared. After 500 ms, the (empty) display box reappeared, and the objects from the stimulus were shown arranged in random order along a horizontal row below the display box. Using their computer mouse, the participant "dragged and dropped" each object to its memorized location. There was no time limit for entering this response. When the participant completed the response, they pressed a button, and then the participant saw a feedback screen with their response in color and the original stimulus in black and white. The participant then pressed a button to proceed to the next trial.

5.4. Results

5.4.1. Euclidean error

Participants' memory errors were quantified using a Euclidean distance error metric which calculates the distance (in pixels) between the object locations in a participant's response and the locations in a stimulus. For each participant, the average error was calculated across incoherent trials and coherent trials. Fig. 6 shows the results for each participant. The three graphs correspond to the 4s, 6s, and 9s stimulus display-time conditions. Within each graph, the horizontal axis plots a participant's average distance error on incoherent trials, and the vertical axis plots the average error on coherent trials. For the three conditions, 53, 63, and 55 participants, respectively, had larger distance errors in incoherent trials.

T-tests were performed to check for statistical significance of the differences in errors between trial types and between stimulus display-time experimental conditions. Differences between incoherent and coherent trial types were significant for 4s, 6s, and 9s conditions ($p < 0.01$ in all cases). Differences between 4s and 6s conditions

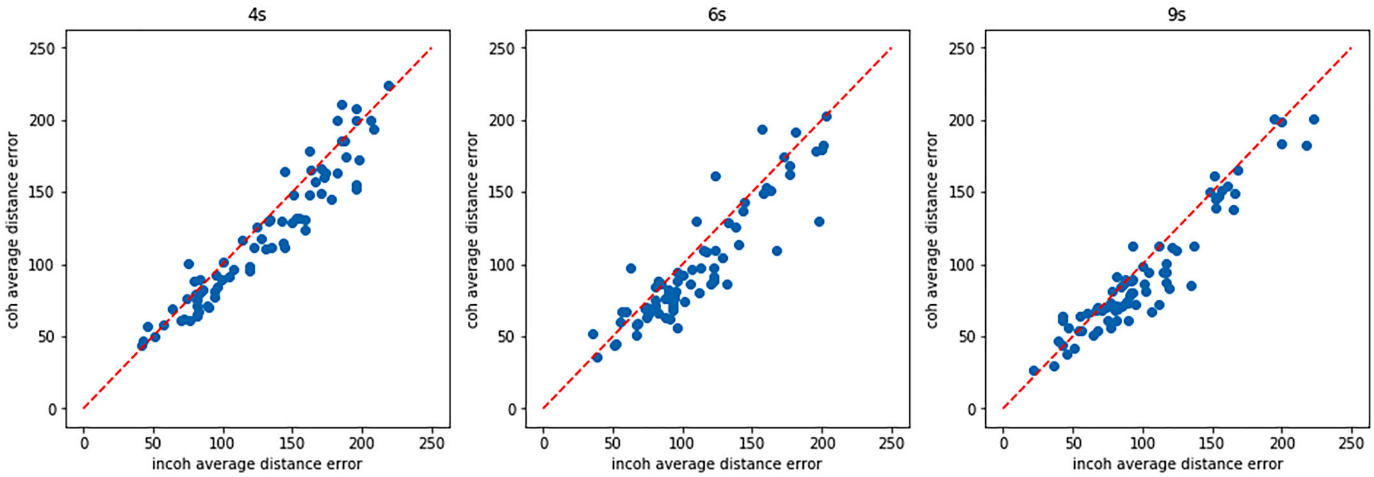

Fig. 6. Each participant's Euclidean distance error averaged across all incoherent trials (horizontal axis) and averaged across all coherent trials (vertical axis) in 4s (left graph), 6s (middle graph), and 9s (right graph) stimulus display-time conditions.

($p < 0.05$ in all cases) and between 4s and 9s conditions ($p < 0.01$ in all cases) were significant. Differences between 6s and 9s conditions were not statistically significant.

Using each participant's errors, we next calculated a participant-specific "semantic index" (SI) to quantify the reduction in error due to adding semantic coherence to the clusters in a visual stimulus (relative to the error in the absence of semantic coherence). We defined SI as follows:

$$SI = \frac{E_I - E_C}{E_I} \times 100$$

where E_I and E_C denote a participant's average errors in the incoherent and coherent trial types, respectively. The average values of SI (averaged across participants) in the 4s, 6s, and 9s conditions are 6.2, 9.0, and 7.1, respectively (differences across conditions were not statistically significant). That is, the addition of semantic coherence leads to a roughly 7.4% reduction in error.

Overall, the experiment found that memory errors were smaller on coherent trials than incoherent trials, indicating that category coherence improved visual memory performance in our participants. This result is consistent with the hypothesis that VWM representations are influenced by semantic regularities. Lastly, although memory errors tended to be larger when stimulus durations were shorter (as expected), we normalized for this effect in our SI measure. Based on this normalized measure, the data do not support the hypothesis that stimulus duration modulates the influence of category regularities on visual memory.

5.4.2. Cluster error count

To further understand participants' performances, we computed a "cluster error count" which is the number of objects that a participant placed in an incorrect cluster. The first step in calculating this measure is to cluster objects in a stimulus and a participant's response. Cluster center locations were calculated using the K-means algorithm (Hastie, Tibshirani, & Friedman, 2009). Then the center locations were used to determine whether an object was placed in an incorrect cluster in a participant's response.

Fig. 7 shows each participant's cluster error count summed across all incoherent trials (horizontal axis) and summed across all coherent trials (vertical axis). For 4s, 6s, and 9s stimulus display-time conditions, 65, 53, and 53 participants, respectively, had larger cluster error counts in incoherent trials. Statistically, the error count in incoherent trials was greater than the count in coherent trials in all conditions ($p < 0.01$ in all cases). In addition, the count decreased as the stimulus display time increased from 4s to 6s, but did not significantly vary as the stimulus display time increased from 6s to 9s ($p = 0.281$ between 6s and 9s; $p < 0.01$ between 4s and 9s; $p < 0.05$ between 4s and 6s).[1]

5.4.3. Cluster size

Finally, we report the "cluster size" which is the average distance of objects to the center of a cluster where, as above, cluster centers were computed using the K-means algorithm (Hastie et al., 2009). In each stimulus display-time condition, there were 74 participants, 36 trials per participant, and 2 clusters per trial, meaning that 5328 clusters appeared in stimuli and responses. The distributions of stimulus and response cluster sizes in each condition are shown in Fig. 8. Response clusters were larger than stimulus clusters in all conditions ($p < 0.01$ in all cases).

Fig. 9 shows each participant's difference between stimulus and response cluster sizes averaged across incoherent trials (horizontal axis) and averaged across coherent trials (vertical axis). In all display-time conditions, incoherent values are larger than coherent values (differences between incoherent and coherent trials are statistically significant at the $p < 0.05$ level in all cases).

[1] This subsection reports an analysis of whether participants placed objects in their correct clusters. A reader might wonder if it's possible to also analyze "within cluster" errors. Unfortunately, we are not aware of a good way of doing so. Consider a trial on which a participant places an object near the center of an incorrect cluster. A within-cluster analysis might assign a small error score—after all, the object is near the cluster center. But a small error score in this situation would be misleading. In fact, the participant made a large error by placing the object in the wrong cluster. One possible approach might be to consider only the subset of trials on which a participant placed all objects in their correct clusters, referred to here as "high performance" trials. Unfortunately, however, this "cherry picking" of trials introduces a bias. There were a relatively large number of high-performance trials on coherent trials, and many fewer on incoherent trials. (For instance, in the 4s stimulus display-time condition, there were 709 high-performance, coherent trials and 440 high-performance, incoherent trials across participants.) This mismatch is a source of bias. Consider the following hypothetical scenario. Suppose that memory performance across trials is quantified as a real-valued score. Further suppose that this score is normally distributed on coherent trials and also normally distributed on incoherent trials. On coherent trials, the mean of the distribution—corresponding to a typical coherent trial—might be a score that yielded a high-performance trial. However, a score yielding a high-performance trial on an incoherent trial would require a score from the right-most tail of the distribution, corresponding to an atypical incoherent trial. This illustrates why we are unaware of a fair and unbiased way to conduct a within-cluster comparison of coherent and incoherent trials.

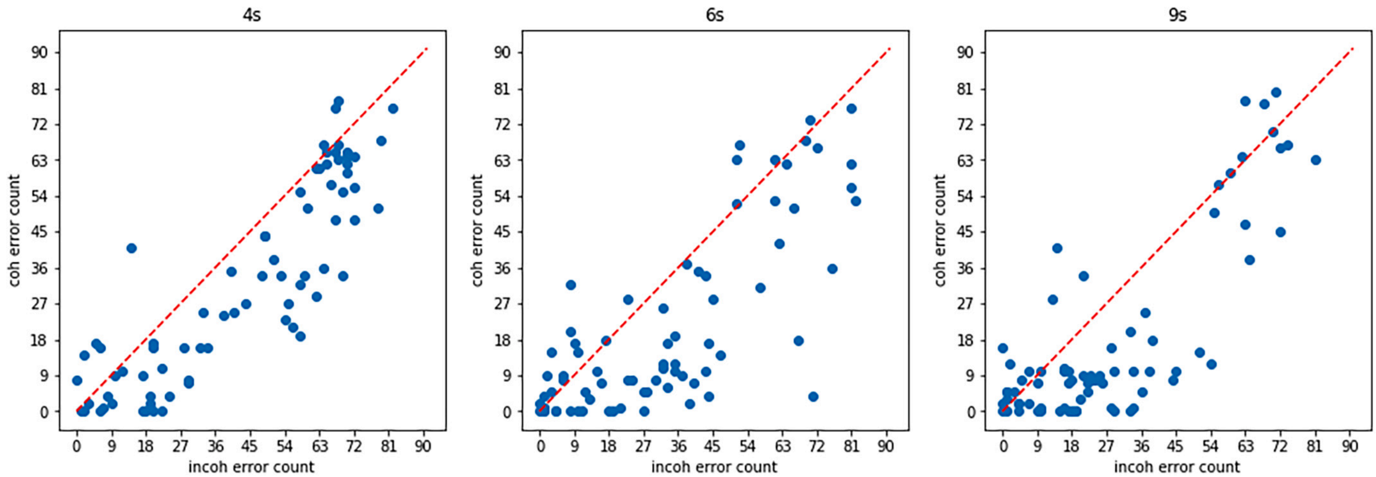

Fig. 7. Each participant's error count summed across all incoherent trials (horizontal axis) and summed across all coherent trials (vertical axis) in 4s (left graph), 6s (middle graph), and 9s (right graph) stimulus display-time conditions.

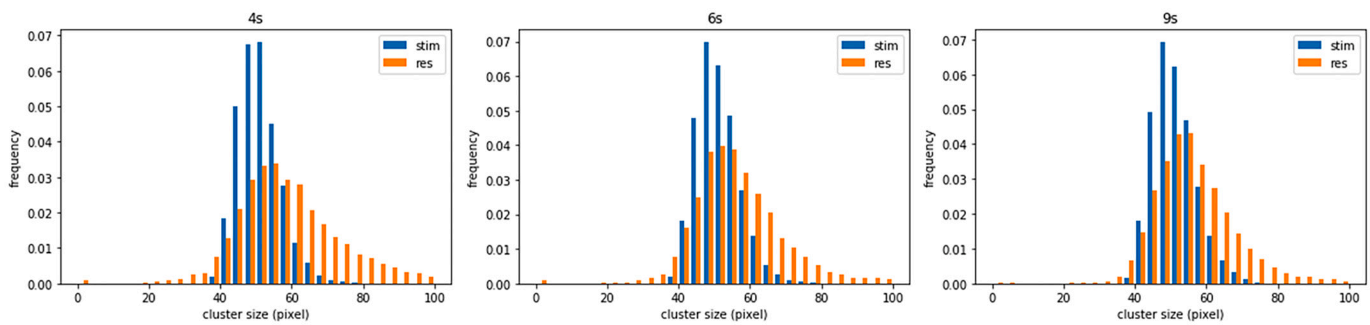

Fig. 8. Stimulus and response cluster size distributions in 4s, 6s, and 9s stimulus display-time conditions.

6. Discussion

In summary, we conducted four experiments studying semantic influence—influence due to regularities in object category membership, in particular—on VWM. Experiments 1–2 used real-world objects around a center fixation mark. Objects came either from the same category (coherent trials) or different categories (incoherent trials). Experiments 3–4 used real-world objects organized into spatial clusters. Objects belonging to a cluster either belonged to the same category (coherent trials) or belonged to different categories (incoherent trials). Experiments 1–3 used a change-detection paradigm. Experiment 4 asked participants to recall the locations of objects. All four experiments yielded the same result: participants showed better VWM performance on coherent trials than on incoherent trials. Taken as a whole, the experiments provide the best (perhaps only) data to date demonstrating that statistical regularities in semantic category membership improve VWM performance.

Our result is inconsistent with recent work by Luu and Stocker

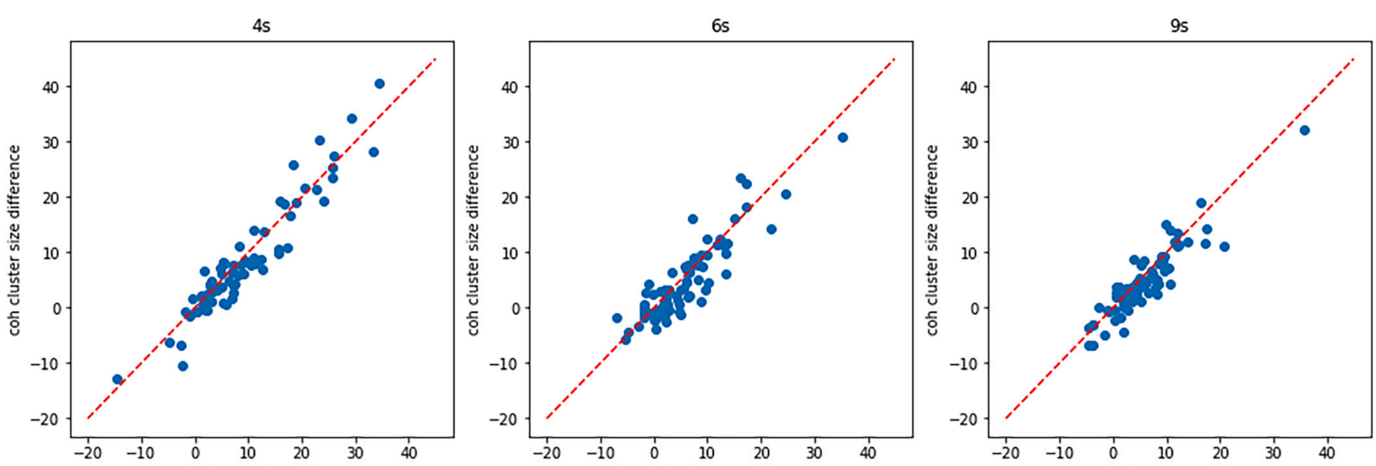

Fig. 9. Each participant's difference between stimulus and response cluster sizes averaged across incoherent trials (horizontal axis) and averaged across coherent trials (vertical axis) in 4s (left graph), 6s (middle graph), and 9s (right graph) stimulus display-time conditions.

(2021) who argued that category information does not influence VWM, although it does influence subsequent sensory recall and inference processes. The experiments reported here used a change-detection paradigm, a common experimental paradigm in the scientific literature, whereas the experiments of Luu and Stocker (2021) used a novel paradigm in which participants had to reverse their categorical judgments about a stimulus feature, if incorrect, before providing an estimate of the feature. It seems likely that these methodological differences account for differences in experimental findings. Future work will need to investigate this further.

Our result is also inconsistent with a conventional perspective in the vision sciences literature in which VWM is sensitive solely to bottom-up visual properties of objects, such as objects' shapes, colors, orientations, and locations. Instead, our result indicates that VWM is also sensitive to top-down semantic information, such as regularities in category membership. This work contributes to a small, but growing, number of studies suggesting a role for semantic information in VWM processing (Asp, Störmer, & Brady, 2021; Brady, Störmer, & Alvarez, 2016; Conci, Kreyenmeier, Kröll, Spiech, & Müller, 2021; Liu et al., 2021).

Most experiments on VWM use meaningless stimuli (e.g., oriented bars, colored squares), thereby avoiding the question of whether semantics influences VWM. Advocates of this approach might define VWM as solely processing and storing visual properties of objects and scenes (e.g., object shape, color, orientation). By this definition, it is nonsensical to consider semantic influence on VWM. Recently, however, there has been a trend in the field of cognitive science to consider more realistic scenarios, including the use of meaningful stimuli such as real-world objects. This raises additional complexities.

Because our work (and the work of others too) indicates a role for semantic information in VWM processing, it suggests that the field of cognitive science may need to modify its conceptualization of VWM so that it is closer to that of "conceptual short-term memory" (CSTM). In articles that have, to date, received relatively little attention, Potter (1993, 2012) argued for the existence of CSTM, and described it as a short-term memory store that is engaged unconsciously and rapidly, and represents current stimuli and their associated concepts from long-term memory. Future work will need to investigate CSTM and its relationship with VWM. What are the properties of CSTM? How is it distinct from VWM? Is the distinction between CSTM and VWM "psychologically real"? Are both concepts needed or is their distinction in cognitive science due to convenience or historical accident?

Acknowledgments

We thank Brady et al. (2008) for making their image collection available on the World Wide Web. This work was supported by NSF research grant DRL-1561335.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104891.

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.
Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological Review, 127*, 891–917.
Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory. *Psychological Science, 22*, 384–392.
Brady, T. F., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 921–929.
Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*, 487–502.
Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America, 105*(38), 14325–14329.
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 of the United States of America, 113*, 7459–7464.
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 and Review* (in press).
Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., , … Chan, P., et al. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavioral Research Methods, 48*, 829–842.
Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning* (2nd ed.). Springer.
Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the weber noise of relative positions. *Journal of Vision, 15*(4), 10, 1–14.
Liu, J., Zhang, H., Yu, T., Ni, D., Ren, L., Yang, Q., … Xue, G. (2021). Stable maintenance of multiple representational formats in human visual short-term memory. *Proceedings of the National Academy of Sciences of the United States of America, 117* (51), 32329–32339.
Luu, L., & Stocker, A. A. (2021). Categorical judgments do not modify sensory representations in working memory. *PLoS Computational Biology, 17*(6), e1008968.
Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition, 122*, 346–362.
Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition, 21*(2), 156–161.
Potter, M. C. (2012). Conceptual short term memory in perception and thought. *Frontiers in Psychology, 3*, 113.
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*, 803–814.
Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review, 119*, 807–830.
Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports, 8*, 16162.