# Journal of Experimental Psychology: Learning, Memory, and Cognition


## Bridging the Gap Between Visual Temporary Memory and Working Memory: The Role of Stimuli Distinctiveness

Roy Shoval, Roy Luria, and Tal Makovski

Online First Publication, October 24, 2019. http://dx.doi.org/10.1037/xlm0000778

American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition



ISSN: 0278-7393

http://dx.doi.org/10.1037/xlm0000778

# Bridging the Gap Between Visual Temporary Memory and Working Memory: The Role of Stimuli Distinctiveness

Roy Shoval
The Open University of Israel

Roy Luria
Tel Aviv University

Tal Makovski
The Open University of Israel

OPEN DATA

Visual working memory (VWM) is traditionally assumed to be immune to proactive interference (PI). However, in a recent study (Endress & Potter, 2014), performance in a visual memory task was superior when all items were unique and hence interference from previous trials was impossible, compared to a standard condition in which a limited set of repeating items was used and stimuli from previous trials could interfere with the current trial. Furthermore, when all the items were unique, the estimated memory capacity far exceeded typical capacity estimates. Consequently, the researchers suggested the existence of a separate memory buffer, the "temporary memory," which has an unbounded capacity for meaningful items. However, before accepting this conclusion, methodological differences between the repeated-unique procedure and typical estimates of VWM should be considered. Here, we tested the extent to which the exceptional set of heterogeneous, complex, meaningful real-world objects contributed to the large PI in the repeated-unique procedure. Thus, the same paradigm was employed with a set of real-world objects and with homogenous sets (e.g., houses, faces) in which the items were meaningful, yet less visually distinct, and participants had to rely on subtle visual details to perform the task. The results revealed a large PI effect for real-world heterogeneous objects, but substantially smaller effects for the homogenous sets. These findings suggest that there is no need to postulate a new memory buffer. Instead, we suggest that VWM capacity and vulnerability to PI are highly influenced by task characteristics, and specifically, by the stimuli distinctiveness.

*Keywords:* proactive interference, visual working memory, temporary memory, heterogeneity

*Supplemental materials:* http://dx.doi.org/10.1037/xlm0000778.supp



Short-term visual memory storage was reported more than 40 years ago by Phillips (1974). In a series of experiments, it was shown that this storage, also known as visual working memory (VWM), is different from the sensory storage (Sperling, 1960) on several parameters. Most importantly, while the sensory storage has a high capacity of information for a very brief duration, the VWM can maintain a limited amount of information but for a longer duration of several seconds. The limited nature of VWM is one of its most researched characteristics, and now days, there is a widespread agreement that it can hold a limited capacity of about three to four items (e.g., Luck & Vogel, 1997; Matsukura & Hollingworth, 2011), although it is still unclear what is the source of this limitation (for reviews of competing mechanisms, see Cowan, 2001; Ma, Husain, & Bays, 2014).

In light of this limited capacity, it is highly important that VWM should effectively remove remembered items that are no longer needed to make room for new ones. That is, it is reasonable to expect that VWM should be relatively immune to cases in which previously stored items are maintained in memory for longer than required by the task demands and consequently interfere with current performance, a phenomenon known as proactive interference (PI; Keppel & Underwood, 1962; Peterson & Peterson,

Roy Shoval, Department of Education and Psychology, The Open University of Israel; Roy Luria, Department of Psychology, Tel Aviv University; Tal Makovski, Department of Education and Psychology, The Open University of Israel.

This research was presented at the 2019 Israeli Conference on Cognition Research.

This research was supported by the Israel Science Foundation (Grant 1344/17). We thank Shiri Tascesme, Ilana Vinizky, and Hagar Cohen for help in data collection. Authors have no conflicts of interest to disclose.

All data is publicly available at osf.io/r85nz.

The data are available at osf.io/r85nz.

Correspondence concerning this article should be addressed to Roy Shoval, Department of Education and Psychology, The Open University of Israel, Derekh ha-Universita 1, Ra'anana 4353701, Israel. E-mail: rshoval@gmail.com

1959). If, however, VWM is not immune to PI, then this might have severe consequences on the capacity estimates of VWM that typically relies on the assumption that this effect plays only a minimal role in standard measures of VWM, such as the change-detection task (Luck & Vogel, 1997). Indeed, the few studies that examined the role of PI in VWM using change-detection tasks found only modest PI effects, if any (Hartshorne, 2008; Lin & Luck, 2012; Makovski & Jiang, 2008; Oberauer, Awh, & Sutterer, 2017; Shipstead & Engle, 2013). These findings suggest that one characteristic of the VWM is that it is relatively immune to PI, and hence, that the capacity estimates of VWM as produced by the change-detection task, are not considerably compromised by PI.

Nonetheless, a recent study critically challenged these conclusions as it found a substantial PI effect in a VWM task, and when the PI was eliminated, VWM capacity was greatly increased (Endress & Potter, 2014). In this study, the researchers used a modified version of the change-detection task and introduced a unique condition in which each item appeared only once throughout the experiment. This condition was compared to a repeated condition in which a limited set of items was used and these items appeared multiple times throughout the experiment (as typically done in VWM experiments). The results showed that performance was much better in the unique condition, in which PI could not occur because each item was novel, than in the repeated condition, in which source memory confusions could occur (Endress & Potter, 2014). Moreover, it was found that when all the items were unique, the estimated capacity of memory was substantially higher compared to previous estimates of VWM capacity, as participants performed above chance level even when 100 items were presented in the memory display. Consequently, when estimating participants' capacity limit (known as *k*; Cowan, 2001; Rouder et al., 2008) using the typical formula, which assumes a fixed capacity, a capacity limit of about 30 items was observed. This pattern of results was replicated in an additional study in which the superior performance in the unique condition was stable and unaffected by the items' encoding durations or by the presence or absence of a parallel verbal suppression task (Endress & Siddique, 2016).

The findings of Endress and colleagues (Endress & Potter, 2014; Endress & Siddique, 2016) may imply that the capacity of VWM is not limited as is widely accepted. Instead, previous estimations of this capacity might be flawed and misleading, due to the involvement of PI in the change-detection tasks that used limited, and hence repeated, sets of items. However, this was not what the researchers claimed. Instead, to bridge the gap between their findings and the limited capacity of VWM, they suggested the existence of an additional memory buffer, which they termed *temporary memory*. Based on real-life situations, such as driving, in which there seems to be no fix limitation on what one can remember, they suggested that this temporary memory specializes in meaningful information and can store an unlimited amount of this information without relying on active maintenance or selective attention, as opposed to working memory that is considered to rely on these mechanisms (e.g., Baddeley, 2012; Conway, Kane, & Al, 2005; Kane & Engle, 2000). Accordingly, the items that were used in their experiments were images of meaningful objects. Yet, they further argued that the information from this temporary memory in their experiment was not consolidated into long-term memory (LTM), as the performance in a surprise memory test at the end of the experiment, was at chance level for items presented in the unique condition. Notably, however, before postulating a new memory buffer, it is imperative to consider some methodological differences between the typical change-detection task and the repeated-unique procedure (RUP; Endress & Potter, 2014; Endress & Siddique, 2016) that might account for the discrepancies in the findings.

First, while the items in a change-detection task are usually presented simultaneously, each at a distinct location, items in the RUP were presented sequentially at the middle of the screen. Indeed, it has been argued that PI might be location specific (Makovski & Jiang, 2008) and accordingly it has been found that the PI effect, as reflected by a superior memory in a unique compared to a repeated condition, was greatly reduced when each item was presented at a distinct location (Makovski, 2016). In that study (Makovski, 2016), the dependency between items' locations and the magnitude of the PI effect was apparent both when the items appeared simultaneously and when they appeared sequentially, suggesting that PI is spatially specific and that spatial distinctiveness reduces PI regardless of the presentation mode. Together, these findings (Endress & Potter, 2014; Endress & Siddique, 2016; Makovski, 2016) imply that the capacity of VWM is not as stable as was typically reckoned, as it is vulnerable to PI. However, they further imply that past estimates of VWM capacity in change-detection tasks were not considerably compromised, as the spatial distinctiveness of the items in these tasks could serve as an effective cue for overcoming PI.

A second fundamental difference between the RUP and a typical change-detection task is the type of items that participants had to remember. In the RUP (Endress & Potter, 2014; Endress & Siddique, 2016), a heterogeneous set of meaningful real-world objects was used, while the majority of the change-detection tasks used homogenous sets of meaningless items or simple visual features (e.g., Luck & Vogel, 1997). Naturally, the use of real-world, meaningful objects, has important ecological justifications, as well as methodological ones, as the unique condition requires a large number of visually distinct items. Nevertheless, it has been found that the capacity limit of such real-world objects is larger compared to simple stimuli (Brady, Störmer, & Alvarez, 2016) and it is currently unknown whether the PI effect can be generalized to other types of stimuli.

Specifically, there are two notable differences between real-world objects and simpler stimuli that might influence the PI effect. First, real-world objects are composed of complex conjunctions of visual features that lead to substantial, and multidimensional, visual differences among the items. This visual heterogeneity, in turn, makes it easier to distinguish between the items based on relatively superficial visual encoding, compared, for example, to a situation in which the stimuli differ across one simple visual feature such as an orientation of a bar, which requires a finer visual encoding. Consider, for example, the distinction between a wall clock and a shovel (both are illustrated in Figure 1). These items are of different sizes, have different colors, are composed of different shapes, and so forth. Hence, an encoding of any of these features can lead to a distinction between the items. On the other hand, to distinguish between two bars with a different orientation, a specific encoding of this continuous feature is required. Indeed, it is well known that confusability that results from the items' visual characteristics severely impairs memory performance (e.g., Brady, Konkle, & Alvarez, 2011; Viswanathan, Perl, Visscher, Kahana, & Sekuler, 2010). As a result, performing the RUP with complex objects might have been substantially easier compared to experiments that used simple stimuli. More importantly for the current purposes, the large visual heterogeneity between the objects, which makes the task much easier in the unique condition, has only a limited effect on the repeated condition because in this condition confusion with previous trials is possible. Therefore, the difference between the two conditions, which is our proxy for the magnitude of the PI effect, might be considerably enlarged when distinct, meaningful objects are tested.

A second major difference between real-world objects and simpler stimuli is that the former increases the likelihood that semantical categorization, and subsequently LTM, would be involved in the task. That is, with these stimuli, Endress and Potter (2014) paradigm, despite the relatively fast rate of presentation, could not properly rule out the possibility that LTM was involved in the task. This is problematic because there is evidence that conceptual information increases capacity estimations (Brady, Konkle, Alvarez, & Oliva, 2008; Feigenson & Halberda, 2008; Jackson & Raymond, 2008). Furthermore, the involvement of semantic information in the task undermines the notion that VWM by itself suffers from PI. Rather, the source of the effect might be due to the involvement of LTM that contributes to greater interference from previous trials. Note that although Endress and Potter (2014) had shown via the surprise memory test at the end of their experiment that the stored information was not consolidated into LTM, this does not mean that LTM was not involved in the task on a trial to trial basis. That is, it is quite possible that conceptual information from LTM was used on each trial of the unique condition but this did not contribute to later memory of the objects.

It is worth noting that these two factors, namely, objects' heterogeneity and object's meaning are not mutually exclusive. This is because the meaning of the object is yet another feature that if encoded is sufficient to discriminate between the stimuli. In other words, in a heterogeneous set of complex stimuli, each stimulus is highly distinct from the other stimuli because of its distinct visual characteristics as well as its distinct semantical meaning. Importantly, the outcome is the same—when a heterogeneous set of complex objects is used, subjects are less encouraged to encode fine visual details, and their performance in the unique condition is enhanced, especially in the unique condition in which confusion from previous trials is impossible. As a result, the difference between the unique and repeated conditions, our estimate of the magnitude of PI, might be inflated.

What is the role of stimuli distinctiveness in creating the PI effect? The present study aimed to test whether the stimuli that were used by Endress and Potter (2014) and specifically their heterogeneity, underlies the large PI effect that was found in the RUP and not in typical change-detection tasks. Naturally, following past studies we expected that reducing the visual distinctiveness of the items should impair overall memory performance (e.g., Brady et al., 2011; Viswanathan et al., 2010). Yet, the focus of the present study was not to compare participants' overall performance in the homogeneous and heterogeneous sets, rather we asked whether the difference between the unique and repeated conditions, which is the proxy of the PI effect, critically depends on the type of the tested stimuli.

Therefore, the first experiment was designed to replicate the basic PI effect using a heterogeneous set of real-world objects. The subsequent experiments aimed to test the influence of different types of stimuli on the magnitude of the PI effect. That is, we asked whether the large PI effect would be generalized to different types of stimuli, specifically, ones that will force participants to rely on subtle visual processing and prevent them from using semantic categorization. To do so, we used homogenous sets of items (houses, faces) instead of the heterogeneous set that has been previously used. Notably, the houses and faces are still meaningful stimuli so if temporary memory is indeed unlimited storage for meaningful information then it should be activated for these stimuli and a large PI should be obtained due to superior performance in the unique condition. If, however, the large PI was found primarily because of the large items' heterogeneity, and not because of the involvement of a different memory buffer, then reducing the visual distinctiveness of the items should lower the performance particularly in the unique condition and this should reduce or eliminate the PI effect.

## Experiment 1

The main goal of the first experiment was to replicate the PI effect using a heterogeneous set of real-world, meaningful objects, and to serve as a baseline for comparison with the subsequent experiments that used homogeneous sets. The secondary goal of this experiment was to test whether reducing the number of repeated items would reduce the PI effect. That is, while previous RUP experiments used a relatively large set of 22 repeated objects (i.e., Endress & Potter, 2014; Endress & Siddique, 2016; Makovski, 2016) standard change-detection tasks have usually tested much smaller sets of repeating items (e.g., up to seven colors were used in Luck & Vogel, 1997). This factor could potentially account for the differences between the two procedures because it is likely that more repetitions of fewer items increased the familiarity of these items. Because it is known that familiarity improves VWM performance (e.g., Blalock, 2015, see also Ngiam, Khaw, Holcombe, & Goodbourn, 2018), it is possible that the relatively high familiarity afforded by the small set of repeated items in the change-detection tasks overcame the PI in these tasks. Hence, testing fewer repeated items in the RUP could lead to higher familiarity with these items and consequently to a better performance in the repeated condition, which will result in a smaller PI effect (e.g., a smaller difference between the unique and repeated conditions). On the other hand, one could also expect that testing fewer repeated items will produce more interference from previous trials and as a result, a larger PI effect, due to an increased probability that items in a given trial had appeared in previous trials. Thus, to rule out the possibility that a large number of repeated items is critical for finding a large PI effect in the RUP, a smaller set of 13 items was tested in the first experiment that was otherwise identical to Makovski's (2016) Experiment 1.

### Method

**Participants.** Participants in all experiments were 18–35 years old, had normal or corrected-to-normal visual acuity, normal color vision, and reported having no attentional, psychiatric, or neurological disorders. Each participant took part in only one experiment. The studies were approved by the ethics committee of the Department of Education and Psychology at the Open Univer-

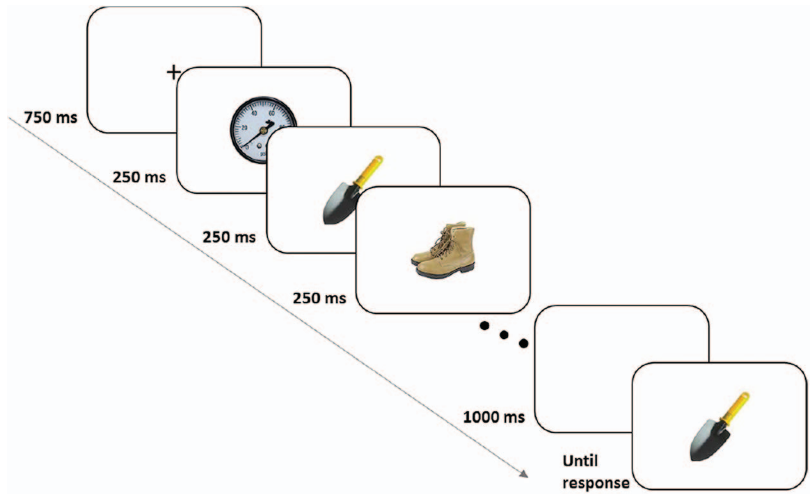


*Figure 1.* An illustration of a trial sequence in all experiments. This is an "old" trial in which the test probe appeared in the sequence. This figure is used with permission. See the online article for the color version of this figure.

sity of Israel. Forty students (seven male; range = 19–34 years old; $M_{age}$ = 25.9) from the Open University of Israel participated in the first experiment for course credit. This sample size was based on Makovski (2016).

**Equipment and stimuli.** Participants were tested individually in a dimly light room. They sat about 67 cm from a 17-in. CRT screen (resolution 1024 × 768, 85 Hz). The experiment was programmed using Psychtoolbox 3.0.14 (Brainard, 1997), implemented in MATLAB R2018a (www.mathworks.com). Stimuli were 2,400 colored images (1.89° × 1.89°) of real-world objects (taken from Brady et al., 2008).

**Procedure.** Each trial began with a black fixation cross (1.3°) appearing at the middle of the screen for 750 ms, followed by four or eight images that were presented sequentially, each for 250 ms. Then, a blank screen appeared for 1,000 ms followed by the test image (see Figure 1). On half of the trials, the test item was an "old" image that appeared in a random position in the sequence, while in the other half it was a "new" image that did not appear in that sequence at all. Participants pressed the "J" key if they thought that the test item was new and the "H" key if they thought it was old. Following a correct response, a green plus sign appeared for 200 ms, while a red minus sign appeared for 1,000 ms after an incorrect response. To reduce verbal recoding, participants repeated a three-letter word aloud, prespecified at the beginning of each 20 trials.

**Design.** Participants underwent two blocks of trials in a counterbalanced order: repeated and unique. In the unique condition, each image appeared only once during the block, while in the repeated condition a set of 13 images was randomly sampled for each participant and was used throughout the block. Each block consisted of 120 trials, randomly and evenly divided into two set size conditions (four, eight) and two correct response conditions (old, new). The two blocks were performed consecutively with breaks given every 20 trials. Importantly, in the new trials of the repeated condition, the test probe was an image from the set of the repeated items but had not appeared in that trial's sequence. For practice, participants completed 20 trials of the unique condition before starting the experiment.

## Results and Discussion

Because of the difficulty of some of the tasks, we set a strict and fixed criterion for excluding data across all of the experiments: First, data from participants whose overall accuracy did not differ from chance level (50%) were excluded. Then, data from outlier participants—defined by an overall accuracy worse than two standard deviations from the mean—were also excluded. Thus, the data from two participants were removed from the analysis of Experiment 1, one participant due to each filter.

Below, we report the results of all experiments in terms of the percentage of correct responses. We also calculated the traditional measures of capacity estimation *k* and these results are reported in the online supplemental materials (Section 1). Overall, the results of these two measures throughout experiments were highly similar.

Figure 2 shows the percentage of correct responses as a function of set size and condition in all experiments. A repeated-measures analysis of variance (ANOVA) with these factors revealed higher

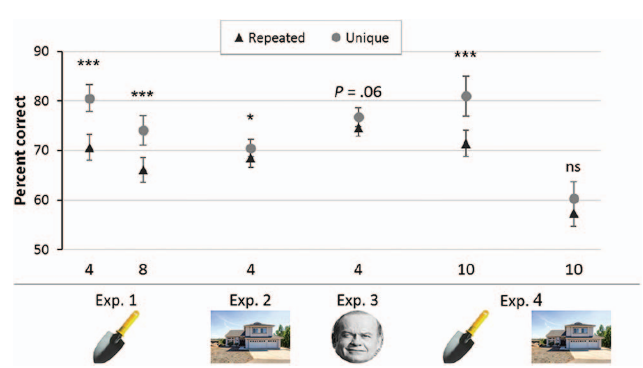


*Figure 2.* Means and 95% confidence intervals for the percent of correct responses across all experiments, pending on the condition and set size. *ns* = not significant. * $p$ < .05. *** $p$ < .001. This figure is used with permission. See the online article for the color version of this figure.

accuracy for the set size of four ($M$ = 75.59, 95% confidence interval [CI: 73.42, 77.76]) compared to set size of eight ($M$ = 70.09, 95% CI [68, 72.18]), $F(1, 37) = 31.01$, $p < .001$, $\eta_p^2 = .46$. More importantly, memory was better in the unique condition ($M$ = 77.3, 95% CI [75.2, 79.4]) compared to the repeated condition ($M$ = 68.37, 95% CI [66.54, 70.2]), $F(1, 37) = 60.03$, $p < .001$, $\eta_p^2 = .62$, and there was no interaction between the factors, $F(1, 37) = 0.92$, $p = .34$. A pairwise comparison confirmed that performance was better in the unique condition for both set size of four (80.53 compared to 70.65), $t(37) = 7.06$, $p < .001$, $d = 1.15$, and set size of eight (74.08 compared to 66.1), $t(37) = 5.7$, $p < .001$, $d = .92$.

A follow-up analysis tested the PI effect as a function of the type of trial (old/new) and the set size, to see if the effect was driven by a specific type of trial. To overcome the concern that the two set sizes have different difficulty baselines in terms of overall task performance, we calculated a PI score measure, which was defined as the percentage in which the performance in the unique condition differed from that of the repeated condition ($\frac{Unique-Repeated}{Repeated}\times 100$). This variable was calculated for each subject in each comparison level, and henceforth, served as the dependent variable when differences in baseline performance might be a concern. However, there were no consistent differences in the PI effect between the old and new trials across experiments, and hence, the results of these analyses are specified in the online supplemental materials (Section 2).

To conclude, Experiment 1 confirmed the existence of a substantial PI effect, as reflected by a superior memory in the unique compared to the repeated condition. Hence, the use of only 13 repeated items did not eliminate the effect, and the difference in the accuracy rates (across set sizes) between conditions was 8.93 ($SD$ = .99). In fact, the size of this effect ($\eta_p^2 = .62$) was larger compared to Experiment 1 in Makovski's (2016) article ($\eta_p^2 = .43$). Hence, it seems that it was not the limited familiarity afforded by the large set of repeated items that drove the effect in previous studies (Endress & Potter, 2014; Makovski, 2016). Rather, it is reasonable to conclude that the use of a smaller set of repeated items, not only did not reduce the PI effect, but if anything, yielded a greater interference in the repeated condition.

## Experiment 2

Experiment 1 confirmed the existence of considerable PI effect in the RUP even when the number of items in the repeated condition was reduced. Experiment 2 aimed to test a major difference between the RUP and typical change-detection tasks—the distinctiveness of stimuli in the memory display. For this purpose, we employed the same paradigm as in the previous experiment but replaced the heterogeneous set of objects with a homogenous set of houses. This change reduced the heterogeneity of the items and therefore forced participants to encode subtle visual characteristics of the stimuli and to rely on these differences throughout the experiment. Additionally, this change minimized the potential use of semantical categorization during the task because all the items were drawn from a single category.

A pilot experiment that used the same design of Experiment 1 and tested sequences of both four and eight items with the images of houses was considerably more difficult and resulted in massive filtering of participants (11 out of 40 participants, 27.5%, were dropped due to overall chance level performance). Thus, only set size of four was tested in Experiment 2, and the method and results of the pilot experiment are reported in the online supplemental materials (Section 3).

Testing only set size of four further enabled us to systematically examine the role of serial position in the PI effect. Previous studies (Endress & Potter, 2014; Makovski, 2016) did not find significant interactions between serial position and the difference between the unique and repeated conditions. However, this issue could not be properly tested in the previous experiment (as well as in the pilot experiment) because the positions in which the items appeared at the sequence were unbalanced and because of the small number of items that were tested in each position. Nevertheless, this issue is important because, in the verbal domain, it was found that only the first items in a list are prone to PI and this finding was taken as evidence supporting the distinction between two memory storages (Craik & Birtwistle, 1971).

Finally, to test whether individual differences in VWM capacity are related to the magnitude of the PI effect, we also measured participants' capacity using a standard color-change-detection task (Allon & Luria, 2017). This enabled us to directly correlate between performance in this task and performance in the RUP. Furthermore, the inclusion of a typical estimation of subject's capacity limit allowed us to test whether individuals with high VWM capacity differ in their PI score from individuals with low VWM capacity. This is important because previous findings that used written words as the visual stimuli showed that low-span individuals are more prone to PI compared to high-span individuals (Kane & Engle, 2000). Furthermore, it was suggested that individual differences in the ability to resolve PI are an important factor that influences capacity measurements of VWM (Bunting, 2006).

### Method

**Participants.** Forty students (10 male; range = 18–33 years old; $M_{age}$ = 24.22) from Tel Aviv University completed the experiment for the pay of ₪30 (about US$8.5).

**Repeated-unique paradigm.** Except for the following changes, the RUP was identical to Experiment 1. It was programmed using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), and was presented on a 23-in. LCD screen (resolution 1920 × 1080). Instead of colored images of real-world objects, the stimuli were colored images of houses, taken from Google images (Figure 3a), and were larger (6.7° × 4.5°) compared to the previous experiment to make it possible to distinguish between them. The fixation cross appeared for 500ms and before and after the presentation of each sequence, a rectangle divided into two black and two gray rectangles was presented for 250ms and served as a mask. Participants were not required to repeat words aloud throughout the experiment. The task included 360 trials with four items in the sequence of each trial. The repeated and unique conditions (blocked, counterbalanced) had 180 trials each, from which 100 were old trials and 80 were new trials. In the old trials, in which the test probe appeared in the sequence, it was randomly and evenly sampled from all positions of the sequence (1–4). For the repeated condition, a set of 12 images was randomly sampled for each participant and was used throughout this condition. For practice, at the be-


*Figure 3.* Illustrations of the houses set (a) that was used in Experiments 2 and 4 (taken from Ahmed & Moustafa, 2016), and of the faces set (b) that was used in Experiment 3 (taken from Bainbridge et al., 2013). This figure is used with permission. See the online article for the color version of this figure.

ginning of each block, participants completed eight trials from the condition of that block.

**Color change-detection task.** To assess their VWM capacity, participants completed this task (Allon & Luria, 2017) prior to the RUP. During the task, four or eight colored squares appeared on a gray screen for 150 ms (memory display), followed by a retention interval of 900 ms. Then, one colored square (test probe) appeared at a location of one of the squares from the memory display and remained on the screen until a response was made. Participants had to decide whether the color of the test probe was the same ("Z" key) or different ("\" key) from the color of the square presented in that location in the memory display. On different trials, the changed square had a color that was not a part of the memory display. The task included 120 trials, randomly and evenly divided into two set sizes (four, eight) and two correct responses (old, new). Each square subtended approximately 1.24° × 1.24° of visual angle and randomly positioned within a 16.6° × 16.6° region, with a minimum distance of 2° between two stimuli (center to center). Colors were randomly selected from a set of nine colors (blue, red, yellow, green, black, brown, orange, pink and cyan). VWM capacity was separately computed for each set size by the standard formula K = S(H – F), in which K is the memory capacity, S is the set size, H is the observed hit rate and F is the false alarm rate (Pashler, 1988).

## Results and Discussion

No participants were dropped from the analysis due to the prespecified criteria. A paired-samples $t$ test indicated that participants memory was better in the unique ($M$ = 70.53, 95% CI [68.75, 72.3]) compared to the repeated condition ($M$ = 68.51, 95% CI [66.58, 70.45]), $t(39) = 2.63$, $p = .01$, $d = .42$, and these results can be seen in Figure 2.

**Difference between homogeneous and heterogeneous sets.** To evaluate the difference between homogeneous and heterogeneous sets, a between-experiments analysis was performed, using only set size of four from Experiment 1 as a baseline for the heterogeneous set. To account for the different baseline of the experiments, the PI score measure was used as the dependent variable. This variable represents the percent of which the performance in the unique condition differed from that of the repeated condition. An independent $t$ test revealed a significant difference in that the PI effect was markedly larger for Experiment 1 ($M$ = 14.95%, 95% CI [10.13, 19.77]) compared to Experiment 2 ($M$ = 3.28%, 95% CI [0.92, 5.64]), $t(76) = 4.47$, $p < .001$, $d = 1.01$. Thus, there was a significant PI effect when a homogeneous set was used, but this effect was substantially smaller compared to the effect that was found in Experiment 1, in which a heterogeneous set was used.

**VWM capacity influence on PI.** The memory capacity of each participant was calculated using her performance in the color-change-detection task (for both set sizes combined). First, we tested if this estimated capacity can predict the overall performance of participants in the RUP, two regression models revealed that participants performance in the RUP could be predicted using their estimated memory capacity in the repeated, $F(1, 38) = .8.34$, $p = .006$, $R^2 = .16$, $\eta_p^2 = .18$, and in the unique condition, $F(1, 38) = 6.31$, $p = .02$, $R^2 = .12$, $\eta_p^2 = .14$. Dividing participants into low and high span individuals based on the median of the VWM estimated capacity ($Mdn$ = 2.4) produced similar results, as the high-span individuals' performance was better than that of low-span individuals in the repeated, $t(38) = 2.97$, $p = .005$, $d = .94$, as well as in the unique condition, $t(38) = 2.37$, $p = .02$, $d = .75$ (summary statistics are presented in Table 1).

Next, and more important for the goals of the current study, we tested if participants' estimated capacity can predict the magnitude of the PI effect. However, a regression model with the VWM capacity as the input variable and the PI score (the percent of the difference between the conditions) as the output variable was insignificant, $F(1, 38) = .4$, $p = .53$. Furthermore, an additional $t$ test found no difference in the PI score between low-span and high-span individuals, $t(38) = .85$, $p = .4$ (see Table 1).

Given the relatively small number of participants for these analyses, it is natural that one should take these findings with cautious. Indeed, in the next experiment, only the lack of a significant connection between capacity and PI was replicated. Thus, these issues are further discussed in the General Discussion in which we combined the results of both experiments.

Table 1
*Means (Standard Deviations) of the Accuracy Rates and the Proactive Interference (PI) Scores in the Repeated-Unique Procedure*

| | | | Repeated (accuracy rate) | | | Unique (accuracy rate) | | | PI score | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Experiment | $n$ | $Mdn$ | Low span | $p$ | High span | Low span | $p$ | High span | Low span | $p$ | High span |
| 2 | 40 | 2.4 | 65.92 (5.21) | ** | 71.11 (5.83) | 68.56 (5.43) | * | 72.5 (5.06) | 4.28 (7.91) | >.1 | 2.28 (6.87) |
| 3 | 40 | 2.23 | 74.02 (5.37) | >.1 | 75.23 (5.3) | 76.93 (4.86) | >.1 | 76.55 (6.93) | 4.48 (10.43) | >.1 | 2.02 (9.51) |
| 2–3 | 80 | 2.33 | 70.19 (6.68) | .06 | 72.92 (5.95) | 72.88 (6.7) | >.1 | 74.4 (6.24) | 4.24 (9.24) | >.1 | 2.36 (8.15) |

*Note.* The $p$ columns show the significance of the two-sample $t$ tests that compared between individuals with low and high memory span, as was measured in the color-change-detection task.
* $p < .05$. ** $p < .01$.

**Serial position.** To test the role of the serial position on participant's performance, a repeated-measures ANOVA was conducted with the correct response rate as the dependent variable and the serial position and the experimental condition as the factors. This analysis included only the old trials (because only in these trials the test probe appeared at a specific position) and revealed a significant effect for the serial position, $F(3, 117) = 100.05$, $p < .001$, $\eta_p^2 = .72$. As can be seen in Figure 4a, the source of this effect was a superior performance for the last position compared to all other positions. This was true for both conditions, as there was no effect of the experimental condition, $F(1, 39) = .36$, $p = .55$, or of an interaction between these factors, $F(3, 117) = .45$, $p = .71$. Importantly, this means that the PI effect was not affected by the item's position in the series, suggesting, instead, that the small PI effect observed in this experiment was driven mainly by new trials (in which the probe was new in the current trial, but was presented in previous trials in the repeated condition). Indeed, a direct comparison between the conditions in the new trials revealed better performance in the unique ($M = 70$, 95% CI [66.58, 73.42]) relative to the repeated condition ($M = 66.47$, 95% CI [62.65, 70.28]), $t(39)$, $p = .04$, $d = .34$, while no differences between the conditions was found in the old trials (unique, $M = 70.95$, 95% CI [67.92, 73.98]; repeated, $M = 70.15$, 95% CI [67.28, 73.02]); $t(39) = .6$, $p = .55$.

## Experiment 3

Experiment 2 revealed a PI effect when a homogenous set was used, but this effect was significantly smaller compared to Experiment 1 in which a heterogeneous set was used. The goal of Experiment 3 was to replicate and generalize the finding that PI is reduced when a homogenous set of objects is tested. To this end, we repeated Experiment 2's design using images of gray-scale faces instead of colored images of houses. The logic was the same as before because homogenous set of faces, like houses, reduces the visual heterogeneity among the items and therefore participants ought to encode subtle visual characteristics of the stimuli to perform the task properly. Importantly, we expected that using the faces instead of the houses should further reduce the visual heterogeneity of the items. This is because the faces had fewer discriminating visual features—they were monochromatic instead of colorful and included only the faces themselves without any background features (see Figure 3). Furthermore, while one might argue that houses can be easily subcategorized based on some semantic features that are present in some houses but not in others (e.g., brick walls, roofing tile, chimney, etc.), such semantic subcategorization seems to be harder with faces. As a result, it was highly unlikely that participants utilized major visual differences or semantical categorization during the task.

### Method

The experiment was identical to Experiment 2 but for the stimuli that were 850 grayscale pictures of faces (4.5° × 4.5°; taken from Bainbridge, Isola, & Oliva, 2013; Figure 3b). Forty-four students (10 male; range = 19–34 years old; $M_{age} = 25.48$) completed the experiment.

### Results and Discussion

One participant was excluded from the analysis due to an overall performance that did not differ from chance level (50%). Three additional participants were excluded because their performance was lower (2 *SD*) from the overall mean. A paired-samples *t* test indicated that participants' memory was better in the unique condition ($M = 76.75$, 95% CI [74.88, 78.62]) compared to the repeated condition ($M = 74.6$, 95% CI [72.9, 76.3]), but this effect was significant only under a one-tailed hypothesis, $t(39) = 1.93$, $p = .06$, $p$(one-sided) = .03, $d = .31$ (see Figure 2).

**Difference between homogeneous and heterogeneous sets.** Again, we used set size of four from Experiment 1 as the baseline for the heterogeneous set and the PI score as the dependent variable. An independent-samples *t* test revealed a larger PI effect in the heterogeneous ($M = 14.95$, 95% CI [10.13, 19.77]) compared to the homogeneous set ($M = 3.31$, 95% CI [0.13, 6.49]), $t(76) = 4.12$, $p < .001$, $d = .93$. Hence, these results replicated those of Experiment 2—there was a PI effect for the homogenous set of faces but this effect was substantially smaller compared to the one found in the heterogeneous set of real-world objects.

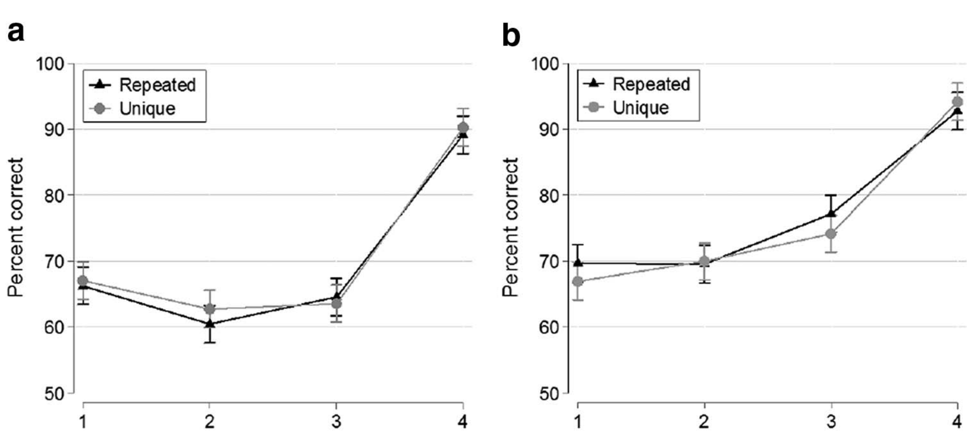


*Figure 4.* Predictive margins with 95% confidence intervals of the accuracy rates in (a) Experiment 2 and (b) Experiment 3, pending on the experimental condition and the serial position of the test item (old trials only).

**VWM capacity influence on PI.** As in the previous experiment, we tested if participants' performance in the RUP can be predicted by their memory capacity, as was measured by the color-change-detection task (for both set sizes combined). However, no connections were found between the variables (see Table 1), as participants' performance in the RUP could not be predicted from their estimated memory capacity, repeated condition, $F(1, 38) = 1.02$, $p = .32$; unique condition, $F(1, 38) = .58$, $p = .45$. Similarly, there was no performance differences between high-span and low-span ($Mdn = 2.23$) individuals, repeated condition, $t(38) = .71$, $p = .48$; unique condition, $t(38) = .2$, $p = .84$. Importantly, this was true also in regards to the PI effect as participants' estimated capacity could not predict their PI score, $F(1, 38) = .03$, $p = .86$, and the PI score was not different for high- and low-span individuals, $t(38) = (–.78)$, $p = .44$.

**Serial position.** A repeated-measures ANOVA with the correct responses rate as the dependent variable revealed a significant effect for the serial position, $F(3, 117) = 120.45$, $p < .001$, $\eta_p^2 = .76$. As can be seen in Figure 4b, the source of the serial-position effect on the correct response rate is that participant's accuracy rate improved as the serial position progressed. Once again, however, there was no effect for the experimental condition, $F(1, 39) = .4$, $p = .53$, nor an interaction between the two factors, $F(3, 117) = 1.18$, $p = .32$. Hence, as in Experiment 2, the PI effect was not affected by the items serial position and was driven mainly by new trials. Indeed, a comparison between the conditions revealed better performance in the unique ($M = 77.25$, 95% CI [73.73, 80.77]) relative to the repeated condition ($M = 71.19$, 95% CI [67.62, 74.76]) in the new trials, $t(39) = 3.05$, $p = .004$, $d = .48$, but no differences between the conditions in the old trials (unique, $M = 76.35$, 95% CI [72.65, 80.05]; repeated, $M = 77.33$, 95% CI [74.73, 79.92]), $t(39) = -.64$, $p = .53$. Notably, this pattern of results, namely, this difference in the PI effect between old and new trials, was not consistently observed across all experiments (see Section 2 in the online supplemental materials), and thus one cannot draw firm conclusions from these findings.

## Experiment 4

The goal of Experiment 4 was to use a different design to further replicate and generalize the previous findings that a smaller yet significant PI effect is present when a homogenous set of items is used. We aimed to design an experiment that will be more similar to the one that was used by Endress and Potter (2014), and to this end, the following changes were made. The crucial independent variable (repeated/unique) was a between-subjects instead of a within-subjects variable, while the type of stimuli was a within-subjects variable. This design allowed us to compare the correct response rates between the heterogeneous and the homogeneous sets without relying on the PI score measure to account for the different baselines of these sets. Additionally, the items for the repeated condition were sampled from a set of 22 items and a large set size of 10 items was used for the memory display to test if we can find high capacity estimations ($K$) for both homogeneous and heterogeneous set of items.

### Method

**Participants, stimuli, and equipment.** Forty students (15 male; 21–35 years old; $M_{\text{age}} = 24.98$) with a normal or corrected-to-normal vision completed the experiment for the pay of ₪20 (about US$5). The equipment and stimuli were identical to those in Experiment 2.

**Procedure and design.** Participants underwent a version of the RUP in which the crucial condition was a between-subjects condition. That is, half of the participants completed the unique condition while the other half completed the repeated condition, which included 22 repeated items for each stimuli type. The experiment included a total of 240 trials with a trial sequence that was identical to that of Experiment 2 but with 10 to-be-remembered items. The trials were evenly divided into two blocks of stimuli types (heterogeneous: objects, 5.6° × 5.6°; homogeneous: houses, 6.7° × 4.5°) that appeared in a counterbalanced order. The trials of each block were evenly and randomly divided into two correct responses (old, new). In the old trials, the test probe was randomly and evenly sampled from Positions 2–3, 5–6, and 8–9 of the sequence. All other parameters of the experiment were identical to those of Experiment 2.

### Results and Discussion

Two participants were excluded from the analysis due to an overall performance that did not differ from chance level (50%) and no participants were excluded due to a performance that was lower (2 *SD*) from the overall mean.

A mixed ANOVA with the experimental condition (repeated, unique) as a between-subjects factor and the type of stimuli (houses, objects) as a within-subject factor was conducted. Not surprisingly, there was a main effect of the type of stimuli, $F(1, 36) = 175.74$, $p < .001$, $\eta_p^2 = .83$, as memory performance for objects ($M = 76.18$, 95% CI [73.39, 78.98]) was much better than memory for houses ($M = 58.9$, 95% CI [56.81, 60.99]). There was also a main effect of the experimental condition, $F(1, 36) = 12.86$, $p = .001$, $\eta_p^2 = .26$, as performance in the unique condition ($M = 70.66$, 95% CI [66.41, 74.9]) was better than the repeated condition ($M = 64.43$, 95% CI [61.5, 67.36]). More importantly, there was a significant interaction between the experimental condition and the type of stimuli, $F(1, 36) = 6.54$, $p = .01$, $\eta_p^2 = .15$. The source of the interaction was a significant PI effect in the heterogeneous set, as the mean of the correct responses was higher in the unique ($M = 80.96$, 95% CI [76.92, 85.01]) compared to the repeated condition ($M = 71.4$, 95% CI [68.76, 74]), $F(1, 72) = 19.39$, $p < .001$, $\eta_p^2 = .21$, but no significant PI effect in the homogeneous set, even though the mean of the unique condition ($M = 60.35$, 95% CI [57, 63.7]) was numerically higher from that of the repeated condition ($M = 57.46$, 95% CI [54.8, 60.11]), $F(1, 72) = 1.78$, $p = .19$.

The PI score measure that was used earlier to compare the PI effect between experiments, could not be computed here because PI could not be measured within participants (as participants did not complete both the repeated and unique conditions). Instead, to compare the benefit of the unique condition regardless of the baseline difficulty level (namely, that the houses were more difficult than the objects), we calculated the percent in which participant's performance with the objects differed from their performance with the houses. A two-sample $t$ test revealed that the performance with objects was better than with houses and that this difference was larger in the unique ($M = 34.96\%$, $SD = 13.94$) compared to the repeated ($M = 25.66\%$, $SD = 18.03$) condition,

$t(36) = 1.78$, $p = .04$ (one-sided), $d = .58$. Hence, as predicted, the unique condition benefited more from the heterogeneity of the set compared to the repeated condition.

As in all experiments, the estimated memory capacity was calculated as well (see Section 1 in the online supplemental materials), but only in this experiment, this calculation yielded a large $k$ of 6.19 (95% CI [5.38, 7]). This datum was observed in the unique condition when the heterogeneous set was used. A one-sample $t$ test confirmed that this $k$ is larger compared to traditional estimations of four items (e.g., Cowan, 2001), $t(18) = 5.69$, $p < .001$, $d = 1.3$, and also compared to the capacity estimation of 4.7 for meaningful items that was reported by Brady et al. (2016), $t(18) = 3.88$, $p = .001$, $d = .89$. Moreover, this datum is in line and even higher compared to the estimations of Endress and Potter (2014) for similar set sizes ($k_{\text{set size 11}} = 4.9$; $k_{\text{set size 12}} = 5.4$). Most importantly, the $k$ was significantly smaller in the unique condition when the homogeneous set was used ($k = 2.07$, 95% CI [1.4, 2.74]); $t(36) = -8.25$, $p < .001$. This suggests that testing meaningful objects in a PI-free condition is not sufficient to yield high memory capacity, but rather, that the heterogeneity of the memory set is a critical factor for both PI in the RUP and for increasing capacity limits. The calculated $k$ in the repeated condition stayed in the range of ≤4, repeated objects' $k = 4.28$, $t(18) = 1.12$ $p = .28$, repeated houses' $k = 1.49$, $t(18) = -9.92$, $p < .001$.

To summarize, as expected, performing a visual memory task with a heterogeneous set was easier compared to a homogeneous set. Importantly, however, a robust PI effect was detected for the heterogonous set of items even though the experimental condition was a between-subjects variable, but the effect failed to reach statistical significance for the homogenous set. This finding is line with our previous experiments showing that even though the PI might not be eliminated, it is greatly reduced when the remembered items are meaningful, yet less visually discriminable.

## General Discussion

There is a widespread agreement that VWM has a very limited capacity (Cowan, 2001; Luck & Vogel, 1997; Matsukura & Hollingworth, 2011). This limited capacity highlights the need for VWM to be highly efficient to function effectively. Accordingly, it has been argued that VWM is not considerably affected by PI (Lin & Luck, 2012; Oberauer et al., 2017; Shipstead & Engle, 2013), suggesting that items that are no longer needed are effectively removed from memory. This notion was recently challenged by Endress and Potter (2014), who devised a novel procedure to test the influence of PI in VWM. Using this procedure, it was found that VWM is severely impaired by PI and these results were further taken as evidence for the existence of a new type of memory entitled "temporary memory." Indeed, having a short-term memory buffer that holds more than three to four items and fits well with our intuition that many real-life situations, as the researchers argued, requires high capacity for meaningful items. Accordingly, their results indicated that applying the formula for calculating the memory capacity in a PI-free condition resulted in what seems to be an unbounded capacity for real-world meaningful objects.

Given the highly important theoretical implications of these conclusions (Endress & Potter, 2014; Endress & Siddique, 2016), the present research aimed to examine whether the large gap between the standard VWM tasks and Endress and Potter's study regarding the role of PI in VWM can be accounted for without assuming a novel memory buffer. Specifically, the experiments were designed to test a notable difference between the RUP and more standard change-detection tasks, which is the heterogeneity of the stimuli. Experiment 1 replicated the basic finding with a heterogeneous set of items in showing that performance in the unique condition was much better compared to the repeated condition. It further ruled out the possibility that the effect is limited to the use of a relatively large set of repeated objects as the effect was found using 13 rather than 22 repeated objects.

The next experiments used homogenous sets of items, as opposed to the heterogeneous set of real-world objects that were used in the first experiment and in previous research (Endress & Potter, 2014; Endress & Siddique, 2016; Makovski, 2016). This change forced participants to encode subtle visual differences and to rely on them to complete the task. Additionally, it impaired their ability to use semantic categorization information because all the stimuli belonged to a single category. Experiments 2 and 3 used a within-subjects design and revealed a significant PI effect even when homogenous sets of houses and faces were tested. Importantly, however, this effect was smaller compared to the effect that was found for the heterogeneous set of real-world objects. The latter finding was also repeated in the final experiment that used a between-subjects design and found a reduced, insignificant PI effect when the homogenous set of houses was tested relative to the robust PI effect found for the heterogeneous set of real-world objects. Together, these results suggest that the visual heterogeneity of the items plays a critical role in modulating the role of PI as tested in the RUP.

The findings of this study imply that there is no need to assume a new memory buffer that specialized in meaningful information (e.g., temporary memory; Endress & Potter, 2014) to explain the large discrepancy regarding the role of PI in VWM tasks. Instead, our results suggest two independent possibilities that can account for this discrepancy: (a) that VWM is much less rigid than what was previously assumed and (b) that the involvement of LTM in the RUP contributed to the large PI effect. Specifically, regarding the first possibility, two characteristics of the VWM seems less definite in light of the recent findings—its capacity limitation and its resistance to PI. First, VWM capacity is indeed limited, as we cannot remember all the visual information that we encounter, but it seems that this limitation is not fixed at a single number. Instead, our memory span is highly dependent on the type of stimuli that we try to remember (e.g., Brady et al., 2016) and in the presence or absence of PI. Second, VWM is more vulnerable to PI than previously thought (Lin & Luck, 2012; Oberauer et al., 2017; Shipstead & Engle, 2013), but this vulnerability is a function of the task characteristics, such as the spatial presentation of the stimuli, as shown by Makovski (2016). The current findings further suggest that the vulnerability to PI is changing as a function of the heterogeneity of the set. Homogeneous sets of meaningful stimuli are prone to PI but the magnitude of this interference is smaller compared to the interference in a heterogeneous set of meaningful stimuli. Independently, it is possible that the RUP enables the involvement of LTM, and specifically of semantic categorization. This involvement might contribute to memory performance the most when a unique set of real-world objects is used. This is because in the unique condition, no confusion with previously acquired labels is possible and because it is easier to categorically label real-world objects compared to houses and faces.

It is important to note that although the present findings suggest that one need not assume a novel memory buffer to explain PI, an explanation might still be needed to account for the large capacity estimates found when a heterogeneous set of objects were tested in a unique condition (Endress & Potter, 2014). Indeed, our results converged with these findings as they revealed a large *k* when participants had to remember 10 real-world objects in the PI-free (i.e., unique) condition of Experiment 4. In contrast, when 10 houses were to be remembered, the estimated capacity *k* was not larger than the traditional estimates, even in the unique condition. Together, these results imply that using meaningful stimuli is not sufficient to produce large capacity estimations, but rather, that the heterogeneity of the set is a crucial factor for this increased capacity. Nevertheless, additional research is needed to clarify whether the findings of large *k*s were due to the involvement of a unique temporary memory buffer, due to the involvement of LTM in the task, or due to other characteristics of the RUP. That is, it is possible that the good memory performance under these conditions was due to the meaning of the objects that increased the involvement of LTM in the task, the visual distinctiveness among the memory items that reduced the interitem interference within a trial, the large visual distinctiveness between the memory items and the probe, or some combinations of the above.

## The Role of Distinctiveness in Proactive Interference

Why did heterogeneity matter for PI as was measured by the RUP? As mentioned in the Introduction, heterogeneous and homogenous sets differ in two important ways. First, in a heterogeneous set, there are more visual differences between the items and it is easier to discriminate among them compared to a more homogenous set. Second, in a heterogeneous set of meaningful items, it is possible to encode the items based on semantical categorization and this categorization is less useful when all the items in a set are drawn from a single category. These differences are not mutually exclusive as the semantic label of a stimulus might be another feature that helps to distinguish it from other stimuli. Accordingly, our manipulation of changing the heterogeneous to homogeneous sets affected both the visual differences between the stimuli and the possibility to assign distinguishing labels to them.

Importantly, both of these factors likely contributed to the fact that the distinctiveness of the real-world objects was higher compared to the homogeneous sets of houses and faces, and this distinctiveness made it possible to perform the task with relatively shallow visual encoding and without focusing on the finer details of each stimulus. Naturally, the unique condition benefited from this distinctiveness more than the repeated condition because in the latter the items' distinctiveness across trials was lower. In contrast, when the homogeneous sets were used, a finer visual encoding was necessary to perform the task, but still, the performance in the unique condition was better for the same reason.

## The *k* Measure Connections to the RUP and PI

The results of the current study and converging evidence suggest that the *k* measure is dependent on the nature of the stimuli that one need to remember. The inclusion of a color-change-detection task (Allon & Luria, 2017) in Experiments 2 and 3 enabled us to further examine the relationship between the standard *k* measure and the overall performance in the RUP. However, only in one of these experiments, the *k* predicted participants' performance in the RUP (in both the repeated and unique conditions). However, it might be the case that the small sample size in these experiments hindered our ability to detect an influence of the *k* on the RUP. For that reason, we combined the results of Experiments 2 and 3 and tested if, in a larger sample ($n = 80$), the *k* will predict participants' performance in the RUP. In this sample, the *k* predicted participants performance in the repeated condition of the RUP, $F(1, 78) = 5.74$, $p = .02$, $R^2 = .06$, $\eta_p^2 = .07$, but had only a marginal predictive power in the unique condition, $F(1, 78) = 3.67$, $p = .06$, $R^2 = .03$, $\eta_p^2 = .04$. Dividing participants into low and high-span individuals based on the *k*'s median ($Mdn = 2.33$) also revealed only a marginal difference between the groups in the repeated condition, $t(78) = 1.92$, $p = .06$, and no difference between the groups in the unique condition, $t(78) = 1.06$, $p = .29$ (see Table 1). These findings suggest that the connections between the *k* measure and the performance in another short-term visual memory task are not as strong as what one might expect. However, more research is needed to fully characterize the connections between the estimated capacity *k* and the RUP.

Of greater interest for the current purposes, the results of the two experiments converged in showing no connection between *k* and the magnitude of the PI. Participants' *k* could not predict their PI score, $F(1, 78) = .28$, $p = .6$, and the PI score was not different for high and low-span individuals, $t(78) = .96$, $p = .34$ (see Table 1). These results contradict the notion that high-span individuals are less prone to PI (Kane & Engle, 2000) and that the ability to resolve PI is what separates between low and high-span individuals (Bunting, 2006). Yet, it is important to note that the PI effect in these experiments was calculated as a difference between blocks (rather than between trials) and that the overall PI was small, which might hinder our ability to detect a reliable connection between *k* and PI.

It is important to acknowledge some limitations of the study. Two noticeable and interrelated limitations are the need to compare the results of different experiments and the need to account for different performance baselines for different stimuli types. As our focus was the PI effect as a function of different stimuli sets, these comparisons were necessary. To enable comparisons across experiments while accounting for the different performance baselines, we used the PI score measure which is the percent of the difference between the unique and repeated conditions. Due to this transformation, the PI score reflects the difference in the magnitude of PI in different stimuli types while, at least partly, controlling for the overall difficulty of the condition. Still, this method is not ideal. In Experiment 4 the type of the stimuli was a within-subject condition and this enabled us to directly compare participants' accuracy rates with different types of stimuli without relying on the PI score. In addition, we used the percent of the difference between the heterogeneous and homogeneous sets and compared this difference in the repeated and unique conditions. Importantly, the result of Experiment 4 converged with the between-experiments comparisons of Experiments 1–3 into a single pattern, and hence, we believe that these limitations are not harming the conclusiveness of its results.

## Conclusions

To conclude, our results show that VWM is vulnerable to PI, but the magnitude of this vulnerability is not fixed. Instead, it is a function of the methodological characteristics, such as the stimuli spatial location (Makovski, 2016), and, as this study revealed, the heterogeneity of the to-be-remembered set of stimuli. This heterogeneity leads to more distinctiveness (visual and/or semantic) among the items that, in turn, can lead to high memory capacity and significant PI effect without assuming new memory storage. On the one hand, these findings suggest that previous estimations of memory capacity that were based on the change-detection task were not severely biased as the stimuli in these tasks are usually homogeneous and appearing at distinct locations. On the other hand, these findings suggest that the VWM's characteristics, and specifically its capacity and vulnerability to PI, might not be strictly rigid as was previously assumed.

Although the existence of separate memory storage for meaningful information (i.e., temporary memory) is possible, the current study suggests that there is no need to assume such storage. Instead, both the PI effect and the remarkably large memory capacity, which were reported by Endress and colleagues (Endress & Potter, 2014; Endress & Siddique, 2016) can be explained solely by the VWM storage or by an interaction between VWM and LTM. With a sufficient amount of visual and semantic distinctiveness among the items, the task can be performed without the need to encode fine visual features. This is particularly pronounced in the unique condition, which subsequently leads to increased memory capacity and to a significant PI effect as is measured by the difference between the unique and repeated conditions.

## References

Ahmed, E., & Moustafa, M. (2016). House price estimation from visual and textual features. In J. J. Merelo, F. Melício, J. M. Cadenas, A. Dourado, K. Madani, A. Ruano, & J. Filipe (Eds.), *Proceedings of the 8th International joint conference on computational intelligence* (pp. 62–68). Setúbal, Portugal: SCITEPRESS.

Allon, A. S., & Luria, R. (2017). Compensation mechanisms that improve distractor filtering are short-lived. *Cognition, 164,* 74–86. http://dx.doi.org/10.1016/j.cognition.2017.03.020

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29. http://dx.doi.org/10.1146/annurev-psych-120710-100422

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142,* 1323–1334. http://dx.doi.org/10.1037/a0033872

Blalock, L. D. (2015). Stimulus familiarity improves consolidation of visual working memory representations. *Attention, Perception & Psychophysics, 77,* 1143–1158. http://dx.doi.org/10.3758/s13414-014-0823-z

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 4. http://dx.doi.org/10.1167/11.5.4

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, 105,* 14325–14329. http://dx.doi.org/10.1073/pnas.0803390105

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 183–196. http://dx.doi.org/10.1037/0278-7393.32.2.183

Conway, A. R. A., Kane, M. J., & Al, C. E. T. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12,* 769–786.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114. http://dx.doi.org/10.1017/S0140525X01003922

Craik, F. I. M., & Birtwistle, J. (1971). Proactive inhibition in free recall. *Journal of Experimental Psychology, 91,* 120–123. http://dx.doi.org/10.1037/h0031835

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General, 143,* 548–565. http://dx.doi.org/10.1037/a0033934

Endress, A. D., & Siddique, A. (2016). The cost of proactive interference is constant across presentation conditions. *Acta Psychologica, 170,* 186–194. http://dx.doi.org/10.1016/j.actpsy.2016.08.001

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences, 105,* 9926–9930. http://dx.doi.org/10.1073/pnas.0709884105

Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. *PLoS ONE, 3,* e2716.

Jackson, M. C., & Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 556–568. http://dx.doi.org/10.1037/0096-1523.34.3.556

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358. http://dx.doi.org/10.1037/0278-7393.26.2.336

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153–161. http://dx.doi.org/10.1016/S0022-5371(62)80023-1

Lin, P., & Luck, S. J. (2012). Proactive interference does not meaningfully distort visual working memory capacity estimates in the canonical change detection task. *Frontiers in Psychology, 3,* 42. http://dx.doi.org/10.3389/fpsyg.2012.00042

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. http://dx.doi.org/10.1038/36846

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

Makovski, T. (2016). Does proactive interference play a significant role in visual working memory tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1664–1672. http://dx.doi.org/10.1037/xlm0000262

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory & Cognition, 36,* 43–52. http://dx.doi.org/10.3758/MC.36.1.43

Matsukura, M., & Hollingworth, A. (2011). Does visual short-term memory have a high-capacity stage? *Psychonomic Bulletin & Review, 18,* 1098–1104. http://dx.doi.org/10.3758/s13423-011-0153-2

Ngiam, W. X., Khaw, K. L., Holcombe, A. O., & Goodbourn, P. T. (2018). Visual working memory for letters varies with familiarity but not complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. http://dx.doi.org/10.1037/xlm0000682

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1–22. http://dx.doi.org/10.1037/xlm0000302

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44,* 369–378. http://dx.doi.org/10.3758/BF03210419

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198. http://dx.doi.org/10.1037/h0049234

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16,* 283–290. http://dx.doi.org/10.3758/BF03203943

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

Shipstead, Z., & Engle, R. W. (2013). Interference within the focus of attention: Working memory tasks reflect more than temporary maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 277–289. http://dx.doi.org/10.1037/a0028467

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74,* 1–29. http://dx.doi.org/10.1037/h0093759

Viswanathan, S., Perl, D. R., Visscher, K. M., Kahana, M. J., & Sekuler, R. (2010). Homogeneity computation: How interitem similarity in visual short-term memory alters recognition. *Psychonomic Bulletin & Review, 17,* 59–65. http://dx.doi.org/10.3758/PBR.17.1.59

Received April 16, 2019
Revision received August 5, 2019
Accepted September 2, 2019 ■