# Massive visual long-term memory is largely dependent on meaning

Roy Shoval[1] · Nurit Gronau[1] · Tal Makovski[1]

## Abstract

Previous research demonstrated a massive capacity of visual long-term memory (VLTM) for meaningful images. However, the capacity and limits of a "pure" VLTM that is independent of conceptual information still need to be determined. In the encoding phase of three experiments, participants viewed hundreds of images depicting real-world objects, along with visually similar images that were stripped of their semantic meaning. VLTM was evaluated using a four-alternative-forced-choice test including old and new images and their counterpart mirror transformations. The results revealed superior memory for meaningful than for meaningless stimuli and importantly, there was no hint of a massive VLTM for the meaningless items. Furthermore, when examining memory recognition of visual properties *per-se* (i.e., original/mirror state), memory was overall poor, and practically negligible for the meaningless items. Taken together, our findings suggest that meaning is critical for massive VLTM and for the ability to store visual properties.

**Keywords** Long term memory · Meaning · Visual memory

Think about all the faces, scenes, and objects that you remember and can vividly visualize. This task seems impossible given the ample amount of information that our visual long-term memory (VLTM) holds. Indeed, there is strong empirical evidence that VLTM has a massive capacity, even for stimuli that are briefly presented in an artificial experimental setting. For instance, studies testing memory of up to 10,000 pictures revealed a remarkable performance, suggesting that there is practically no upper bound to VLTM capacity (Shepard, 1967; Standing, 1970, 1973). This is consistent with the wide-held view that one of the hallmark characteristics of LTM is its unbounded capacity, particularly compared with the restricted capacity of visual short-term memory (e.g., Brady et al., 2011).

More recent studies further showed that not only does VLTM have a massive capacity, but this type of information is stored with high fidelity (Bainbridge et al., 2019; Brady et al., 2008; Hollingworth, 2005; Konkle, Brady, Alvarez, & a,, & Oliva, A., 2010b). For example, after viewing 2,500 images of real-world objects, participants demonstrated high recognition performance of the old images even when the foils were drawn from the same semantic category as the memory targets, or when they depicted the same object in a different state (e.g., an open vs. closed dresser; Brady et al., 2008). Visually precise massive memory was also reported for natural scenes, as participants could recognize almost 3,000 such images regardless of the number of same-category scenes that were presented at encoding, and even though the test foils were drawn from the same category as the memory targets (Konkle, Brady, Alvarez, & a,, & Oliva, A., 2010b). These results indicate that VLTM stores not only basic-level categorical information but also a massive amount of visual details.

Importantly, all the aforementioned studies used real-world, meaningful images of objects and scenes (i.e., everyday stimuli that are easily identified/labeled). This raises the question regarding the extent to which massive VLTM relies on semantic information—can we remember visual details of stimuli that bear no semantic meaning? Is memory for the surface features of an item (e.g., color, orientation, or state) dependent on the encoding of its meaning? Semantic

✉ Roy Shoval
  rshoval@gmail.com

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

meaning is long known to support verbal (e.g., Besner & Davelaar, 1982; Hulme et al., 1991) as well as visual memory. For instance, memory for ambiguous shapes was found to be better when these were associated with labels that reduced their ambiguity (Bower et al., 1975; Koutstaal et al., 2003). In addition, a conceptually meaningful arrangement of several objects (as opposed to a meaningless object arrangement) allowed for improved memory of objects' visual details, even when viewed for a mere glimpse (Gronau & Shachar, 2015). Other studies have further shown that images' memorability was primarily predicted by semantic information, while basic visual details such as shapes or colors did not predict (Isola et al., 2014), or could hardly predict memorability rates (Hovhannisyan et al., 2021). Finally, preexisting knowledge and categorical information largely improve performance also in visual short-term memory tasks (Asp et al., 2021; Brady et al., 2016; Brady et al., 2019; Conci et al., 2021; Olsson & Poom, 2005; Sahar et al., 2020; Shoval & Makovski, 2022; Wiseman & Neisser, 1974).

Meaning may enhance visual memory performance in several ways. One possibility is that semantic knowledge affects how images are encoded and preserved in memory. That is, meaning may act as a "conceptual hook" that supports the long-term representation of the visual details (Brady et al., 2011; Konkle, Brady, Alvarez, & Oliva, 2010a). A related notion is that preexisting conceptual knowledge activates additional brain regions (relative to meaningless stimuli) that support the formation of visually detailed memories (Asp et al., 2021; Brady et al., 2008; Brady & Alvarez, 2015). Similarly, the stimulus' meaning can support performance by incorporating additional memory systems such as verbal LTM (e.g., Paivio, 1990), semantic working memory (Shivde & Anderson, 2011), or conceptual memory (Endress & Potter, 2014; Potter, 1976).

The important role that meaning plays in VLTM does not entail that visual information is strictly negligible, rather, both the images' "gist" and their visual details could support massive VLTM (Cunningham et al., 2015). A recent study provided evidence for short-term as well as long-term memory of "purely" meaningless visual stimuli—scrambled versions of natural scenes—suggesting that visual memorability can be driven by perceptual features per se (Lin et al., 2021). Note, however, that VLTM in this study was tested with a rather small set of 48 images. In addition, memory-recognition performance was tested only at a coarse level (i.e., detection of an old item), and not at a more fine-tuned perceptual level (e.g., recognition of specific visual details of the items). Hence, the degree to which participants can remember a *massive* amount of arbitrary *visual details*, in the absence of supportive semantic information, remains to be determined.

Here, we aimed to test whether a massive VLTM exists for meaningless visual stimuli, and if so, whether it is sensitive to relatively subtle perceptual transformations—the stimulus orientation. To this end, participants encoded hundreds of real-world objects or their scrambled versions, and their memory was tested for the items' exact appearance. The scrambling manipulation maintained most of the low-level visual statistics of the objects (color, size, brightness, etc.), and allowed similar levels of inter-item distinctiveness within the meaningful and scrambled sets (Shoval et al., 2020; Viswanathan et al., 2010), while dramatically reducing the objects' meaning.

## Experiment 1

The first experiment examined the extent to which meaning affects memory performance, at both the "coarse" and the "fine" levels. To manipulate meaning, two sets of stimuli were used—a *meaningful* set of real-world objects images and a *lightly scrambled* version of these images. We hypothesized that performance with the meaningful set would be better than with the lightly scrambled set, at both the coarse and the fine levels. This is because meaning might act as a "hook" that binds together the visual features of a stimulus, thus enabling more efficient and stable storage of the visual information.

Notably, this experiment was not aimed at testing "massive" memory; rather, it was our first approximation of VLTM performance when using the lightly scrambled images. It was further used to compare performance among participants who were explicitly notified of the upcoming LTM test with those who were not. We hypothesized that explicit instructions may affect the depth level at which visual information is encoded and stored in VLTM, mainly enhancing fine-detailed memory (e.g., Draschkow et al., 2019).

## Method

### Participants

Participants in all experiments were 19–34 years old, with normal or corrected-to-normal vision, normal color perception, and without any attentional, psychiatric, or neurological disorders. Each participant completed only one experiment. All three experiments were approved by the Ethics Committee of the Open University of Israel.

Forty-five individuals (18 males, mean age = 26.29) participated in Experiment 1 for a payment of 30 New Shekels (~$9.5). This sample provides a power of 0.91 to find a within-between interaction effect ($\eta_p^2$) with a size of 0.25 or larger in a mixed analysis of variance (ANOVA) model.

## Apparatus and stimuli

Participants sat about 67 cm from a 23-inch LCD screen (resolution 1,920 × 1,080) and were tested individually. The experiment was programmed with MATLAB R2018a (www.mathworks.com) and PsychToolbox (Version 3.0.14; Brainard, 1997). The meaningful objects set included 480 stimuli (3.17º × 3.17º; Fig. 1a), selected from a larger pool of images (adopted from Brady et al., 2008; https://brady lab.ucsd.edu/stimuli.html), whereas the lightly scrambled set included distorted versions of the same images, transformed as follows: Each image was split to half and recombined by inverting one of its halves (Fig. 1b; see also Brady & Störmer, 2021; Makovski, 2018; Sahar et al., 2020). This scrambling technique slows down the labeling and reduces the subjectively perceived meaningfulness of the stimuli (Makovski, 2018).

These stimuli were screened by four independent observers (three research assistants and the first author), such that the final set of 480 image pairs (meaningful images *and* their corresponding scrambled versions) excluded highly symmetrical stimuli (i.e., could not produce a distinguishable mirrored version), included written words, and/or included significant spaces between object parts in a way that compromised their "objecthood." In addition, 100 images of real-world objects were gathered from Google Images to serve as the filler task stimuli (see Procedure)—half remained intact and half were lightly scrambled.

Both the test and the filler task stimuli were randomly assigned to each participant, with no overlap between the meaningful and the lightly scrambled sets. That is, each participant viewed either the meaningful or the lightly scrambled version of an image, but not both. All stimuli used across experiments are publicly available online (https://osf.io/z93qk/).

## Procedure

The experiment started with an encoding phase. In each trial, an individual image was presented at the screen center for 500 ms, followed by a 1,500-ms blank interstimulus interval (ISI; Fig. 2a). Two-hundred forty images (120 meaningful, 120 lightly scrambled) were randomly presented during encoding and were later included in the memory-recognition test phase. To ensure that participants attended to the stimuli, we incorporated a filler task that included 32 additional images (16 meaningful, 16 lightly scrambled) that were presented twice during the encoding sequence (Brady et al., 2008). The repetitions occurred after intervals of 1, 2, 4, 8, 16, 32, 64, or 128 images (four stimuli in each interval), and the participant's task was to press the space key as fast as possible upon the detection of a repeated image. When a press was made, performance feedback appeared—a green plus sign for a correct response (hit) and a red minus sign for an incorrect response (false alarm)—and remained on the screen until the end of the ISI. Overall, 304 images were presented during the encoding phase, which lasted approximately 10 minutes.

Subsequently, participants received a brief explanation and began the memory-recognition test. A four-alternative forced-choice (4AFC) paradigm was used, in which each trial comprised an *old* image (shown at the encoding phase),



**Fig. 1** Examples of the meaningful (**a**) and the lightly scrambled (**b**) stimulus sets that were used in Experiments 1–2, and the fully scrambled stimulus set (**c**), used in Experiment 3
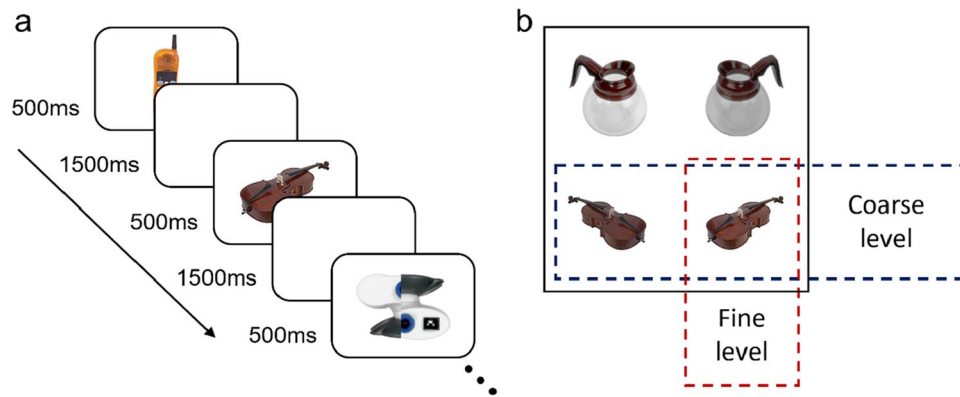
**Fig. 2** Illustration of the encoding phase sequence (**a**) and a 4AFC trial from the test phase (**b**). Note that recognizing either one of the violins was considered a coarse correct response. To the extent that the category was correctly identified, selecting the correct orientation was considered a fine correct response. Image examples are presented for illustrative purposes and are not scaled to their real size

a *new* image (never shown before), and the mirror transformations of these stimuli (see Fig. 2b). The four images always belonged to the same stimulus set (meaningful or lightly scrambled). Participants were instructed to choose the old image as it appeared in its *original orientation*. For each 4AFC trial, the locations of old and new images (upper/lower quadrant), and correct/incorrect orientations (left/right quadrant) were randomly determined. Stimuli were positioned 4.11 degrees diagonally from the screen center. After a response, a 500-ms blank interval preceded the next test trial.

### Design

A mixed design was used, with the type of stimulus (meaningful/meaningless) as a within-subject factor and participants' knowledge of the upcoming test phase (informed/uninformed about the test phase at encoding) as a between-subject factor. In both the informed and uninformed conditions, participants were told that their goal was "to remember the presented images and their details as accurately as possible." However, participants in the ininformed group (*n* = 26) were unaware of the upcoming test phase and were notified only about the repetition detection task, while participants in the Informed group (*n* = 19) were notified in advance about the memory test phase that would follow the repetition detection task. Note, however, that no specific instructions were given about the orientation of the images, in either group.

### Data analysis

We first tested participants' performance during the encoding phase and compared repetition detection between the meaningful and scrambled stimuli. Responses from the test phase were then analyzed at two levels—coarse and fine.

For the coarse level, we tested whether participants remembered the identity of the old stimuli regardless of their orientations. Thus, responses were considered correct if the participant chose either the correct or incorrect orientation of the old stimulus. Then, to test performance at a fine, perceptual level, we used a conditional analysis, in which only responses that were correct at the coarse level were used. In other words, we examined whether participants remembered the fine details of a stimulus (i.e., its orientation), given that its identity was correctly remembered (for a similar analysis see, e.g., Swallow & Jiang, 2010).

### Results

Participants exclusion criteria were predetermined and identical across all experiments. In Experiment 1, three participants were excluded from the analysis because their hit rate in the repetition detection task was lower (2 *SD*) than the overall mean. One additional participant was excluded because her performance in the test phase at the coarse level, was lower (2 *SD*) than the overall mean. Thus, data from 41 participants (informed group, *n* = 17; uninformed group, *n* = 24) were included in the following analyses.

We first examined the detection rate and reaction time (RT) to the repeated items during encoding (i.e., the filler task). A mixed ANOVA with the stimulus type (meaningful/meaningless) as a within-subject factor, and the encoding type (informed/uninformed) as a between-subject factor, revealed higher and faster hit rates for the meaningful compared with the lightly scrambled stimuli—accuracy, $F(1, 39) = 15.06$, $p < .001$, $\eta_p^2 = .28$; RT, $F(1, 39) = 9.57$, $p = .004$, $\eta_p^2 = .2$ (Table 1). As expected, the hit rate for repeated pairs with short intervals between images was higher than that of long-interval pairs (see Supplemental Materials for the complete analysis). Furthermore, the false alarm (FA) rate (i.e., erroneously detecting a nonrepeated item as a repeated

one) was higher for the lightly scrambled compared with the meaningful stimuli, $F(1, 39) = 24.35$, $p < .001$, $\eta_p^2 = .38$ (Table 1). There was no effect for the encoding type condition, nor an interaction between the conditions in any of these analyses (all $F$s < 3.12, all $P$s > .08).

Next, we examined the accuracy performance in the test phase at its coarse level (old vs. new stimuli, regardless of their orientation). A mixed ANOVA model revealed, once again, a robust main effect of Stimulus type, $F(1, 39) = 86.55$, $p < .001$, $\eta_p^2 = .69$, stemming from improved performance in the meaningful compared with the lightly scrambled condition (Fig. 3a). The encoding type factor and the interaction between the two factors were nonsignificant—encoding type, $F(1, 39) = 0.95$, $p = .34$; interaction, $F(1, 39) = 0.4$, $p = .53$.

Then, we used only responses that were correct at the coarse level and tested the performance at the fine level (i.e., examining whether the original orientation was chosen). Once again, a mixed ANOVA model revealed only a strong effect for the Stimulus type, $F(1, 39) = 11.29$, $p = .002$, $\eta_p^2 = .22$, indicating better performance with the meaningful than the lightly scrambled stimuli (Fig. 3b). All other effects were nonsignificant—informed/uninformed, $F(1, 39) = 0.51$, $p = .48$; interaction: $F(1, 39) = 0.5$, $p = .48$.

Finally, two one-sample $t$ tests compared participants' fine-level performance to chance level (50%) with both stimuli types, across the informed/uninformed condition. Due to clear directional hypotheses, the $t$ test here and in all subsequent analyses were one-sided. These analyses confirmed that the mean correct response at the fine level was significantly higher than chance for both the meaningful, $t(40) = 6.45$, $p < .001$, $d = 1.01$, $BF_{10} > 2.7*10^5$, and the lightly scrambled sets, $t(40) = 2.27$, $p = .01$, $d = .35$, $BF_{10} = 3.29$.

## Discussion

Experiment 1 revealed a strong main effect of Stimulus type across all analyses: stimulus meaning affected repetition detection performance during the encoding phase, and most importantly, it modified long-term memory at both the coarse and the fine levels. At the fine level, although memory for perceptual details (i.e., stimulus orientation) was better than chance, it was overall quite poor (<60%). Still, it was even worse for the Lightly scrambled stimuli than the Meaningful stimuli.

Interestingly, explicit instructions to memorize the stimuli for a future memory test did not affect any of the results. This result is consistent with several previous findings (e.g., Hyde & Jenkins, 1973; Makovski et al., 2020; Oberauer & Greve, 2021) and hence this factor was not further tested and only the explicit instructions condition was used in the next experiments.

## Experiment 2

Experiment 2 was designed to allow a more reliable examination of a potential *massive* VLTM. Hence, the experiment was identical to Experiment 1, aside from the fact that all participants were notified about the upcoming memory test, and the number of encoded stimuli was more than doubled. We expected that the overall performance level would decrease, compared with Experiment 1, due to the enlargement of the stimulus set and the increase in cognitive load during encoding. Still, we hypothesized that the main pattern of results would remain, that is, performance with the meaningful set would be higher than with the Scrambled set, at both levels of analysis.

### Method

#### Participants

Based on the effects found in Experiment 1, we calculated that a minimum of 20 participants is needed for finding a strong effect size ($d = 0.8$) with a power of 0.95 in a
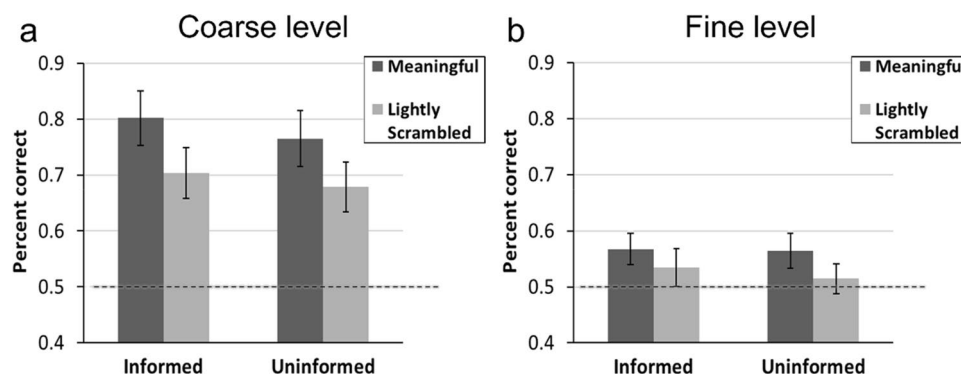


**Fig. 3** Mean recognition performance and 95% confidence intervals at the coarse level (regardless of orientation) (**a**) and the fine level (given a coarse-level correct response) (**b**), in Experiment 1. The dotted line represents chance level

one-tailed *t*-test. Twenty-five individuals (10 males, Mean age 26.92) participated in Experiment 2 for a payment of 90 New Shekels (~$28).

### Apparatus and stimuli

The Apparatus and stimuli were identical to those of Experiment 1, but now each set (Meaningful, Lightly scrambled) included 1,216 images.

### Procedure and design

The experiment was identical to Experiment 1, aside from the following changes: First and foremost, during the encoding phase, 768 images were presented. Six hundred and eight images (304 Meaningful, 304 Lightly scrambled) appeared once and were later included in the test phase. The set of the repetition-detection filler task included 80 image pairs that appeared in the following intervals: 14 image pairs with intervals of 1 & 2, 12 with an interval of 4, 10 with an interval of 8, 8 with intervals of 16 & 32, 6 with an interval of 64, 4 with an interval of 128, and 2 with intervals of 256 & 512 (Brady et al., 2008).

Second, due to the length of the experiment, two one-minute breaks were given during the encoding phase (before trials 256 and 512), and three self-paced breaks were given during the test phase (before trials 152, 304, and 456). Thus, the encoding stage lasted ~28 minutes, and completing the entire experiment took approximately 75 minutes. Finally, all participants completed only the Informed condition, and thus a within-subject design with the stimulus type as a single factor was used.

## Results

One participant was excluded from the analysis due to poor detection of repeating items during the encoding phase. No participants were dropped due to poor performance at the coarse level in the test phase.

As in Experiment 1, we first assessed participants' performance in the filler task. Two paired-samples t-tests revealed greater and faster hit rates for the Meaningful compared with the Lightly scrambled stimuli (Accuracy—$t(23) = 6.42$, $p<.001$, $d = 1.31$, $BF_{10} = 12,410.63$; RT—$t(23) = 3.2$, $p = .002$, $d = 0.65$, $BF_{10} = 10.52$; Table 1). The FA rate was again higher for the Lightly scrambled compared with the Meaningful stimuli, $t(23) = 4.49$, $p<.001$, $d = 0.92$, $BF_{10} = 173.23$ (Table 1).

In the main VLTM task, a paired-samples t-test revealed a better coarse-level performance with the Meaningful compared with the Lightly scrambled stimuli, $t(23) = 7.61$, $p<.001$, $d = 1.55$, $BF_{10}>3.01*10^5$. However, for the fine level, the difference between the Meaningful and Lightly

**Table 1** Descriptive statistics (means and 95% confidence intervals) of the performance in the repetition-detection task across all experiments

| Exp. | Measurement | Hits | | False alarms | | Hits | False alarms |
| | | Meaningful | Lightly scrambled | Meaningful | Lightly scrambled | Fully scrambled | Fully scrambled |
|---|---|---|---|---|---|---|---|
| 1 | Accuracy (percent) | 81.86 (76.34–87.39) | 73.17 (67.12–79.22) | 2.98 (2.28–3.68) | 5.38 (4.1–6.66) | | |
| | RT (milliseconds) | 746.51 (712.46–780.53) | 778.07 (748.1–808.04) | | | | |
| 2 | Accuracy (percent) | 80 (73.12–86.88) | 71.04 (64.12–77.97) | 2.4 (1.65–3.15) | 3.86 (2.74–4.99) | | |
| | RT (milliseconds) | 761.43 (718.26–804.6) | 788.41 (743.11–833.71) | | | | |
| 3 | Accuracy | | | | | 42.04 (34.53–49.55) | 11.47 (8.67–14.27) |
| | RT (milliseconds) | | | | | 818.45 (764.8–872.1) | |
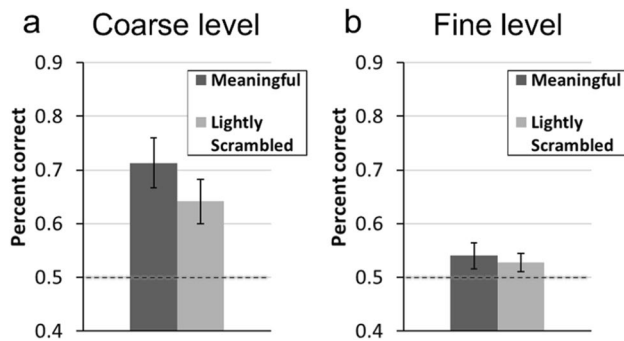
**Fig. 4** Mean recognition performance and 95% confidence intervals at the coarse level (**a**) and the fine level (**b**), in Experiment 2. The dotted line represents chance level

scrambled stimuli was insignificant, $t(23) = 1.21$, $p = .12$, $BF_{10} = 0.72$ (see Fig. 4). Although the overall performance was relatively poor, two one-sample t-tests confirmed that the mean correct response at the fine level was still higher than chance (50%) for both the Meaningful ($t(23) = 3.37$, $p = .001$, $d = .69$, $BF_{10} = 29.78$) and the Lightly scrambled stimuli ($t(23) = 3.44$, $p = .001$, $d = .7$, $BF_{10} = 34.55$).

## Discussion

As in Experiment 1, performance in the repetition detection task during encoding, and performance of long-term memory at the coarse level, were better with the Meaningful than with the Lightly scrambled stimuli. However, unlike Experiment 1, there was no evidence for a difference in memory at the fine level between the Meaningful and the Lightly scrambled stimuli. This might originate from a floor effect because, although memory for fine visual details was better than chance in both stimulus types, it was overall very poor (<55%).

More importantly, when asked to remember hundreds of meaningless images, long-term recognition accuracy, even at the coarse level, dropped to about 65%. This finding suggests that in the absence of meaning, VLTM is rather limited in capacity. This conclusion is further consistent with the poor performance at the fine level for both stimulus types, which indicates that memory for arbitrary, 'meaningless' visual features (orientation) is far from having a massive capacity.

## Experiment 3

Experiment 3's goal was to provide a more stringent test for the capacity of VLTM when meaning is further minimized. To that end, we used a different meaning-distorting manipulation (Stojanoski & Cusack, 2014) that scrambles all of an object's parts, and hence reduces the likelihood that intact object parts can reveal its identity (Fig. 1c; Brady & Störmer, 2021; Shoval

& Makovski, 2022). Because this scrambling technique was applied to the same Meaningful set of objects that was tested in the previous experiments, the Meaningful set was not included in this experiment. Importantly, we expected that this strong distortion manipulation would impose even stronger limitations on visual memory capacity, yielding an overall lower performance than the previous experiments, at both the coarse and fine levels.

## Method

### Participants

Thirty-three individuals (10 males, Mean age 25.33) participated in the experiment for a payment of 30 New Shekels (~$9.5), providing a power of 0.88 to find an effect size of $d = 0.5$ in a one-tailed, one-sample t-test.

### Apparatus and stimuli

The Apparatus was identical to the previous experiments. Only one new stimulus set was tested—a Fully-Scrambled set. To create this set, 662 meaningful images that were used in the previous experiments were subjected to a Diffeomorphic Scrambling technique (Stojanoski & Cusack, 2014). A scrambling level of five was used as it seemed sufficient to distort the objects' meaning and transform all of the objects' parts while preventing them from appearing as random color blobs (see Fig. 1c). Six hundred and twelve of these images were randomly selected to serve as the stimuli for all participants, and the remaining 50 images were used in the filler repetition-detection task.

### Procedure and design

Apart from the inclusion of a single stimulus set, and the employment of an Informed condition without an Uninformed one, the experiment was identical to Experiment 1. To clarify, as in Experiment 1, 304 images, randomly selected for each participant, were presented during encoding. This included the 240 test images (all Fully-Scrambled) and the 32 repeating images (with the same intervals as in Experiment 1) for the filler task. All 240 test images were later tested in the 4AFC test phase.

### Results

One participant was excluded from the analysis due to poor repetition-detection performance. No participants were dropped due to poor coarse level performance in the test phase.

The repetition-detection rate was low, likely because the stronger scrambling manipulation made the stimuli more difficult to remember (see Table 1). Accordingly, detection

accuracy was lower than the repetition-detection of the Lightly scrambled set in Experiment 1, $t(70) = 6.63$, $p<.001$, $d = 1.58$, $BF_{10} = 1.52*10^6$. Similarly, the FA rate with the Fully-Scrambled stimuli was high, and higher than the FA rate of the Lightly scrambled set in Experiment 1, $t(70) = 4.36$, $p<.001$, $d = 1.04$, $BF_{10} = 462.03$. Finally, the RT of hit responses with the Fully-Scrambled stimuli was only numerically higher than with the Lightly scrambled set in Experiment 1, $t(69) = 1.43$, $p = .16$, $BF_{10} = 0.59$.

Most importantly, even at the coarse level memory was now very low (M = 54.78%; $CI_{95\%}$ = 53.06-56.51), although higher than chance level, $t(30) = 5.66$, $p<.001$, $d = 1.02$, $BF_{10} = 10,680$. The fine level performance was also very poor (M = 51.5%; $CI_{95\%}$ = 50.27-52.73) but still slightly better than chance, $t(30) = 2.49$, $p = .01$, $d = 0.45$), $BF_{10} = 5.24$.

## Discussion

Experiment 3 revealed that when stimuli are largely stripped of their meaning, visual memory performance is severely impaired. Specifically, even at the coarse level performance was very low, indicating that only a handful of the 240 images was remembered. Memory performance was even poorer at the fine level, demonstrating almost no memory for visual details. Overall, these results substantiate the conclusion that meaning is imperative for VLTM.

## General discussion

This study examined the possible existence of a massive-capacity VLTM that is independent of conceptual information. Participants viewed hundreds of images from stimulus sets that differed in their meaning but were comparable in terms of low-level visual statistics. Then, memory was tested at a coarse level that contrasted an old image with a novel one, and at a fine level that contrasted two orientations of the same image.

Consistently with past findings (Asp et al., 2021; Konkle, Brady, Alvarez, & Oliva, 2010a), we found that meaning enhanced VLTM. Performance at the coarse level was better with the meaningful than with the Lightly scrambled stimuli (Exp. 1 & 2), which in turn, was better than the Fully-Scrambled stimuli (Exp. 3). Experiment 1 further provided some evidence that meaning contributed to LTM at the fine level, but this effect was smaller compared with the coarse level and was likely restricted by the overall poor performance of fine details memory. Most importantly, across the three experiments, there was no indication of a truly massive, 'pure' visual LTM. By large, participants were unable to remember the fine visual details (i.e., the orientation) of most stimuli, regardless of stimulus type. Even when examining the coarse level only, the need to encode a massive

amount of lightly scrambled stimuli yielded rather poor memory performance (Experiment 2). Further scrambling the objects and abstracting them from conceptual meaning nearly abolished recognition ability at both the coarse and the fine levels (Experiment 3).

As discussed in the Introduction, researchers have suggested several ways in which semantic meaning and preexisting knowledge might assist visual memory. The present results further indicate that meaning not only assists visual memory but is rather critical for remembering a massive amount of visual information. In visual working memory, for instance, it was shown that prior knowledge enables larger memory capacity (Brady et al., 2009; Feigenson & Halberda, 2008) and the same mechanism might facilitate encoding to Long-Term Memory as well. That is, the ability to encode and store visual information in the absence of meaning is quite restricted, perhaps because semantic/conceptual information is necessary for binding together independent visual features. This 'gluing' process could take place during either encoding or memory storage. Alternatively (yet not mutually exclusively), conceptual meaning could serve as an efficient retrieval cue that facilitates accurate recognition. Clearly, further research is required to scrutinize the role of semantics in visual memory.

It is noteworthy that in both Experiments 1-2, VLTM for *meaningful* items at the *coarse* level was lower compared with past reports (e.g., >90%; Brady et al., 2008). This discrepancy may be explained by the use of a short encoding duration (500ms vs. 3,000ms), which likely made it harder for participants to fully utilize the meaning of the stimuli (e.g., identify or label them). Interestingly, this may potentially imply that the effect of meaning on VLTM might be even larger than we report because it was recently found that meaningful items are more likely to benefit from longer and deeper encoding than scrambled objects or simple features (Brady & Störmer, 2021; Brady et al., 2016; but see Li et al., 2020).

As mentioned above, our findings further revealed that memory at the *fine* visual level was extremely poor, not only for the meaningless stimuli but also (albeit somewhat better) for the meaningful stimuli. This finding stands in contrast to past reports arguing that VLTM for real-world objects stores massive information with high fidelity (e.g., Brady et al., 2008). This inconsistency might be explained by the fact that a specific 'state' of a real-world object (e.g., an open vs. closed dresser, or a full vs. empty mug) is likely represented both at the visual and the conceptual levels. In contrast, an arbitrary visual feature or detail, such as an object's orientation, is encoded and represented only (or mostly) at the visual-perceptual level, resulting in rather poor memory performance levels. However, dedicated research is needed to examine this possibility.

Taken together, our data challenge the existence of a purely visual, massive LTM that can store hundreds of meaningless images or arbitrary visual details. Instead, we argue that for

storing ample visual information, memory must be 'hooked' onto semantic information. These conclusions further blur the distinction between visual short-term memory, which was traditionally tested with meaningless stimuli, and VLTM that is usually tested with meaningful stimuli. When meaningful objects were used in a short-term memory task, the capacity limit seemed almost unbounded (e.g., Endress & Potter, 2014). Here, we show a complementary finding - when VLTM is tested with meaningless items, capacity is severely limited. Thus, both memory systems are greatly affected by meaning, and both are highly limited when meaning is absent.

**Data availability** The raw data files and stimuli sets that were used across experiments are publicly available online (https://osf.io/9k532/). The codes for the experiments as well as the analyses codes will be sent by email upon request.

## Declarations

**Conflicts of interest** The authors have no conflicts of interest to disclose. Parts of this research were presented at the 62nd Annual Meeting of the Psychonomic Society (2021), at the 29th Object Perception, Attention, and Memory conference (OPAM; 2021), and at the 9th Israeli Conference on Cognition Research (ISCOP; 2022).

**Ethics approval** Approval was obtained from the ethics committee of the Open University of Israel. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

**Consent to participate** 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, 33*(5), 902–918.

Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2019). Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nature Communications, 10*(1), 1–13.

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

Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory & Cognition, 3*(2), 216–220.

Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision, 15*(15), 1–24.

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, 48*(7), 942. 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.

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

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), 1–34.

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.

Brady, T. F., Alvarez, G. A., & Störmer, V. S. (2019). The role of meaning in visual memory: Face-selective brain activity predicts memory for ambiguous face stimuli. *Journal of Neuroscience, 39*(6), 1100–1108.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

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

Cunningham, C. A., Yassa, M. A., & Egeth, H. E. (2015). Massive memory revisited: Limitations on storage capacity for object details in visual long-term memory. *Learning & Memory, 22*(11), 563–566.

Draschkow, D., Reinecke, S., Cunningham, C. A., & Võ, M. L. H. (2019). The lower bounds of massive memory: Investigating memory for object details after incidental encoding. *Quarterly Journal of Experimental Psychology, 72*(5), 1176–1182.

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General, 143*(2), 548–565.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants ' memory capacity. *Proceedings of the National Academy of Sciences, 105*(29), 9926–9930.

Gronau, N., & Shachar, M. (2015). Contextual consistency facilitates long-term memory of perceptual detail in barely seen images. *Journal of Experimental Psychology: Human Perception and Performance, 41*(4), 1095–1111.

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

Hovhannisyan, M., Clarke, A., Geib, B. R., Cicchinelli, R., Monge, Z., Worth, T., Szymanski, A., Cabeza, R., & Davis, S. W. (2021). The visual and semantic features that predict object memory: Concept property norms for 1,000 object images. *Memory & Cognition, 49*(4), 712–731.

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.

Hyde, T. S., & Jenkins, J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior, 12*(5), 471–480.

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(7), 1469–1482.

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

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

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.

Li, X., Xiong, Z., Theeuwes, J., & Wang, B. (2020). Visual memory benefits from prolonged encoding time regardless of stimulus type. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(10), 1998.

Lin, Q., Yousif, S. R., Chun, M. M., & Scholl, B. J. (2021). Visual memorability in the absence of semantic content. *Cognition, 212*, 104714.

Makovski, T. (2018). Meaning in learning: Contextual cueing relies on objects' visual features and not on objects' meaning. *Memory & Cognition, 46*(1), 58–67.

Makovski, T., Michael, S., & Chajut, E. (2020). How does a threatening stimulus affect the memory of the display? *Quarterly Journal of Experimental Psychology, 73*(5), 676–687.

Oberauer, K., & Greve, W. (2021). Intentional remembering and intentional forgetting in working and long-term memory. *Journal of Experimental Psychology : General.* Advance online publication. https://doi.org/10.1037/xge0001106

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.

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

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

Sahar, T., Sidi, Y., & Makovski, T. (2020). A metacognitive perspective of visual working memory with rich complex objects. *Frontiers in Psychology, 11*, 1–14.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 156–163.

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.

Shoval, R., & Makovski, T. (2022). Meaningful stimuli inflate the role of proactive interference in visual working memory. *Memory & Cognition, 50*, 1157–1168.

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.

Standing, L. (1970). Perception and memory for pictures: Single trial learning of 2500 visual stimuli. *Psychonomic Science, 19*(2), 73–74.

Standing, L. (1973). Learning 10 000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207–222.

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.

Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition, 115*(1), 118–132.

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*(1), 59–65.

Wiseman, S., & Neisser, U. (1974). Perceptual organization as a determinant of visual recognition memory. *The American Journal of Psychology, 87*(4), 675–681.