# Meaningful stimuli inflate the role of proactive interference in visual working memory

Roy Shoval[1] · Tal Makovski[1]

## Abstract

The use of meaningful daily objects in visual working memory (VWM) tasks revealed two uncharacteristic findings: enlarged memory capacity, and strong proactive interference (PI), which was previously believed to play only a modest role in VWM. To disassociate the roles of meaning and visual complexity in these effects, a set of stimuli composed of meaningful daily objects was compared to visually similar meaningless sets. These sets were included in a Repeated (PI-prone) condition in which stimuli were repeatedly drawn from a limited set of items, and in a Unique (PI-free) condition in which each stimulus appeared only once. In line with past findings, the results consistently showed superior memory for meaningful stimuli. Importantly, they also showed a stronger PI-effect for meaningful stimuli as the difference between the Repeated and Unique conditions was greatly reduced (Experiment 1) or eliminated (Experiment 2) for meaningless stimuli. Together, these results strongly imply that meaning, and not visual complexity, plays a key role not only in boosting memory capacity but also in inflating the role of PI in VWM.

## Introduction

Visual working memory (VWM), the memory buffer that actively holds a limited amount of visual information for short durations (Phillips, 1974), is a fundamental and extensively studied cognitive construct. Much of this research has focused on the VWM capacity and the nature of its limit (Luck & Vogel, 2013; Ma et al., 2014). Another, somewhat related, line of research has focused on the interactions between VWM and long-term memory (LTM) and specifically on the role of proactive interference (PI) in VWM. PI is the process in which no longer relevant information acquired in previous trials interferes with current performance (Keppel & Underwood, 1962; Peterson & Peterson, 1959). Finding PI in VWM tasks would entail that LTM interferes with VWM performance and that VWM is less efficient than typically thought. Importantly, however, the occurrence and magnitude of PI in VWM are still debatable.

Most of the previous studies that tested PI in VWM compared participants' recognition of a stimulus depending on whether this stimulus appeared or did not appear in the previous trial (or few trials), and concluded that PI has only a modest role in limiting VWM capacity (Hartshorne, 2008; Lin & Luck, 2012; Makovski & Jiang, 2008; Oberauer et al., 2017). However, these studies used limited sets of stimuli that were repeating throughout the tasks, and hence a long-lasting PI build-up could occur, but a comparison to a truly PI-free condition was impossible. For that purpose, a dedicated paradigm was devised – the Repeated-Unique Paradigm (RUP; Endress & Potter, 2014).

In the RUP participants are viewing serial presentations of stimuli, followed by an old/new recognition test. Crucially, this paradigm compares the overall performance in a Repeated condition that includes a limited set of repeating stimuli to that of a Unique condition in which each stimulus appears only once and is therefore PI-free. Indeed, performance in this Unique condition was superior compared to the Repeated condition (Endress & Potter, 2014; Endress & Siddique, 2016; Shoval & Makovski, 2021). Moreover, using a large set-size of 100 stimuli per trial in the Unique condition revealed a memory capacity (known as K; Pashler, 1988) of 30 items – much higher than

✉ Roy Shoval
  rshoval@gmail.com

1   Department of Education and Psychology, The Open
    University of Israel, University Rd. 1, 4353701 Ra'anana,
    Israel

the upmost conventional estimations of four items (Cowan, 2001). These findings suggest that PI is a crucial factor that limits VWM capacity and that in the absence of PI this capacity might be unbounded.

Although the rationale for measuring the difference between the Unique and Repeated condition (hereafter PI-effect) seems valid, several factors that might influence this difference should be considered. For instance, it had been shown that spatial distinctiveness modulates the PI-effect that was reduced when stimuli were presented at different locations on the screen compared to when they appeared (sequentially) at the same location (Makovski, 2016). In addition, it was shown that heterogeneity – the distinctiveness of the stimuli within the set – modulates the magnitude of the PI-effect. PI was smaller when the stimuli were drawn from homogenous sets composed entirely of houses or faces compared to when a heterogeneous set of everyday objects was used (Shoval et al., 2020). Note, however, that these stimulus sets differ not only in their visual distinctiveness, but also in their semantic heterogeneity, and thus it is not clear whether and to what extent semantics plays a role in the PI-effect.

Indeed, there are good reasons to suggest that the type of stimuli, and specifically, their meaning, should also influence the PI-effect. Up to this point, only meaningful stimuli of real-world objects were used in the RUP. In contrast, the studies that used a single Repeated condition and showed a minor effect of PI in VWM typically used simple, impoverished stimuli such as color patches (e.g., Lin & Luck, 2012; Makovski & Jiang, 2008). There are certainly good justifications for testing both types of stimuli. Memorizing simple stimuli mostly involves low-level visual processing of a bottom-up nature, thus enabling a relatively pure characterizing of VWM with minimal involvement of higher-order mechanisms such as semantic categorization. Meaningful stimuli, conversely, provide higher ecological validity. Importantly, however, the difference between these sets can dramatically affect performance, as even the foremost level of initial neuronal processing of real-world stimuli differs from artificial ones in fundamental ways (Drew et al., 2018).

Consequently, superior memory for meaningful compared to meaningless stimuli was reported with both verbal (e.g., Besner & Davelaar, 1982; Hulme et al., 1991) and visual stimuli (Brady et al., 2016; Olsson & Poom, 2005; Sahar et al., 2020; but see Quinlan & Cohen, 2016). It was further found that VWM capacity for daily objects exceeded the utmost conventional estimation of four items, as a $k$ of ~4.7 was detected (Brady et al., 2016). Notably, daily objects differ from simple stimuli not only in their meaning but also in their visual complexity and the inter-item distinctiveness they provide. Yet, it has been recently shown that memory is enhanced for meaningful stimuli even when their visual properties are similar to those of meaningless stimuli (Asp

et al., 2021; Brady & Störmer, 2021; Conci et al., 2021; Sahar et al., 2020).

The difference between meaningful and meaningless stimuli may also affect the magnitude of the PI-effect because meaning can differently affect the Repeated and Unique conditions. One possibility is that the difference between the Repeated and Unique conditions would be greater with meaningful objects. This might be because meaningful items produce stronger memory representations that could enhance the inter-stimuli distinctiveness. This distinctiveness can elevate VWM performance, particularly in the Unique condition because when a small set of items is repeatedly presented (namely, in the Repeated condition), this distinctiveness is likely reduced (Shoval et al., 2020). Aside from the inter-stimuli distinctiveness, meaningful items could lead to more confusion in the Repeated condition, which would also result in a greater PI-effect, because their strong memory representations could linger in the system for a longer duration. In contrast, it is also possible that the difference between the Repeated and Unique conditions would be greater with meaningless, rather than meaningful, stimuli. This could occur because, for instance, meaningless objects that are poorly remembered are likely to cause more confusion during testing, resulting in even worse performance in the Repeated condition.

The current study aimed to examine the role of meaning in VWM tasks, and specifically its involvement in PI while controlling for the visual distinctiveness of the stimuli. To that end, the RUP (Endress & Potter, 2014) was employed with meaningful and meaningless stimuli with similar visual distinctiveness. If meaning is the main reason for the increased memory performance and PI in VWM, we would expect to find superior memory for Meaningful compared to Meaningless stimuli, and for this difference to be stronger in the Unique (PI-free) condition. Alternatively, if visual distinctiveness, and not meaning, is driving the exceptional findings of daily objects in VWM performance (i.e., increased capacity and PI), then similar findings should be found for Meaningless and Meaningful stimuli.

## Pilot experiments

We first tested two pilot experiments that employed both conditions of the RUP with meaningful and meaningless stimuli with similar visual properties. The Meaningful set included images of daily objects, whereas the Meaningless set included distorted versions of the same images. Specifically, one-half of each object from the Meaningful set was flipped. This manipulation largely kept the items' "objecthood," as well as most of the visual statistics of the meaningful stimuli (colors, size, brightness, etc.), while their meaning was largely distorted (Fig. 1b, the full stimuli set is publicly available at https://osf.
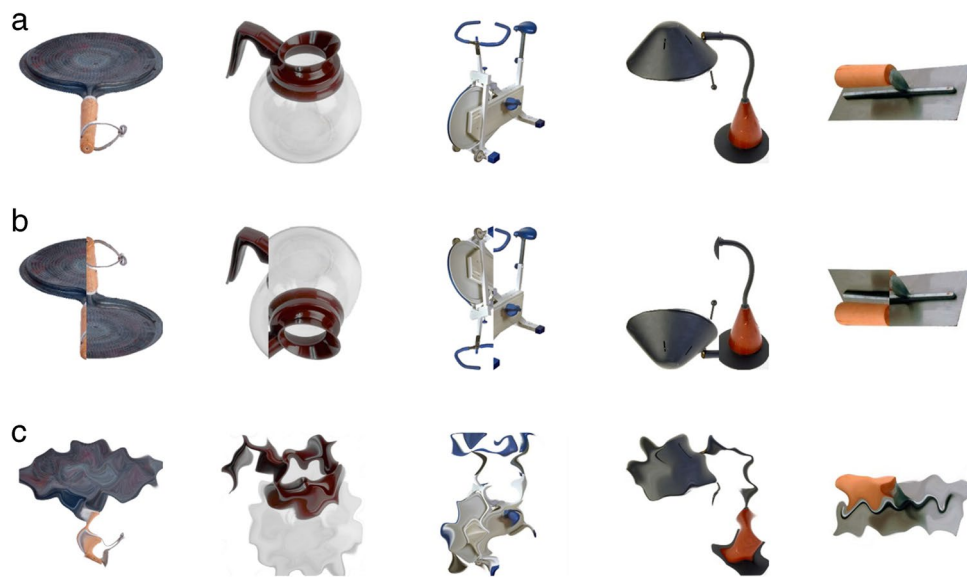
**Fig. 1** Examples from the Meaningful set that was used in both experiments (**a**), the Flipped Meaningless set that was used in Experiment 1 (**b**), and the Scrambled Meaningless set that was used in Experiment 2 (**c**)

io/ns8p3/). This distortion of meaning was verified in a previous study, as it was shown that compared to the Meaningful set, it is harder to verbally name stimuli from the Meaningless set that were also explicitly rated as less meaningful (Makovski, 2018; for a similar use of this meaning manipulation see also Brady & Störmer, 2021, and Sahar et al., 2020). Hence, by controlling for the visual distinctiveness of both stimuli sets, we can assess whether meaning, by itself, increases VWM capacity estimates as well as PI.

The two pilot experiments included a within-between design and a rather small set size of four items. Both pilots showed better performance with the Meaningful set and suggested that the PI-effect is stronger for meaningful stimuli (see S1 and S2 in the Online Supplemental Materials (OSM) for their full report). However, the statistical interaction between the Susceptibility to PI (Repeated/Unique) and Meaning was weak for some dependent variables in those experiments. This might be because of statistical noise produced by the mixed designs, because the tasks were quite easy, and because PI was small as the difference between the memory set size (four) and the repeating items pool (nine) was quite large.

## Experiment 1

Experiment 1 employed the Pilots' procedure, but to reduce statistical noise and make the task harder a full within-subjects design and a larger set-size were used. Additionally, the relative difference between the set-size and the size of the

repeating items pool was reduced to increase PI (Endress, 2022).

## Method

### Participants

In all experiments, participants were 18–35 years old, with normal or corrected-to-normal vision, normal color perception, and without any attentional, psychiatric, or neurological disorders. All experiments were approved by the Ethics Committee of the Open University of Israel, and each participant completed only one experiment. Forty individuals (25 males, mean age 25.78 years) participated in Experiment 1 for a payment of 30 New Shekels (~8.5 $). This sample provides a power of 0.85 to find a within-subject interaction effect with a size of 0.2 or larger.

### Equipment and stimuli

Participants were tested individually in a dimly lit room. They sat about 67 cm from a 23-in. LCD screen (resolution $1,920 \times 1,080$, 85 HZ). The experiment was programmed with MATLAB R2018a (www.mathworks.com) and Psychtoolbox 3.0.14 (Brainard, 1997). The Meaningful-set included 600 images ($1.89\,^{\circ}$ x $1.89\,^{\circ}$, taken from Brady et al., 2008), whereas the Meaningless set included the distorted versions of these images (Makovski, 2018; Fig. 1).
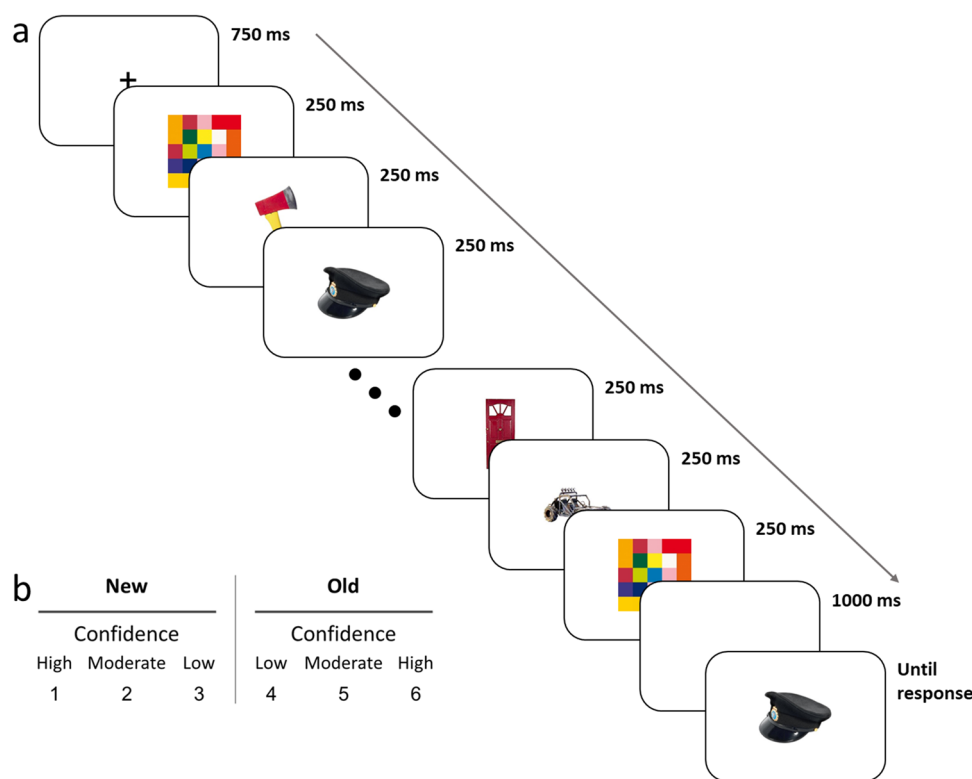
**Fig. 2** An illustration of an "old" trial sequence with Meaningful stimuli (**a**), and the confidence scale that appeared under the test probe (**b**)

## Procedure

Each trial began with a fixation cross (750 ms, 1.3°), followed by a colorful mask pattern (1.89° x 1.89°, 250 ms), the memory display, another mask image (250 ms), a blank screen (1,000 ms), and a test probe (Fig. 2). The memory display included eight images that appeared sequentially in the middle of the screen (250 ms per image), and when the test probe appeared participants had to indicate if this is an "old" image that appeared or a "new" image that did not appear at the preceding memory display.

To evaluate the memory sensitivity under different response criteria a receiver operating characteristics (ROC) procedure was used (Van Zandt, 2000; Yonelinas & Parks, 2007). Participants responded using a 1–6 scale (Fig. 2) depending on their degree of confidence, when 1, 2, 3 represented "new with high/medium/low confidence" (respectively), and 4, 5, 6 represented "old with low/medium/high confidence" (respectively). To minimize the use of verbal re-coding, participants repeated aloud a three-letter word. A new word was presented on inter-trials breaks that were given every 16 trials.

## Design

Each participant completed four within-subject blocks that included all combinations of the Susceptibility to PI and the Meaning conditions. The first (and last) two blocks were of the same Meaning condition (counterbalanced), and the same (counterbalanced) order of the Susceptibility to PI condition was used in each of the Meaning conditions (i.e., the Repeated block was the first in both the Meaningful and Meaningless conditions for half of the participants and the second for the other half).

Participants completed 64 trials in each block, evenly and randomly divided into two correct responses (old, new). The serial position in which "old" items had appeared in the memory display was randomly selected. In the Unique condition, each image appeared only once throughout the experiment while in the Repeated condition, two different sets of ten images were randomly selected for the Meaningful and Meaningless blocks. Note that the test probes in "new" trials of the Repeated condition were drawn from the repeating items pool but did not appear in the memory displays of that trial. For practice, each participant completed four trials that matched her first block in the experiment.
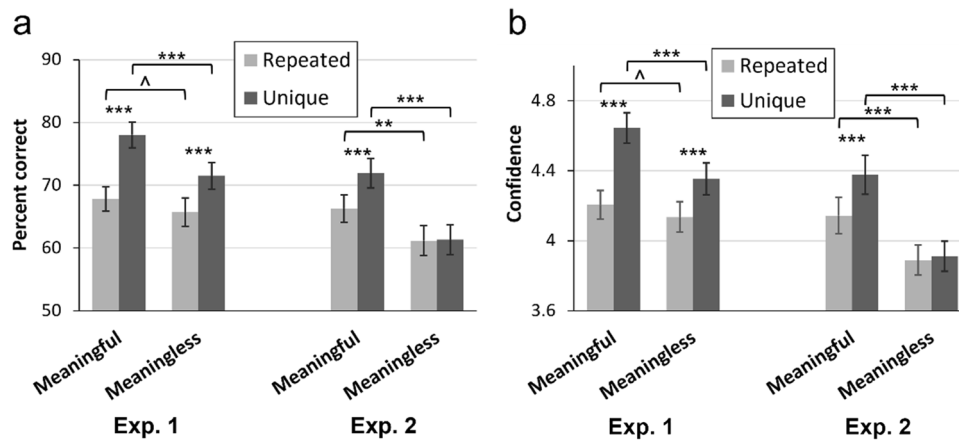
**Fig. 3** Means and 95% confidence intervals of the unweighted (**a**) and weighted (**b**) percent of correct responses, as a function of the stimulus type, Susceptibility to PI, and experiment. *** reflects $p < .001$ ** reflects $p < .01$, and ^ reflects $p < .1$

## Results

Two participants were excluded from the analysis because their performance was not better than chance (a one-sample test of proportion deemed indifferent from chance level, 50%; $z \leq 1.645$). Another participant was excluded due to performance that was lower (2 SD) than the overall mean. These exclusion criteria were set before the experiments and were used in our other studies as well (Shoval et al., 2020; Shoval & Makovski, 2021).

We first tested the percent of correct responses, independent of participants' confidence levels (i.e., options 1–3 from the confidence scale were coded as New, and options 4–6 were coded as Old). A repeated-measures ANOVA with the Meaning and Susceptibility to PI as factors revealed two main effects and a significant interaction (Meaning, $F(1,36) = 20.97$, $p < .001$, $\eta_p^2 = .39$; Susceptibility to PI, $F(1,36) = 82.84$, $p < .001$, $\eta_p^2 = .7$; Interaction, $F(1,36) = 6.87$, $p = .01$, $\eta_p^2 = .16$). As can be seen from the conditions means and 95% confidence interval (CI; Fig. 3a), performance was better in the Unique compared to the Repeated condition, and with the meaningful compared to the meaningless stimuli. Crucially, the source of the interaction was that the PI-effect was stronger in the Meaningful condition ($t(36) = 8.46$, $p < .001$, $d = 1.39$) than in the Meaningless condition ($t(36) = 4.75$, $p < .001$, $d = .78$). When integrating participants' confidence in their responses by assigning different weights for each 1–6 response, similar but slightly stronger effects were detected (Fig. 3b; the complete report of this analysis is presented in section S3 of the OSM).

To account for the possibility that our results were driven by different baseline performances for each type of stimuli, we used the percent in which the performance in the Repeated condition differed from that of the Unique condition ((Unique - Repeated)/Repeated × 100; see Shoval et al.,

2020; Shoval & Makovski, 2021). Analyzing this variable did not change the conclusions as the PI-effect with the meaningful stimuli (M = 15.72%, $CI_{95\%}$ = 11.67–19.78) was larger relative to the PI-effect of the meaningless stimuli (M = 9.55%, $CI_{95\%}$ = 5.52–13.59), t(36) = 2.21, $p = .03$, $d = 0.36$.

Finally, a ROC curve was created for each subject in each condition by plotting the accumulated hit rate (y-axis) against the accumulated false-alarm rate (x-axis) on each response criterion (i.e., the six levels of confidence that participants had in their response). Then, the area under each participant's curve was converted to the sensitivity measure $d'$ using Salgado's (2018) tables (Fig. 4). A repeated-measures ANOVA with this $d'$ as the dependent variable revealed, once again, two main effects and a significant interaction (Meaning, $F(1,36) = 23.08$, $p < .001$, $\eta_p^2 = .39$; Susceptibility to PI, $F(1,36) = 72.43$, $p < .001$, $\eta_p^2 = .67$; Interaction, $F(1,36) = 13.07$, $p < .001$, $\eta_p^2 = .27$; Fig. 4). This suggests that the results were driven primarily by different sensitivities and not by criterion shifting.

## Discussion

Consistent with previous findings, participants performed better with the Meaningful compared to the Meaningless stimuli (e.g., Brady et al., 2016), and in the Unique compared to the Repeated condition (e.g., Endress & Potter, 2014). Importantly and in line with the Pilot experiments, the interaction between these factors was significant in that the PI-effect was stronger in the Meaningful condition. These results were found even though verbal re-coding was reduced via articulatory suppression. In addition, the possibility that differences in task difficulty drove these results seems unlikely not only because the results held when using a relative measure (performance difference in percent), but
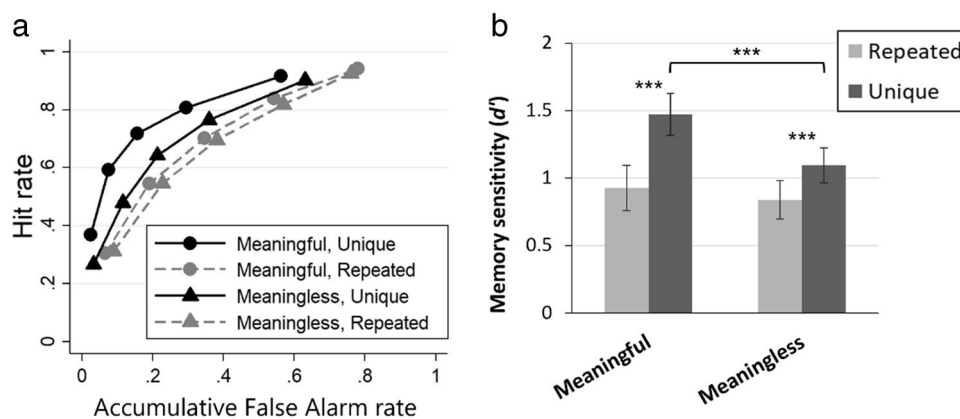
**Fig. 4** Receiver operating characteristic (ROC) curves of Experiment 1 as a function of the experimental conditions (**a**) and the means and 95% confidence intervals of the resulting $d'$ (**b**). *** reflects $p < .001$

also because former research indicated that task difficulty, as manipulated for example by different set sizes or different retention durations, did not affect the magnitude of the PI effect (Endress & Potter, 2014; Shoval et al., 2020; Shoval & Makovski, 2021). Therefore, because the visual distinctiveness was similar within both sets, we can conclude that meaning, and not visual distinctiveness, played a critical role both in improving VWM performance and in increasing the PI effect.

These results cannot be explained merely by the manipulation of forming the meaningless set. The claim, for instance, that the meaningless set required subjects to remember more object parts instead of a single integrated object is not very likely because observers tend to remember items as a whole when they share spatial-temporal properties (e.g., Baddeley et al., 2011; Delogu et al., 2012; Karlsen et al., 2010; Zeki, 2001). Moreover, even if the manipulation was not powerful enough and some semantic information was still available in the meaningless set, then the reported findings undermine the role of semantics in VWM, and these effects, if they exist, should be even larger for purely meaningless stimuli as was indeed observed in the next experiment that used a different, more powerful scrambling manipulation.

## Experiment 2

In the former experiment, we strived to use sets of stimuli that differ in their meaning but are similar in their visual properties. However, using sets of stimuli that differ *only* in their meaning is a great challenge. Thus, to generalize our results, in Experiment 2 we used a different manipulation to create the meaningless set. Specifically, we used the phase-scrambling technique of Stojanoski and Cusack (2014) that again preserves images' perceptual properties

while removing their meaning (see also Brady & Störmer, 2021). In addition, as Experiment 1 uncovered above chance level performance even for eight meaningless items, we used a larger memory set size of ten items.

## Method

### Participants

Forty-two individuals (14 males, mean age 25.38 years) participated in the experiment for a payment of 30 New Shekels (~8.5 $). This sample provided a power of 0.87 to find a within-subject interaction effect with a size of 0.2 or larger.

### Equipment, stimuli, procedure, and design

Except for the following changes, Experiment 2 was identical to Experiment 1. First and foremost, instead of the meaningless set that was previously used we used another meaning-distorting manipulation, specifically we used the diffeomorphic transformations (Stojanoski & Cusack, 2014) that should achieve the same goal – distorting images' meaning while keeping their visual properties constant. To do so we chose 1,603 images of daily objects (1.89 º x 1.89 º, taken from Brady et al., 2008) for our meaningful set, while Stojanoski and Cusack's (2014) scrambling method was applied for the same stimuli and acted as the meaningless set. We used a scrambling level of five, which seems to be sufficient to distort objects' meaning while at the same time preventing them from looking like color blobs (Fig. 1c). In addition, each trial consisted of ten rather than eight memory items, the repeated sets included 13 items, and each block included 60 instead of 64 trials.
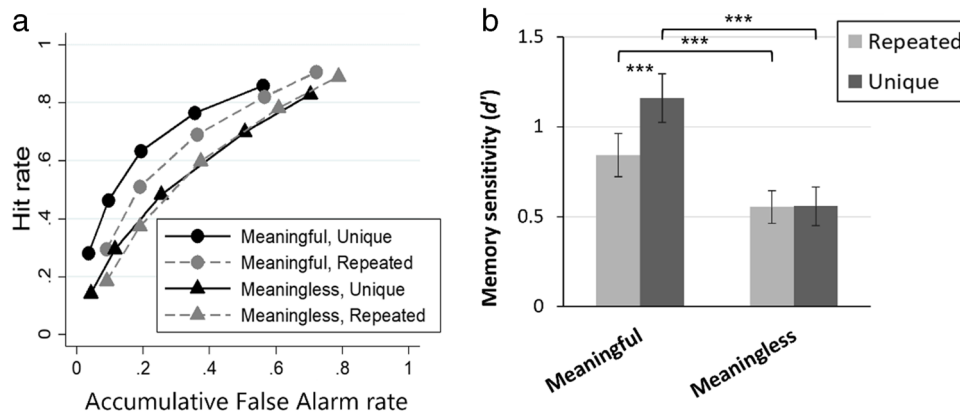
**Fig. 5** Receiver operating characteristic (ROC) curves of Experiment 2 as a function of the experimental conditions (**a**) and the means and 95% confidence intervals of the resulting *d'* (**b**). *** reflects *p* < .001

**Table 1** Means and 95% confidence intervals of the memory capacity k (set-size x (Hits – False alarms)) depending on the conditions in both experiments

|  | Exp. 1 | | Exp. 2 | |
|---|---|---|---|---|
|  | Meaningful | Meaningless | Meaningful | Meaningless |
| Unique | 4.49 (4.15–4.83) | 3.44 (3.11–3.77) | 4.39 (3.91–4.87) | 2.27 (1.8–2.74) |
| Repeated | 2.85 (2.54–3.16) | 2.52 (2.16–2.88) | 3.26 (2.79–3.73) | 2.24 (1.81–2.68) |

## Results

Four participants were excluded from the analysis due to performance that did not differ from chance level (50%), and two due to performance that was lower (2 SD) than the overall mean. Once more, we applied a repeated-measures ANOVA with Meaning and Susceptibility to PI as factors and the percent of correct responses independent of confidence levels as the dependent variable. As in Experiment 1, both main effects and their interaction were significant (Meaning, $F(1,35) = 67.36$, $p < .001$, $\eta_p^2 = .66$; Susceptibility to PI, $F(1,35) = 5.49$, $p = .025$, $\eta_p^2 = .14$; Interaction, $F(1,35) = 7.69$, $p < .01$, $\eta_p^2 = .18$; see Fig. 3a for the conditions' means and 95% CIs). However, as opposed to Experiment 1, while a significant PI-effect was found in the Meaningful condition, $t(35) = 4.02$ $p < .001$, $d = 0.67$, there was no PI-effect whatsoever in the Meaningless condition, $t(35) = 0.1$ $p = .92$. These findings were even stronger when confidence was integrated into the results (Fig. 3b and S3 in the OSM).

As in Experiment 1, the results were not driven by a difference in baseline performance, as also in terms of percentage, the PI-effect with the meaningful stimuli (M = 9.35%, 95% CI = 4.97–13.74) was larger relative to the (absence of) PI-effect of the meaningless stimuli (M = 1.72%, 95% CI = (-4.42) –7.86), t(35) = 2.35, p = .03, d = 0.39.

ROC curves (Fig. 5a) were also created and the area under each curve was converted to the sensitivity measure

*d'* (Salgado, 2018; Fig. 5b). A repeated-measures ANOVA with this measurement as the dependent variable revealed, once again, a main effect for the Susceptibility to PI, $F(1,35) = 5.92$, $p = .02$, $\eta_p^2 = .14$, the meaning of the stimuli, $F(1,35) = 92.17$, $p < .001$, $\eta_p^2 = .72$, and the interaction between these factors, $F(1,35) = 13.18$, $p < .001$, $\eta_p^2 = .29$.

Finally, to evaluate memory capacity in comparison to previous reports, *k* was calculated using the binary measurement of correct responses (independent of participants' confidence levels), by the standard formula of $k =$ set-size x (Hits – False alarms).[1] The data of both experiments are reported in Table 1.

Interestingly, in both experiments, only in the Unique condition with the meaningful stimuli did memory capacity exceed the utmost conventional capacity estimation of four items. However, because k was limited by set-size in these experiments, we ran a follow-up study that included only the Unique condition with a larger set size of 15 items. The results revealed a large *k* of 6.32 (95% CI = 5.69–6.95) for Meaningful stimuli, while the *k* of the Flipped-Meaningless stimuli (from Experiment 1) and of a homogeneous

---

[1] To calculate the *k* measure we used the binary correct variable, independent of participants confidence in their responses. However, because the ROCs are curvilinear, this analysis should be interpreted with caution (Robinson et al., 2020).

set composed entirely of faces was significantly lower and stayed in the range of 4 (see S4 in the OSM for a complete report of this experiment). Hence, in the absence of PI, both semantic and visual distinctiveness were necessary to significantly elevate memory capacity.

## Discussion

As in Experiment 1, participants performed better with the Meaningful than with the Meaningless stimuli, and their overall performance was better in the Unique compared to the Repeated condition. Most importantly, the interaction between these factors was once again robust. This finding substantiates the conclusion that meaning improves VWM performance mainly in the Unique condition. Furthermore, the ROC analyses in both experiments indicate that these effects of meaning, PI as well as their interaction, are driven largely by changes in memory sensitivity.

Notably, the small PI-effect that was present with the meaningless stimuli of Experiment 1 was completely absent in the current experiment. The difference between the Flipped and Scrambled sets can potentially explain these results – while in the Flipped set some objects' parts are still visible, the scrambling method transforms all the objects' parts, and hence it is possible that it distorted the meaning to a greater degree (see Brady & Störmer, 2021, for a similar view). Hence, the lack of PI-effect in the Meaningless-Scrambled condition raises the possibility that meaning is *necessary* for this effect to emerge. However, more research is needed to validate this strong claim.

## General discussion

To what extent does PI play a role in VWM performance? Contrary to previous studies that did not use a truly PI-free condition as a baseline (e.g., Hartshorne, 2008; Lin & Luck, 2012; Makovski & Jiang, 2008; Oberauer et al., 2017), studies that employed the RUP repeatedly found better performance in a PI-free (Unique) relative to a PI-prone (Repeated) condition. However, the magnitude of this effect is not consistent but dependent on factors like spatial (Makovski, 2016) and visual distinctiveness (Shoval et al., 2020). Because up to this point only meaningful stimuli were used in the RUP, the current study aimed to test whether and to what degree the use of meaningful stimuli is an essential factor in producing the PI-effect. Two experiments that also converged with the results of two Pilot experiments revealed superior memory for meaningful compared to meaningless stimuli, but this effect was stronger in the Unique condition. Consequently, the PI-effect (i.e., the difference between the Unique and Repeated conditions) for the meaningless stimuli was reduced (Experiment 1) or vanished (Experiment

2) compared to the Meaningful set. As this was the case with two different meaningless sets of stimuli that preserved most of the visual statistics of the meaningful set, we can conclude that the meaning of the stimuli, and not their visual characteristics, is the crucial factor in modulating the PI-effect.

Furthermore, in both experiments as well as in the follow-up experiment, only meaningful stimuli tested in the Unique condition demonstrated a large memory capacity that exceeded most previous estimations. These results suggest that in the absence of PI, memory for meaningful stimuli is qualitatively different than for meaningless stimuli and that this condition is likely to drive the large PI-effect typically found in the RUP. In line with this conclusion, the fact that some PI-effect was still detected for the Meaningless-Flipped set but not for the Meaningless-Scrambled set, which presumably distorts meaning to a greater degree (Brady & Störmer, 2021), further implies that meaning is a critical factor for finding PI in VWM, and when reduced, PI plays only a small role in VWM (e.g., Lin & Luck, 2012).

What is it about meaningful stimuli that drive these results? For one, meaningful stimuli can be labeled and it was shown that labeling during a VWM task (Souza & Skóra, 2017), as well as higher categorical distinctiveness between these labels (Souza et al., 2021), enhances performance. It is also often assumed that the meaning of the stimuli is encoded together with other features of the stimuli as a conceptual hook (Konkle et al., 2010; Koutstaal et al., 2003). This leads to stronger memory representations and increased inter-item distinctiveness, which can explain why the PI-effect is larger for meaningful than for meaningless items, in several, not mutually exclusive, ways. One possibility is that it is harder to clear these strong representations of the meaningful stimuli compared to the weaker representations of the meaningless stimuli, even after they are no longer needed. This possibility implies that a potential source of PI is overload due to "leftovers" in the VWM buffer that impairs the encoding and retaining processes. That is, if these leftovers occupy VWM they further limit its already restricted capacity, thus leaving less "room" for new representations, while also reducing the cognitive resources that are available for the memorization of both new and current information. This possibility is supported by evidence from the verbal memory domain showing that PI is not only affecting the testing phase but also affects the encoding of new information (e.g., Kliegl et al., 2015; Pastötter et al., 2011). In contrast, recent findings from our lab suggest that mainly the testing stage (retrieval), and not the encoding or retention, is affected by PI in the RUP (Shoval & Makovski, 2021).

Another possibility is that the stronger representations of meaningful stimuli are more likely to lead to confusion during testing because they linger in LTM for longer durations. This is consistent with the temporal discrimination theory
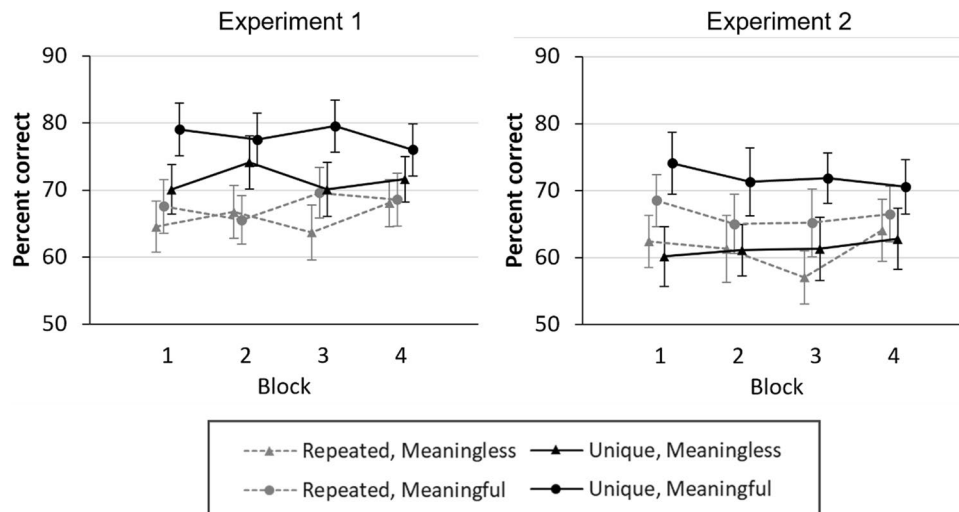
**Fig. 6** Means and 95% confidence intervals of the percent of correct responses, as a function of the stimulus type, the Susceptibility to PI, and the progress of the experiment

(Wixted & Rohrer, 1993), which postulates that participants are unable to restrict their focus to the relevant trial. This explanation, however, stands in contrast to the suggestion that working memory is an efficient memory system that mainly takes advantage of LTM when needed (Mızrak & Oberauer, 2021). For example, the Hebb repetition paradigm (Hebb, 1961) demonstrated that LTM contributes to working memory performance, as memory is better for lists of words that are repeating during the experiment compared to lists that are presented only once. Similarly, in the visual domain, Oberauer et al. (2017) found that after extensive learning of 120 color-object associations participants were better in a VWM task while PI was not increased. Conversely, this interpretation of our results entails that VWM is not as efficient as proposed and that previous information held in LTM can both assist and impair performance.

A somewhat different possibility is that the difference between the Repeated and Unique conditions is inflated when meaningful stimuli are tested regardless of PI. Specifically, the memory benefits of the meaningful stimuli are diminished in the Repeated condition, not because of increased confusion due to PI, but likely because the distinctiveness, both semantic and visual, which is afforded by the meaningful items in the Unique condition is greatly reduced when a repeated set of a few items is used (Shoval et al., 2020). This notion is consistent with the finding that a significant PI effect is observed mainly in conditions that afford semantic involvement (Endress & Potter, 2014; Makovski, 2016), whereas studies that used simpler, meaningless stimuli attributed a smaller role of PI in VWM (e.g., Hartshorne, 2008; Lin & Luck, 2012; Makovski & Jiang, 2008; Shipstead & Engle, 2013). Most importantly, this notion suggests that it is not that the lack of PI affords high capacity

in the Unique condition (Endress & Potter, 2014), rather the meaning of the stimuli affords both increased capacity and what only appears to be an increased influence of PI.

To further test the possibility that the difference between the Repeated and Unique conditions reflects more than PI built-up, we tested one of the key assumptions underlying the RUP. Specifically, we examined whether PI is accumulating throughout the task, thus influencing the performance of the Repeated condition as a whole. To that end, we divided each condition in Experiments 1 and 2 into four mini-blocks of 16 trials in Experiment 1 and 15 trials in Experiment 2. If there is indeed an accumulated PI build-up then performance in the Repeated condition should deteriorate as the experiment proceeds. Surprisingly, however, absolutely no indication for this pattern was found (Fig. 6). Even when performing eight independent repeated-measures ANOVA models, one for each condition of each experiment, and not controlling for multiple comparisons, none of the models was significant (all $Fs > 1.$, $ps > .13$). In other words, a time-dependent effect was not detected although the analysis gave it the highest chance to unfold. Thus, it seems that the PI-effect is not accumulated as the experiment advances, and is, with high probability, dependent on the immediate trials that preceded the current one. Consequently, this conclusion indirectly supports the notion that the larger difference between the Repeated and Unique conditions in the meaningful compared with the meaningless sets is not due to a larger PI build-up per se, but rather reflects differential effects of the meaning manipulation on the Repeated and Unique conditions.

As our focus was on visual memory, we manipulated meaning in two distinct ways that kept visual properties across the stimuli sets constant. This allowed us to infer that

meaning, and not low-level visual factors, drove the reported results. Still, one might argue that the meaningless stimuli shared a common visual feature (a hard edge in the middle of the stimuli in Experiment 1; curvature in Experiment 2) that reduced the visual distinctiveness of the set compared to the meaningful condition. Importantly, however, it is not clear whether, in fact, the Scrambled sets were less distinctive than the Meaningful set as most of the other key visual statistics (e.g., colors, orientation, brightness) were similar across the sets. Naturally, this is an empirical question and dedicated research is needed to determine whether there is a substantial difference in visual distinctiveness between the sets. However, we believe that it is not very likely that a possible reduction in visual distinctiveness was responsible for the better overall performance in the meaningful set compared to the meaningless set, and more importantly for the reduction of the PI in the latter. Indeed, that PI is affected by distinctiveness was exemplified in a previous study that showed a larger PI when the items were heterogeneous rather than homogenous (e.g., all items were houses; Shoval et al., 2020). Yet, the distinctiveness manipulation in that study affected both visual and semantic distinctiveness, and to a much greater degree, relative to the possible reduction in visual distinctiveness in the present study. Thus, the extent to which PI is modulated merely by visual distinctiveness still needs to be determined, and it is still possible that visual distinctiveness affected some portion of the results. Nonetheless, the current study showed that even when the visual distinctiveness is much more similar between the sets, reduced semantic distinctiveness (and possibly other meaning-related factors) suffices to largely reduce the PI effect.

All the same, meaning is a complex concept and even though we used two different distortion manipulations, it is inherently confounded with factors such as familiarity, categorical distinctiveness, or verbal encoding. Potentially, each of these factors might partially account for our findings. However, the use of articulatory suppression reduced the likelihood that verbal encoding underlies the present findings. There is also some evidence suggesting that familiarity cannot account for these findings. First, performance in the Repeated conditions, both with the meaningful and the meaningless sets, remained relatively stable as the experiments progressed (Fig. 6), suggesting that this factor did not affect our results to a significant degree. Second, familiarity effects are rather weak and mostly affect the consolidation of the stimuli primarily when the encoding duration is short (Blalock, 2015; Shoval & Makovski, 2021; Xie & Zhang, 2017, 2018). Nevertheless, it is still possible that these and other factors inherently related to the concept of meaning were in play.

Relatedly, one should also consider the possibility that semantic information afforded by the Meaningful set triggered the involvement of other memory constructs that were not involved or involved to a lesser degree with the Meaningless sets (e.g., conceptual memory (Hu & Jacobs, 2021; Potter, 1976); verbal long-term memory (Paivio, 1990); semantic memory (Shivde & Anderson, 2011)). However, this involvement can potentially explain mainly the improved performance with the meaningful compared to the meaningless stimuli, as it leaves open the discussion about VWM susceptibility to PI. The current dataset cannot verify or dismiss this possibility because if additional memory systems were involved, they were probably operating alongside VWM due to the visual nature of our stimuli. Thus, additional research is required to examine the perceptual and mnemonic differences between meaningful and meaningless stimuli and to characterize the complex interactions among the related memory constructs.

In conclusion, our results indicate that the meaning of the stimuli, and not their unique visual distinctiveness, leads to high VWM capacity and to what appears to be a stronger influence of PI. Importantly, this PI-effect is driven mainly by the qualitatively different performance with meaningful stimuli in a PI-free condition. Together with the finding that PI is not accumulated during a Repeated condition, our results suggest that VWM per se is affected by PI only to a minor degree.

In a broader perspective, the increasing use of daily stimuli in visual paradigms has major ecological justifications as our world is not composed of simple abstract stimuli. However, these stimuli add complexity to the study of fundamental cognitive processes, because they raise the likelihood that high-level cognitive processes would distort our conclusions. In the case of VWM, our results indicate that the use of meaningful stimuli leads to a large capacity as well as to what appears to be stronger PI in VWM tasks, whereas, for the meaningless stimuli, typical VWM findings were observed. These conclusions challenge the validity of using meaningful objects when trying to characterize "pure" fundamental processes, and further highlight the caution needed in the transfer from simple, artificial stimuli to complex, daily objects.

## Declarations

**Ethics approval and informed consent** All procedures performed in this study were approved by the Institutional Review Board (IRB) of the Open University of Israel (ethics approval number: 2930). Informed consent was obtained from all individual participants included in the study.

## 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, 1–17.* https://doi.org/10.1162/jocn_a_01693

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia, 49*(6), 1393–1400. https://doi.org/10.1016/j.neuropsychologia.2010.12.042

Besner, D., & Davelaar, E. (1982). Basic processes in reading: Two phonological codes. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 36*(4), 701–711. https://doi.org/10.1037/h0080665

Blalock, L. D. (2015). Stimulus familiarity improves consolidation of visual working memory representations. *Attention, Perception, and Psychophysics, 77*(4), 1143–1158. https://doi.org/10.3758/s13414-014-0823-z

Brady, T. F., & Störmer, V. S. (2021). The role of meaning in visual working memory: Real-world objects, but not simple features, benefit from deeper processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0001014

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*(38), 14325–14329. https://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*(27), 7459–7464. https://doi.org/10.1073/pnas.1520027113

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

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

Cowan, N. (2001). The magical number 4 in short term memory. A reconsideration of storage capacity. *Behavioral and Brain Sciences, 24*(4), 87–186. https://doi.org/10.1017/S0140525X01003922

Delogu, F., Nijboer, T. C. W., & Postma, A. (2012). Binding "when" and "where" impairs temporal, but not spatial recall in auditory and visual working memory. *Frontiers in Psychology, 3*(MAR), 1–6. https://doi.org/10.3389/fpsyg.2012.00062

Drew, T., Williams, L. H., Jones, C. M., & Luria, R. (2018). Neural processing of repeated search targets depends upon the stimuli: Real world stimuli engage semantic processing and recognition memory. *Frontiers in Human Neuroscience, 12*(460), 1–14. https://doi.org/10.3389/fnhum.2018.00460

Endress, A. (2022). Memory and proactive interference for spatially distributed items. Memory and Cognition, 1–35. https://doi.org/10.31234/osf.io/h8ge4

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General, 143*(2), 548–565. https://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. https://doi.org/10.1016/j.actpsy.2016.08.001

Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. *PLoS One, 3*(7). https://doi.org/10.1371/journal.pone.0002716

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *brain mechanisms and learning* (pp. 37–46).

Hu, R., & Jacobs, R. A. (2021). Semantic influence on visual working memory of object identity and location. *Cognition, 217*, 104891. https://doi.org/10.1016/j.cognition.2021.104891

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*(6), 685–701. https://doi.org/10.1016/0749-596X(91)90032-F

Karlsen, P. J., Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2010). Binding across space and time in visual working memory. *Memory and Cognition, 38*(3), 292–303. https://doi.org/10.3758/MC.38.3.292

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*(3), 153–161. https://doi.org/10.1016/S0022-5371(62)80023-1

Kliegl, O., Pastötter, B., & Bäuml, K. H. T. (2015). The contribution of encoding and retrieval processes to proactive interference. *Journal of Experimental Psychology: Learning Memory and Cognition, 41*(6), 1778–1789. https://doi.org/10.1037/xlm0000096

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General, 139*(3), 558–578. https://doi.org/10.1037/a0019165

Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: Testing the semantic categorization account. *Journal of Experimental Psychology: Learning Memory and Cognition, 29*(4), 499–510. https://doi.org/10.1037/0278-7393.29.4.499

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. https://doi.org/10.3389/fpsyg.2012.00042

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

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

Makovski, T. (2016). Does proactive interference play a significant role in visual working memory tasks? *Journal of Experimental Psychology: Learning Memory and Cognition, 42*(10), 1664–1672. https://doi.org/10.1037/xlm0000262

Makovski, T. (2018). Meaning in learning: Contextual cueing relies on objects' visual features and not on objects' meaning. *Memory and Cognition, 46*(1), 58–67. https://doi.org/10.3758/s13421-017-0745-9

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory and Cognition, 36*(1), 43–52. https://doi.org/10.3758/MC.36.1.43

Mızrak, E., & Oberauer, K. (2021). What is time good for in working memoryy? *Psychological Science, 32*(8), 1325–1337. https://doi.org/10.1177/0956797621996659

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), 1. https://doi.org/10.1037/xlm0000302

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences of the United States of America, 102*(24), 8776–8780. https://doi.org/10.1073/pnas.0500810102

Paivio, A. (1990). *Mental representations: A dual coding approach.* Oxford University Press.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44*(4), 369–378. https://doi.org/10.3758/BF03210419

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K. H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning Memory and Cognition, 37*(2), 287–297. https://doi.org/10.1037/a0021801

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individualverbal items. *Journal of Experimental Psychology, 58*(3), 193–198. https://doi.org/10.1037/h0049234

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*(2), 283–290. https://doi.org/10.3758/BF03203943

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509–522. https://doi.org/10.1037/0278-7393.2.5.509

Quinlan, P. T., & Cohen, D. J. (2016). The precategorical nature of visual short-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 42*(11), 1694–1712. https://doi.org/10.1037/xlm0000274

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

Sahar, T., Sidi, Y., & Makovski, T. (2020). A metacognitive perspective of visual working memory with rich complex objects. *Frontiers in Psychology, 11*, 1–14. https://doi.org/10.3389/fpsyg.2020.00179

Salgado, J. F. (2018). Transforming the area under the normal curve (AUC) into cohen's d, pearson's rpb, odds-ratio, and natural log odds-ratio: Two conversion tables. *European Journal of Psychology Applied to Legal Context, 10*(1), 35–47. https://doi.org/10.5093/ejpalc2018a5

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*(1), 277–289. https://doi.org/10.1037/a0028467

Shivde, G., & Anderson, M. C. (2011). On the existence of semantic working memory: Evidence for direct semantic maintenance. *Journal of Experimental Psychology: Learning Memory and Cognition, 37*(6), 1342–1370. https://doi.org/10.1037/a0024832

Shoval, R., & Makovski, T. (2021). The locus of proactive interference in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 47*(5), 704–715.

Shoval, R., Luria, R., & Makovski, T. (2020). Bridging the gap between visual temporary memory and working memory: The role of stimuli distinctiveness. *Journal of Experimental Psychology: Learning Memory and Cognition, 46*(7), 1258–1269. https://doi.org/10.1037/xlm0000778

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition, 166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Souza, A. S., Overkott, C., & Matyja, M. (2021). Categorical distinctiveness constrains the labeling benefit in visual working memory. *Journal of Memory and Language, 119*, 104242. https://doi.org/10.1016/j.jml.2021.104242

Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision, 14*(12), 1–16. https://doi.org/10.1167/14.12.6

Van Zandt, T. (2000). ROC curves and confidence judgemennt in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 582–600. https://doi.org/10.1037/0278-7393.26.3.582

Wixted, JT., & Rohrer, Doug. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1024–1039. https://doi.org/10.1037/0278-7393.19.5.1024

Xie, W., & Zhang, W. (2017). Familiarity speeds up visual short-term memory consolidation. *Journal of Experimental Psychology: Human Perception and Performance, 43*(6), 1207.

Xie, W., & Zhang, W. (2018). Familiarity speeds up visual short-term memory consolidation: Electrophysiological evidence from contralateral delay activities. *Journal of Cognitive Neuroscience, 30*(1), 1–13. https://doi.org/10.1162/jocn

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory : A review. *Psychological Bulletin, 133*(5), 800–832. https://doi.org/10.1037/0033-2909.133.5.800

Zeki, S. (2001). Localization and globalization in conscious vision. *Annual Review of Neuroscience, 24*(1), 57–86.

**Open practices statement**

The data of all experiments are publicly available via the Open Science Framework and can be accessed at https://osf.io/ju6m9/, and the stimuli used in the experiments are publicly available at https://osf.io/ns8p3/. The Experiments' programing-codes, as well as the scripts for the data analysis, will be sent via e-mail upon request. None of the experiments was preregistered.