

# Memory sensitivity improves with larger image size during naturalistic encoding

Kreichman Olga<sup>1,2,\*</sup>, Masarwa Shaimaa<sup>1,2</sup>, Brook Limor<sup>1,2</sup>, Gilaie-Dotan Sharon<sup>1,2,\*</sup>

<sup>1</sup> *School of Optometry and Vision Science, Bar Ilan University, Ramat Gan, Israel*

<sup>2</sup> *The Gonda Multidisciplinary Brain Research Center, Bar Ilan University, Ramat Gan, Israel*

\* Corresponding authors:

Sharon Gilaie-Dotan, School of Optometry and Vision Science, Faculty of Life Science, Bar Ilan University, Ramat Gan, 5290002 Israel. E-mail: [shagido@gmail.com](mailto:shagido@gmail.com), [sharon.gilaie-dotan@biu.ac.il](mailto:sharon.gilaie-dotan@biu.ac.il)

Olga Kreichman, School of Optometry and Vision Science, Faculty of Life Science, Bar Ilan University, Ramat Gan, 5290002 Israel. E-mail: [oplaksin88@gmail.com](mailto:oplaksin88@gmail.com)

**Abstract**

We have recently found that people remember bigger images more than smaller images during naturalistic-like encoding. Since these findings were established with accuracy levels which may be confounded by response biases, and since memory performance was apparently low, here we addressed these issues. By reanalyzing our old lab dataset ( $n=101$ ) and in a new set of online experiments ( $n=199$ ), we estimated accuracy,  $d'$  prime memory-sensitivity that accounts for potential response biases (criterion), and RT-based ROC measures. Our findings, which were consistent across experiments, showed that overall memory performance as estimated by  $d'$  prime and ROC-based analyses was modest but significantly above chance, and these indicate that certain level of image information was encoded and learnt. Furthermore, larger images (in the  $3^\circ$ - $12^\circ$  range) were better remembered as evident in  $d'$  prime and accuracy. In the online experiments, memory for  $1.5^\circ$  images did not differ from that of  $3^\circ$  images and memory for images likely bigger than the screen size ( $24^\circ$ ) dropped, likely reflecting limitations of display size. Memorability also increased with image size. Our findings provide converging measures supporting the idea that image size plays an important role in image memory during naturalistic encoding for images in the size range of  $3^\circ$ - $12^\circ$ . Our study also emphasizes the importance of testing an effect across multiple measures and methods.

**Keywords:** image memory, image size, naturalistic encoding, signal detection theory, RT-based ROC, image memorability

## Introduction

While much is known about the outstanding capacity of visual image memory (e.g. Standing, 1973; Brady et al., 2008), most image memory studies employ an image-encoding task when participants can be either informed or not informed that they need to memorize the images during that task (e.g. Brady et al., 2008; Bilton, 2011; Bainbridge, 2017; Bönstrup et al., 2020). Therefore, much less is known about what influences image memory during more naturalistic-like visual behavior as for example when participants are freely viewing the images without being instructed to perform any image encoding task, and when participants are not informed of any memory aspect of the study (note this may differ from incidental memory paradigms that typically instruct participants to perform some encoding task). Physical image properties which are measurable properties that physically describe an image (as image size, contrast, luminance (overall brightness), and color (e.g. Wandell, 1995; Marr, 2010) are not always considered to play a dominant role in visual memory (e.g. image size, contrast and luminance are often not accounted for in image memory studies, but this is not always the case (e.g. Fiser and Biederman, 2001; Wolfe and Kuzmova, 2011) while higher-level properties as visual categories and semantic contents are assumed to play a more consistent role in memory across studies (Sato and Yoshikawa, 2013; Isola *et al.*, 2014; Khosla *et al.*, 2015). However, the physical property of image size is known to modulate the extent of visual system activity associated with processing an image such that bigger images activate more visual system processing resources starting from the retina and in retinotopic brain areas (Tootell *et al.*, 1997; Grill-Spector and Malach, 2004; Schneider, Richter and Kastner, 2004). Therefore, we have recently assumed that during more naturalistic-like vision, free of task-related modulations (Hillyard, Vogel and Luck, 1998; Huk and Heeger, 2000; Saalmann and Kastner, 2009), the enhanced visual system processing resources devoted to larger images in early processing stages will be translated to better quality signals transmitted downstream (to higher-order visual- and memory-related areas) and will result in better memory than that of smaller images. While different visual

categories are supported by different higher-order visual areas that are each sensitive to different aspects of the visual input and some are modulated by visual eccentricity (e.g. Levy *et al.*, 2001; Grill-Spector and Malach, 2004; Gilaie-Dotan, Nir and Malach, 2008; Kreichman and Gilaie-Dotan, 2024), we assumed that the effect of image size on image memory would affect most visual categories even if to different extents. We have recently developed a more naturalistic encoding paradigm (resembling incidental encoding but without any instructed task) where participants freely view the images, have no encoding task, and at this viewing phase are not informed about the memory aspect of the study (Masarwa, Kreichman and Gilaie-Dotan, 2022). This paradigm, which we will hereafter refer to as “naturalistic encoding”, allowed us to test our hypothesis that without task-related modulations larger images are better remembered, and we found strong evidence for that across multiple experiments (overall  $n = 182$  participants in 7 different experiments (Masarwa, Kreichman and Gilaie-Dotan, 2022)). In that study we relied on recognition memory accuracy for already-seen images (% correct) to estimate memory. However, it has been suggested that this measure in recognition memory studies may be potentially affected by individual response biases (Brady *et al.*, 2023) and this indicates that recognition memory accuracy for older images that we measured may actually reflect a change in decision criterion (e.g. more liberal criterion with growing image size) rather than a genuine change in recognition memory sensitivity. In addition, in our earlier study (Masarwa, Kreichman and Gilaie-Dotan, 2022), while significant within-individual effect of image size on recognition memory was consistently found across experiments, the overall memory levels were low (between ~35-40% accuracy for smallest images and ~55-65% for the largest images). These findings raise the question of whether during naturalistic encoding participants can successfully encode memory representations of images and this is especially relevant for smaller images.

To address these issues, here in a reanalysis of our earlier lab findings (Experiments 1-3 and 5,  $n=101$ ) and in a new set of 3 online experiments (overall  $n=199$ ), we (i) estimated accuracy to test replication of

our earlier findings, (ii) assessed  $d'$  to test whether memory-sensitivity is affected by image size (beyond any potential effects on response biases), and in the online experiments we also (iii) computed RT-based ROC measures (Weidemann and Kahana, 2016). The  $d'$  and ROC-based analyses allowed us to verify that overall memory performance was above chance or guessing performance. In addition, since high-level visual categories are associated with different neural substrates (e.g. Levy *et al.*, 2001; Grill-Spector and Malach, 2004; Kreichman and Gilaie-Dotan, 2024), we used images from multiple visual categories and for category-specific investigations we focused on faces, people, indoor and outdoor scenes. Furthermore, since our earlier findings were all obtained in lab-based settings, the online experiments allowed us to examine whether our lab results replicate and extend to a bigger image size range in online (less controlled) settings. This was important given the profound replication crisis that experimental science has undergone (e.g. Ioannidis, 2005) with very low replication rates in cognitive psychology.

## **Methods**

### **Reanalysis of lab-based data**

#### **Participants**

Data from Experiments 1-3 and 5 (101 participants, mean age  $27.17 \pm 6.86$  (SD) years, 58 women, all with normal or corrected to normal vision as checked in the lab before the experiment began, 93 right handers) that were already reported in our previous lab-based study (Masarwa, Kreichman and Gilaie-Dotan, 2022) were included in the reanalysis here that includes both accuracy of old (already seen) images, and signal detection theory measures ( $d'$  and decision criterion). We refer here to these experiments as Lab3°-21° (Experiments 1-3) and Lab3°-24° (Experiment 5). Note that data from Experiment 4 were not included as the experimental conditions involved several viewing distances, and

data from Experiments 6 and 7 were also not included since they entailed more complex designs and stimuli that were not used in the online experiments (see below). The experimental protocol was approved by the Bar-Ilan University Human Subjects Institutional Review Board (HSIRB). All the lab participants signed written informed consent before their participation.

### **Setup and experimental design**

We hereby provide a short description of the setup of the lab experiments (see further details in (Masarwa, Kreichman and Gilaie-Dotan, 2022)). Participants performed the experiments on a 24" screen (that could accommodate images of  $24^\circ \times 24^\circ$ ) at a viewing distance of 60 cm and the experiments were performed in a dark room. All experiments started with instructions to freely view and attend the images without notifying them about a memory test that will follow. The experiment started with an exposure phase of 160 presented images (2s/images, 500ms ISI) of different sizes (see below) and was followed by a surprise old-new recognition memory test presenting 320 images (160 old and 160 new, all sized  $\sim 8^\circ \times 8^\circ$ ), each presented for 500ms and participants were required to report whether they remembered seeing that image earlier or not. Experiments 1-3 included 2 image size conditions (smaller images subtending  $3.15^\circ \times 3.15^\circ$  and large images subtending  $20.78^\circ \times 20.78^\circ$ ) presented in four fixed-order blocks of 40 randomly ordered images each (2 blocks of bigger images and 2 blocks of smaller images). The images included different visual categories (faces, people, hands, animals, food, flowers, indoor places, outdoor places, and vehicles), and the images of each visual category were distributed equally between the small and large image sets. Experiment 1 (n=17) block order was large–small–small–large, Experiment 2 (n=16) had the same block order as Experiment 1 but with swapped image sets between the small and large image conditions, and Experiment 3 (n=17) included the same images sets as Experiment 1 but the block order was changed to small–large–large–small. Experiment 5 included 4 image size conditions in the exposure phase ( $3^\circ \times 3^\circ$ ,  $6^\circ \times 6^\circ$ ,  $12^\circ \times 12^\circ$ , and  $24^\circ \times 24^\circ$ ) that were presented each in one separate block (blocks and image order within a block were randomly ordered).

Images were taken from LaMem dataset (Khosla *et al.*, 2015) and included 4 categories (faces, people, indoor and outdoor scenes; see more details below in the online experiment description). Importantly, all image size blocks included 10 images from each category, and the new images were also matched in category to the old image category distribution.

## **Analyses**

Reanalysis of the data from Experiments 1-3 and 5 (termed here Lab3°-21° and Lab3°-24°) were based on the data shared on <https://osf.io/7sr3c/>. These included accuracy and signal detection theory measures of  $d'$  and decision criterion (see below in the Analyses section of the online experiments for further description of these procedures).

## **New online experiments**

### **Participants**

A total of 199 participants participated in the 3 online experiments (78 women, aged 18-35 (mean age  $26.29 \pm 4.35$  (SD) years), 21 left-handed (according to self-report), and all with normal or corrected to normal vision (according to self-report); see more details at Supp. Fig. S1 at [osf.io/m59zf/](https://osf.io/m59zf/)). In the 3°-15.5° experiment there were 132 participants, in the 1.5°-12° experiment there were 34 participants, and in the 3°-24° experiment there were 33 participants. Participants were recruited via Prolific online platform (<https://www.prolific.co/>) with the following recruitment criteria: age (18-35 years old), fluency in English, credible participation experience on Prolific (at least 200 runs and acceptance rate of 98% and above), and participation with a desktop computer/laptop (not handheld devices) which allowed presenting the images at sizes and viewing distances similar to those used in the lab. Each participant participated only once to ensure she/he were naïve to the study's memory aspects (for

additional performance-based exclusion criteria see Analyses below). While recruitment criteria and participant demographics were available via Prolific (demographic details of 4 participants were missing from Prolific, see [osf.io/m59zf/](https://osf.io/m59zf/)), the identity of the participants was completely anonymized such that we were blind to any identifying details (each participant was known by her/his Prolific-encrypted alphanumeric string). The experiments were approved by the Bar-Ilan University Human Subjects Institutional Review Board (HSIRB) (approval numbers ISU202007003 from 14.08.2020 and 060125410 from 06.01.2025) and all research was performed in accordance with the HSIRB and the Declaration of Helsinki guidelines. Prior to their participation participants were provided with an overall overview of the experiment (not exposing any information about the memory aspect of the experiment) as is common on Prolific. They were then asked to provide their informed consent to participate by agreeing that their data will be used (without identifying details) for research purposes before they could proceed to the actual experiment (upon their consent they were automatically directed to the actual experiment hosted on Pavlovia (see below)). All participants were paid at the end of their participation according to the criteria set by Prolific (regardless of their level of performance) and this was performed via the Prolific interface.

We have previously estimated sample size for lab-based experiments based on six experiments run in the lab from our previous study (Masarwa, Kreichman and Gilaie-Dotan, 2022) with a similar paradigm in which we tested the effect of picture size (large ( $21^\circ/24^\circ$ ) vs small ( $3^\circ$ )) on memory recognition accuracy (%) during naturalistic encoding. Specifically, we tested the difference in accuracy between large and small picture conditions for each of 16-26 individuals (i.e., paired samples). Based on the standard deviations of the differences and expecting in a lab-based study a mean difference of at least 15% in accuracy between the biggest and smallest image size conditions ( $24^\circ$  vs  $3^\circ$  with  $\alpha = 0.005$ ,  $\beta = 0.9$ ), the largest sample size required was estimated to be  $n = 24$ . Since these estimates were based on lab-based experiments but different online visual memory experiments sometimes recruit many more participants

taking into account the potential noise present in online data collection (e.g.  $n=100$  in each experiment in (Brady and Alvarez, 2015), and 219, 40, 61, 227 participants in the experiments in (Li, Bainbridge and Bakkour, 2022)), here we decided to recruit  $\sim 60$  participants for each version of the  $3^\circ$ - $15.5^\circ$  online experiment and smaller group sizes of  $\sim 30$  participants were recruited for the additional experiments. Exclusion criteria and participant exclusion details are provided in Analyses.

### **Experimental Design**

The experiments were created in PsychoPy (<https://www.psychopy.org/online/> (Peirce *et al.*, 2019) PsychoPy Builder v2021.2.3) and hosted on its online platform Pavlovia (<https://pavlovia.org/> (Gallant and Libben, 2019; Sauter, Draschkow and Mack, 2020)).

### ***Stimuli***

Images used in the experiment were precisely as those used in Experiment 5 in our lab-based experiment (see details above and full description can be found at (Masarwa, Kreichman and Gilaie-Dotan, 2022) and that study's experiments and results can be found at <https://osf.io/7sr3c/>). The images were taken from the "LaMem" Dataset (<http://memorability.csail.mit.edu/>) that includes 60,000 images (real world photographs) from different visual categories, many of them with per-image memorability scores (Khosla *et al.*, 2015). The four image categories selected from LaMem images were faces, people, indoors and outdoors. Face images included mainly a face without body parts while people included body parts and could also include more than one individual. Indoor and outdoor images did not include people in them, and face and people images were either indoors, outdoors, or with ambiguous background (face images: 19 with indoor background, 14 with outdoor background, 7 with ambiguous background; people images: 16 with indoor background, 22 with outdoor background, 2 with ambiguous background). While the LaMem image memorability scores for the images used in the exposure phase of this study varied (all images: 0.32-0.97, faces: 0.67-0.97, people: 0.69-0.94, indoors:

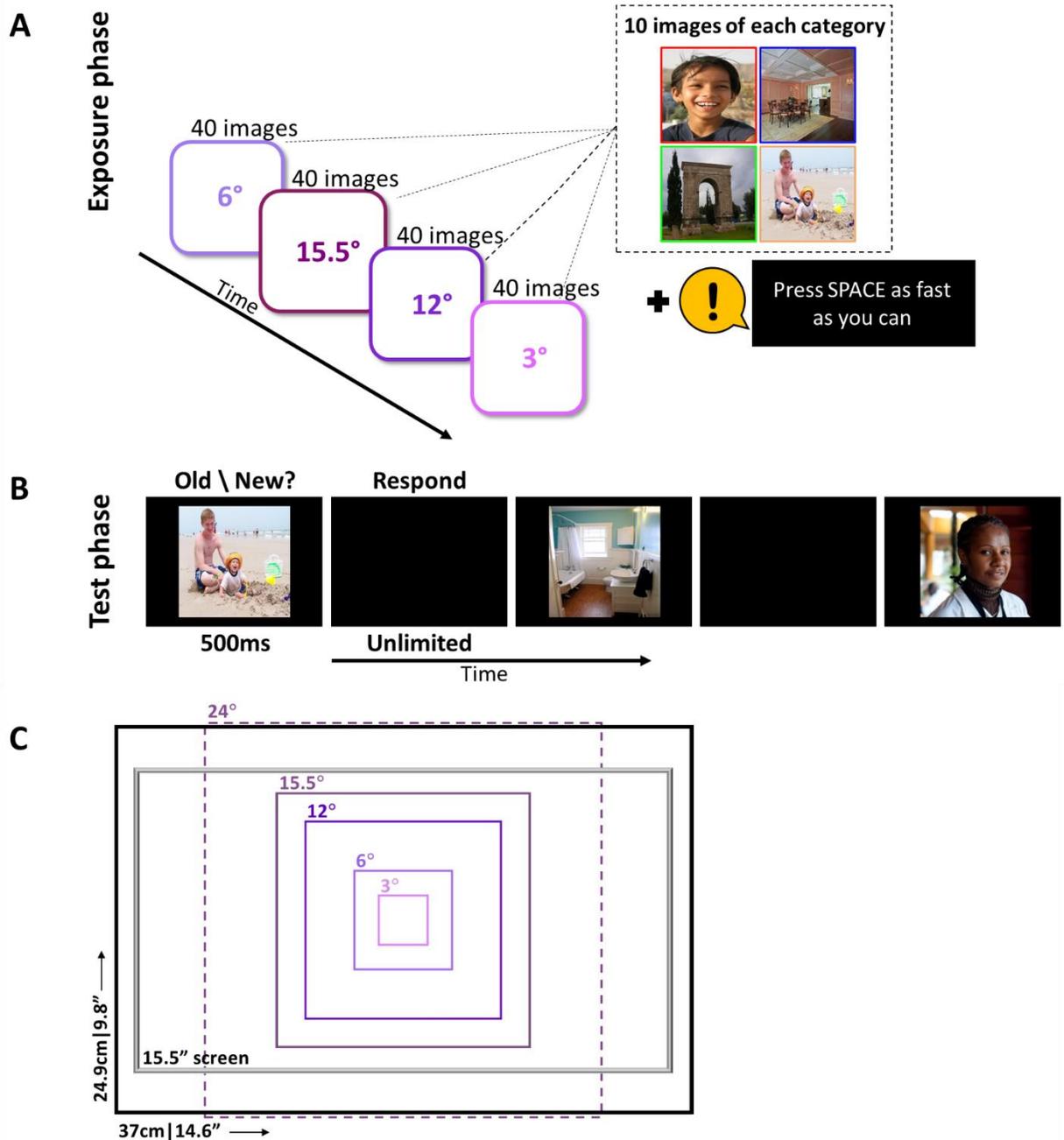
0.54-0.88, outdoors: 0.32-0.88), we made sure that the memorability scores of the images were comparable across size conditions. This was achieved by selecting for each visual category (faces, people, indoors, and outdoors) images of comparable memorability scores for each image size condition (see Fig. S1A in (Masarwa, Kreichman and Gilaie-Dotan, 2022)) and this assured that LaMem memorability scores across sizes were comparable. All images were resized to 900 × 900 pixels. These uniformly sized images were then scaled to be displayed according to the experimental condition (exposure: 3°, 6°, 12°, 15.5°, test: 8.5°) using the calibration procedure at the beginning of each experiment. Throughout the experiment the screen background color was black. We choose the largest images to be of size 15.5° since most desktop screens can fit 16 cm x 16 cm (6.3 inch x 6.3 inch) images without overflowing, corresponding to ~16.5°/15.5°/14.94°/14.5° when viewed from 55 cm/59 cm/61 cm/63 cm distance (height-wise (DisplaySize)). The test size condition remained at 8.5° as in the lab-based experiment to maintain the test-exposure size relations for the 3°, 6°, and 12° conditions.

### ***Procedure***

The following procedures describe the procedure of the 3°-15.5° online experiment and specific changes that took place in the other two online experiments appear afterwards. Participants were first presented with a demographics questionnaire (age range, gender, handedness, vision, highest level of education, see Fig. S1 at [osf.io/m59zf/](https://osf.io/m59zf/)), then continued to calibrating image size on their screen (see more details below under "Screen calibration"), and to consenting having their data used (without identifying details) for research purposes. They then proceeded to the actual experiment. The experiment started with the exposure phase where participants were asked to observe each image on the screen and follow any instructions that may appear (without any information about the memory aspect of the experiment that were to follow later). The instructions were presented until the space key was pressed, after which images were sequentially presented (2000 ms/image followed by 500ms ISI). This exposure phase

included 160 images presented in four blocks, each block of a fixed image size (either 3°, 6°, 12°, or 15.5°) presented 40 images in that size (10 from each visual category: faces, people, indoors, and outdoors); block order and image order within each block were random (see Fig. 1). After every 32 images an attention check that they were not notified about earlier appeared (black screen with “Press SPACE as fast as you can”) where they were asked to press the spacebar as fast as they could in order to continue the experiment (see more details below under “Attention checks during exposure phase”). There were 4 attention checks during the exposure phase. After the exposure phase (160 presented images) additional instructions (for the test phase) appeared. Participants were presented with surprise recognition memory instructions (to press leftward arrow if they recall seeing the image earlier or rightward arrow if not) and asked to press the shift key to proceed (the key here was different than that used for the attention checks in order for the participants not to miss this instruction screen by quickly pressing the space key assuming that it was another attention check). This was followed by sequential presentation of 320 mid-sized images (one 8.5° × 8.5° image per trial, 50% of the images were old, 500ms each, randomly ordered, new images matched for the categories of the old images). After the participant responded the next image was presented (no feedback was provided). After every 100 images a take-a-short-break screen appeared that allowed participants to take a short break and resume the experiment by pressing the space bar (overall there were 3 take-a-short-break screens during the experiment). The experiment ended with an end of experiment screen and participants were then directed back to Prolific with a completion code.

We ran two versions of the exposure phase where the version difference only entailed swapping image sets between size conditions (all 40 images presented in 3° in version 1 (i.e., the face, people, indoor and outdoor images) were presented in 15.5° in version 2 and vice versa, and all 40 images presented in 6° in version 1 (i.e., the face, people, indoor and outdoor images) were presented in 12° in version 2 and vice



**Figure 1. Online 3°-15.5° experimental paradigm.** (A) Exposure phase: participants viewed 160 images (from “LaMem” Dataset) in four blocks, each block included 40 images of a specific size (3°, 6°, 12°, 15.5°) with 10 images of each visual category (faces [red], people [orange], indoors [blue], and outdoors [green]). Block order and within-block image order were randomized across participants. Images were presented for 2 seconds followed by a black screen of 500ms. Participants were instructed to freely view the images without any memory or encoding instructions and were not informed of any memory-related task that would follow (“naturalistic encoding”). No response was required. Several pseudo-random attention checks (independent of the viewed images; black screen with “Press SPACE as fast as you can” instructions) appeared during this exposure phase to assess participant’s engagement in the experiment. Participants whose response times in these attention checks were too slow were excluded

from further analyses (see Methods). (B) Test phase: after the exposure phase participants were given a surprise old/new recognition memory test for sequentially presented 320 (160 old) intermediately-sized (8.5°) images. Participants were required to report for each image if they recall seeing it earlier (“old”) or not (“new”). Images were presented in random order for 500ms each after which a black screen appeared until a response was given (with no time limit). No feedback was given. The experimental paradigms of the other two online experiments (1.5°-12° and the 3°-24° experiments) and of the Lab3°-24° experiment were very similar (see Methods for details). (C) A visualization of the image size conditions (during the exposure phase) used in the online experiments as they appeared on a typical 15.5” screen. The borders of the actual 15.5” screen are indicated by the gray line, and the black rectangle delineates the laptop itself (its height and width at the bottom left of the outer black rectangle are 24.9 cm/9.8” high, 37cm/14.6” wide). The lighter (smaller) to darker (larger) purple squares represent the size conditions (in visual angle) relative to that screen size. Note that the 24° image size extends beyond the actual screen size indicating that the upper and lower borders of these images were not visible during the exposure phase in that condition in the 3°-24° online experiment.

versa). Thus, each experimental version presented 4 image size blocks where each block consisted of images from each of the 4 visual categories.

The procedures for the smaller images 1.5°-12° experiment were precisely as those of the 3°-15.5° experiment except for following differences: (i) it was only based on the version 1 of the 3°-15.5° experiment, (ii) exposure phase image sizes were 1.5°, 3°, 6°, or 12° and (iii) test phase image size was 4.25°.

The procedures for the bigger images 3°-24° experiment were precisely as those of the 3°-15.5° experiment except for (i) being based on the version 1 of the 3°-15.5° experiment, and (ii) exposure phase image sizes were 3°, 6°, 12°, or 24°.

### **Screen calibration**

We aimed to ensure image size on the participant’s screen in the online experiments was consistent across participants and similar to that presented in the lab. To that end we calibrated the screen size in the beginning of the experiment which allowed us to present the experimental stimuli in sizes similar to those used in the lab assuming a 55-65 cm viewing distance (Rempel *et al.*, 2007; Bilton, 2011; Jaiswal *et*

*al.*, 2019). This was done using a screenscale module in PsychoPy, see example at <https://run.pavlovia.org/Wake/screenscale/>. While we cannot be sure of the precise viewing distance of the participants and therefore the visual angles that the stimuli occupied for each of them, our study modulated image sizes for each participant, and this allowed us to examine how differences in image size affected memory in a within-participant fashion.

### **Attention checks during exposure phase**

Most online cognitive experimental studies employ an experimental task when stimuli are being presented (Brown *et al.*, 2014; Bainbridge, 2017; Kardosh *et al.*, 2022; Sahar and Yeshurun, 2024) typically requiring participants to attend and oftentimes provide responses to the presented stimuli. Since our paradigm relies on more naturalistic-like encoding which is less controlled by its nature (freely viewing images for ~10 min without any expected response such that participants can even be away from the screen during this exposure period) and is much less common in scientific investigations, it was necessary for us to assess in some way that the participants were genuinely viewing the images. To address this issue, we incorporated pseudo-random simple attention checks during the exposure/encoding phase of the experiment allowing us to detect and exclude participants who were likely away from the computer or inattentive. The attention checks appeared after every 32 images where each block consisted of 40 images (one attention check in each block, after the 32<sup>nd</sup> image of block 1, after the 24<sup>th</sup> image in block 2, after 16<sup>th</sup> image in block 3, after 8<sup>th</sup> image in block 4). Since the participants were not notified in advance that there will be such checks during the experiment, they seemed to the participants to occur randomly.

### **Analyses**

Data of each participant were recorded as a CSV file and this included response times and correctness per each presented image. These CSV files were then analyzed with R (Rstudio 2021.9.0.351 (RStudio

Team, 2020) in-house scripts) to obtain per-condition results across participants, per-image (memorability-related) results, as well as statistical analyses.

**Performance-based exclusion criteria (independent of our research question).** The following were applied to reduce potential noisy data. We first excluded participants with 0% performance for the old images and 100% in the new condition (no participants were excluded), then participants who failed our **attention checks** (i.e. failed 1 or more of the 4 attention checks (taking longer than 8 seconds to respond to the first attention check or more than 3 seconds to respond to any of the 3 consecutive attention checks)). This resulted in 5 (out of 34 in the 1.5°-12° exp., 17.2%), 13 (out of 132 in the 3°-15.5° exp., 9.8%), and 5 (out of 33 in the 3°-24° exp., 15.1%) participants being excluded.

**Analysis rational.** Each experiment included 4 image sizes that reflect its main experimental conditions such that together the 3 experiments included 6 experimental conditions (image sizes 1.5°-24° reflecting the 6 levels of the independent variable). The dependent variables were recognition accuracy to test replication of our earlier lab-based findings (Masarwa, Kreichman and Gilaie-Dotan, 2022), *dprime* to examine whether memory sensitivity is significantly above chance level and is affected by image size, decision criterion to examine whether image size affects individual response biases, and RTs to test whether image size influences performance speed. Additionally, we also examined whether overall memory (with *dprime* and AUC measures) were significantly above chance. In the 3°-15.5° experiment, data from both experimental versions were analyzed together.

**Accuracy and RT.** Memory recognition accuracy was calculated as the proportion of correct answers for the old (already seen) images while the “new” images accuracy was calculated as the proportion of correct answered for the unseen (“noise”) images. Reaction times (RTs) reflect the median RT relative to image onset for all the images in the test phase.

**Signal detection theory (SDT)-based analyses.** Individual  $d'$  ( $d'$ ,  $Z(\text{hits}) - Z(\text{FA})$ ) and criterion ( $-0.5(Z(\text{hits}) + Z(\text{FA}))$ ) measures were calculated both (i) across all the experimental trials, and (ii) for each size condition separately. Since during the test phase all the images were presented in the same size and not according to the sizes in the exposure phase, all the new images served as noise for the SDT calculations (i.e. there was no a-priori selection bias of specific new images to serve as noise for any of the old image sizes).

**AUC of the RT-based ROC.** Since we did not collect confidence ratings in our study, we followed a method proposed by Weidemann and Kahana (Weidemann and Kahana, 2016) that uses RTs as an estimate of confidence (faster RTs for higher confidence, slower RTs for lower confidence) to calculate RT-based ROC curves. Individual ROC curves were calculated and from them also the area under the ROC curve (AUC). This curve was created for each participant based on all the trials of the experiment taking all the old images (“targets”) as the signal and all the new images (“lures”) as the noise.

**Statistical analyses.** Repeated measures ANOVAs were performed using the R `anova_test` and `get_anova_table` functions where `get_anova_table` automatically applies the Greenhouse-Geisser sphericity correction only to factors violating the sphericity assumption (i.e., Mauchly’s test significant  $p$ -value  $\leq 0.05$ ). Effect size estimates were obtained using the `anova_test` function with the `effect.size` parameter set to “ges” which provides generalized eta squared values. Friedman rank sum non-parametric test was used on median RTs (`friedman.test` R function).

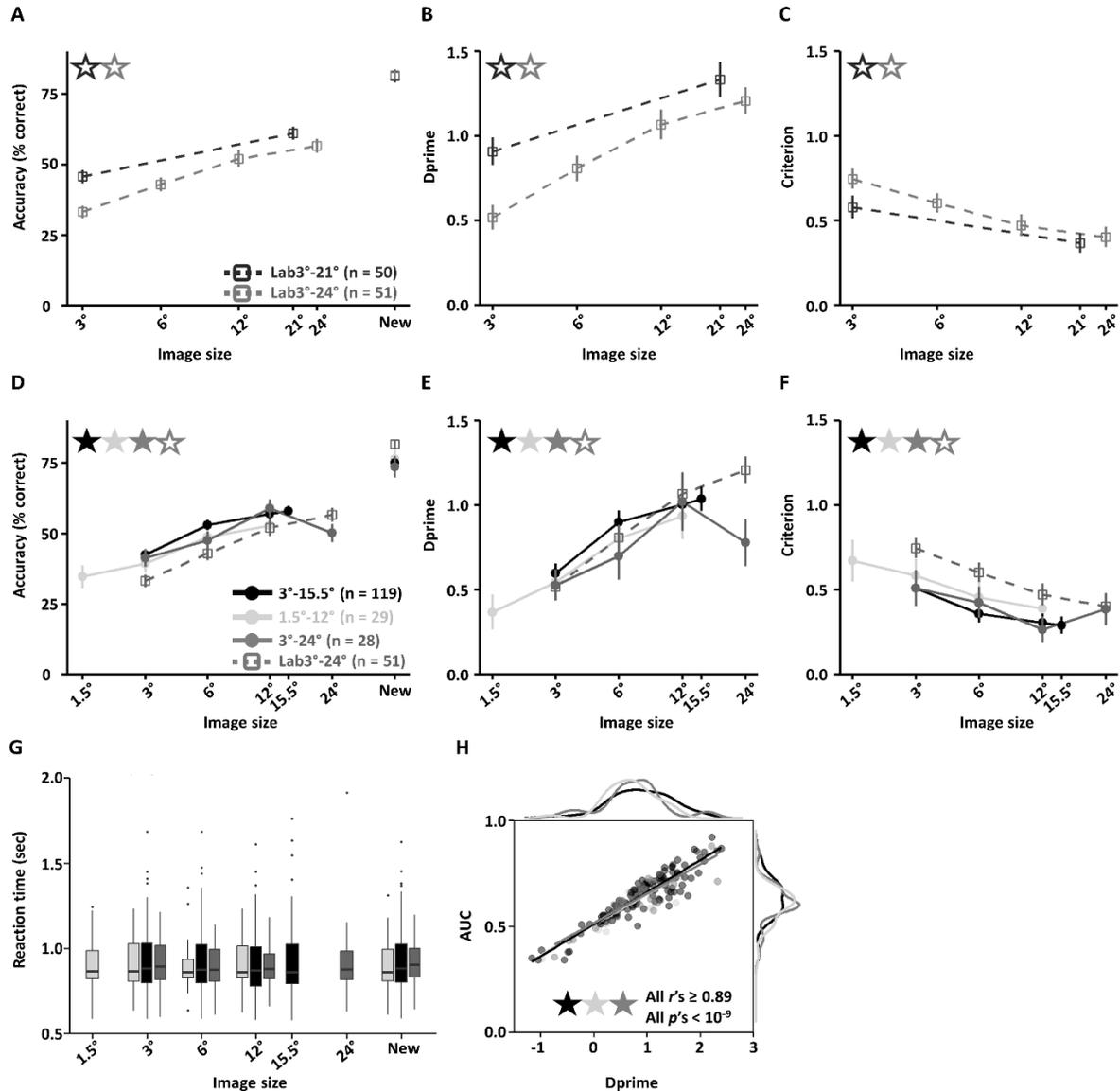
**Memorability analyses.** For each image and presentation size its memorability was defined as the proportion of participants that remembered it out of the group that underwent that experimental version. First, we ran consistency analysis to examine if memorability was consistent across different groups of participants. In each experiment we split the participant group (before exclusion) into two sub-groups of equal sizes and calculated for each image its memorability according to that specific sub-

group. These two different memorability scores (one per sub-group) were then compared for consistency by testing their correlation (Pearson). In addition, we also took advantage of the fact that in the 3°-15.5° experiment each image was presented in a different size in each of the two experimental versions (images presented in 3° in version 1 were presented in 15.5° in version 2 and vice versa, and images presented in 6° in version 1 were presented in 12° in version 2 and vice versa) to examine whether images were more memorable when presented in larger sizes relative to smaller size presentation (each image was presented in two different sizes in the two experimental versions). Using these memorability values we first correlated these values to assess whether image memorability was correlated across presentation sizes. We then also estimated the difference in memorability associated with difference in presentation size by subtracting for each image its memorability when presented in smaller size (in one of the experimental versions) from its memorability when presented in larger size (in the other experimental version).

## Results

### Bigger images are better remembered in the 3°-12° size range

The memory performance based on the reanalysis of the data from our earlier lab experiments (Masarwa, Kreichman and Gilaie-Dotan, 2022) by accuracy,  $d'$ , and criterion are presented in Figure 2 A-C and in Table 1. Since accuracy levels were relatively low, we examined whether memory sensitivity ( $d'$ ) was significantly above chance or guessing behavior. First, we found that the overall  $d'$  was modest but significantly above chance (i.e. above 0 (Lab3°-21°:  $M = 1.12$ , 95% confidence interval [0.94, 1.29],  $t(49) = 12.59$ ,  $p < 2.2 \cdot 10^{-16}$ ; Lab3°-24°:  $M = 1.09$ , 95% confidence interval [0.95, 1.23],  $t(50) = 15.49$ ,  $p < 2.2 \cdot 10^{-16}$ , 1-sampled  $t$ -tests, see Tables 1 and 3)) and this was also the case even for the smallest image size condition of 3° (Lab3°-21°:  $M = 0.91$ , 95% confidence interval [0.74, 1.07],  $t(49) =$



**Figure 2. Image size affects memory performance across experiments and measurements ( $n_{\text{lab}} = 101$ ,  $n_{\text{online}} = 176$ ).** (A)-(C) Reanalysis of lab-based results, (D)-(H) Online experimental results. (A) Test phase accuracy (% correct) for the surprise old/new recognition memory task (on the y-axis) by image size condition (x-axis) for the lab experiments (Lab3°-24° experiment in lighter gray, Lab3°-21° experiment in darker gray). Significant effect of image size on image memory accuracy was found in each experiment (all  $p$ 's  $< 10^{-10}$ ). (B, C) The same as in A but for memory-sensitivity (dprime) in B and decision criterion in C showing image size significantly improved memory sensitivity (in B,  $p < 10^{-10}$ ) beyond the significant shift to a more liberal criterion with growing image size (in C,  $p < 10^{-10}$ ). (D) Same as in A for each of the 3 experiments (3°-15.5° experiment in black, 1.5°-12° experiment in light gray and 3°-24° experiment in dark gray). Significant effect of image size on image memory was found in each experiment (all  $p$ 's  $< 10^{-5}$ ). The results of Lab3°-24° from A are also superimposed on the graph as reference. (E, F) The same as in D but for memory-sensitivity (dprime) in E and decision criterion in F showing image size significantly improved memory sensitivity (in E,  $p < 10^{-4}$ ) beyond the significant shift to a more liberal criterion with growing image size (in F,  $p < 10^{-4}$ ). Here too the dprime (B) and criterion (C) curves of Lab3°-24° (lighter gray) are superimposed as references. (G)

Reaction time median results per condition with boxplot notations (boxes from the first quartile to the third quartile, the vertical line within the box represents the median, whiskers denote  $1.5 \times \text{IQR}$  (Tukey, 1977; McGill, Tukey and Larsen, 1978), outliers indicated by dots; see Methods); no effect of size on RT was found. (H) Area under the curve (AUC) of RT-based ROC curves (y-axis (Weidemann and Kahana, 2016)) were significantly predictive of  $d'$  (memory sensitivity, x-axis). Each point represents one participant with the same color notations as in C-F. Marginal distributions of  $d'$  (above graph) and of AUC (on right) for each of the experiments. Both  $d'$  and AUC were significantly above chance (guessing performance;  $d' > 0$ ,  $\text{AUC} > 0.5$ , see Results). Error bars in A-F represent SEM. Asterisks in (A)-(F) denote significant effects of image size and in (H) correlation significance.

$11.23$ ,  $p = 3.7 \times 10^{-15}$ ; Lab3°-24°:  $M = .52$ , 95% confidence interval [0.37, 0.66],  $t(50) = 7.24$ ,  $p = 2.6 \times 10^{-9}$ , 1-sampled  $t$ -tests, see Tables 1 and 3). We then examined whether signal detection theory measures were influenced by image size, and found that  $d'$  significantly increased with image size (Lab3°-21°:  $t(49) = 8.47$ ,  $p = 3.6 \times 10^{-11}$ , paired two-tailed  $t$ -test; Lab3°-24°:  $F(3,150) = 55.85$ ,  $p = 2.68 \times 10^{-24}$ ,  $\eta_G^2 = .184$ , 1-way repeated measures ANOVA; see Table 2) and the criterion significantly decreased with image size (Lab3°-21°:  $t(49) = 8.47$ ,  $p = 3.6 \times 10^{-11}$ , paired two-tailed  $t$ -test; Lab3°-24°:  $F(3,150) = 55.85$ ,  $p = 2.68 \times 10^{-24}$ ,  $\eta_G^2 = .086$ , 1-way repeated measures ANOVA; see Table 2). These indicate that memory sensitivity increased with image size beyond the changes in decision criterion that image size entailed.

To investigate whether these results generalize in online settings which may be more variable settings than those in our lab, we ran an online experiment that mimicked the parametric image-size experiment from the lab (Lab3-24, Experiment 5 from (Masarwa, Kreichman and Gilaie-Dotan, 2022)). We kept the 3 smaller image sizes at exposure (3°, 6°, 12°) as in the lab and decreased the largest image size at exposure to 15.5° (from 24°) for it to fit most laptop screens. The image size at test remained as in the lab (8°) so that the relative test-to-exposure image size would be the same as it was in the lab for the 3 smaller image size conditions. The memory performance as tested in the test phase for each image size condition by accuracy,  $d'$ , criterion, and also AUC as well as reaction times are presented in Figure 2D-H and Table 1. In this online experiment (3°-15.5° experiment, in black in Figure 2), we found that our lab-based findings for recognition memory accuracy for already seen images replicated such that

memory accuracy for larger images was significantly better than for smaller images (one-way repeated measures ANOVA: main effect of image size:  $F(2.73, 322.02) = 46.8, p = 1.98 \times 10^{-23}, \eta_G^2 = .088$ ;

Bonferroni/Dunn post hoc:  $3^\circ$  vs.  $6^\circ$ :  $p = 2.27 \times 10^{-10}$ ,  $6^\circ$  vs.  $12^\circ$ :  $p = .049$ ; see Table 2). Here too, we examined whether memory performance was significantly above chance or guessing behavior.

Consistently with the results from the lab (see above), we found that the overall  $d'$ prime was significantly above chance (i.e. above 0,  $M = 0.88$ , 95% confidence interval [0.76, 1.00],  $t(118) = 14.01, p < 2.2 \times 10^{-16}$ , 1-sampled  $t$ -test, see Tables 1 and 3). We also found that the AUC was significantly above chance (i.e. above 0.5,  $M = 0.64$ , 95% confidence interval [0.62, 0.66],  $t(118) = 13.9, p < 2.2 \times 10^{-16}$ , 1-sampled  $t$ -test). In addition, AUC was significantly correlated with  $d'$ prime ( $r = 0.92, t(117) = 25.8, p = 4.5 \times 10^{-50}$ ). These results indicate that memory performance was modest but significantly above chance. Examining whether signal detection theory measures are influenced by image size, we found again that  $d'$ prime (memory sensitivity) significantly increased with image size ( $F(2.72, 320.92) = 44.79, p = 1.67 \times 10^{-22}, \eta_G^2 = .052$ , see Table 2) and the criterion significantly decreased with image size ( $F(2.72, 320.92) = 44.79, p = 1.67 \times 10^{-22}, \eta_G^2 = .022$ ). No effect of image size on RT was found (see Figure 2G and Tables 1 and 3 for details). This supports our lab-based findings that memory sensitivity increased with image size beyond the changes in decision criterion that image size entailed.

To further test whether image size effects on memory sensitivity was influenced by the platform the experiment was run on (lab vs online) we ran a 2-way mixed ANOVA with platform (non-repeated) and image size (repeated) for the  $3^\circ, 6^\circ, 12^\circ$  image size conditions that overlapped between the platforms.

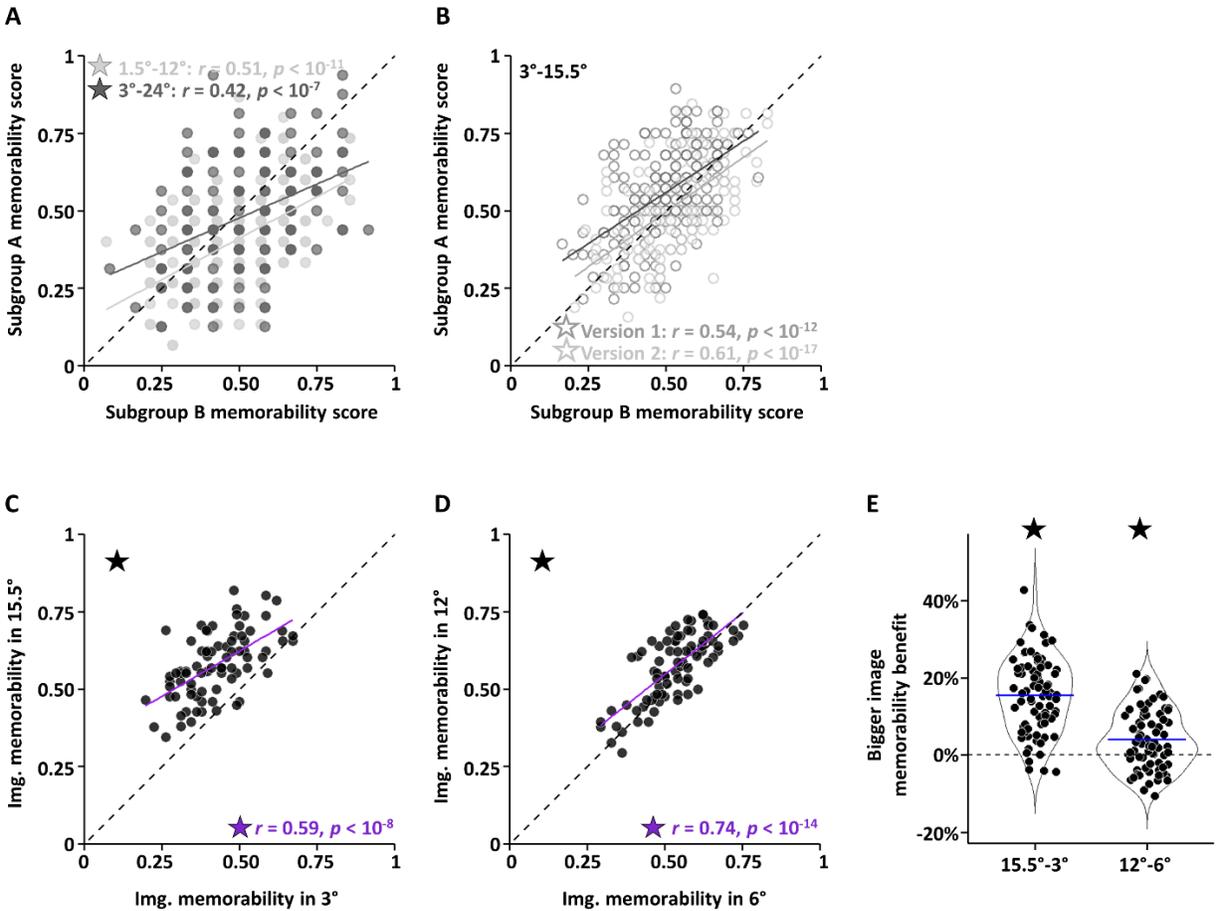
While as expected we found that image size affected memory sensitivity ( $F(2, 336) = 77.39, p = 2.27 \times 10^{-28}, \eta_G^2 = .066$ ), no effect of platform ( $F(1, 168) = 0.109, p = 0.74, \eta_G^2 = .0005$ ) and no interaction between platform and image size was found ( $F(2, 336) = 2.46, p = 0.087, \eta_G^2 = .002$ ).

To further examine the ranges of image sizes that influence memory performance we ran two additional experiments. The results of the first with image sized 1.5°-12° are presented in Figure 2 in light gray shades and of the second with image sized 3°-24° are presented in Figure 2 in dark gray shades. In both experiments we found that image size significantly affected memory performance across all measures in the same manner as in the 3°-15° experiment (see Figure 2 and Table 2). In addition, in both experiments we also found that memory was above chance both in  $d'$  and in AUC measures and even for the smallest image sizes (see Tables 1 and 3). Furthermore, AUC was significantly correlated with  $d'$  in each of these experiments (Table 3). Post-hoc examinations revealed that memory performance was not significantly different between 1.5° and 3° images (accuracy:  $p = 0.88$ ,  $d'$ :  $p = 0.62$ , see Tables 1 and 2) and that it declined for images sized 24° relative to 15.5° (accuracy:  $p = 0.037$ ,  $d'$ :  $p = 0.04$ , see Tables 1 and 2). We assume that the latter can be attributed to overscan during the exposure phase due to screen-size limitations relative to the screen size used in our lab-based investigations. In addition, another control experiment where images were scaled relative to the participant's screen also replicated these findings (see <https://osf.io/m59zf/>).

Consistent with earlier studies and our lab-based memory-accuracy findings, there was a significant effect of image category on memory sensitivity that was evident in face and people images being remembered more and outdoor and indoor images significantly less across all the experiments we ran (see details in <https://osf.io/m59zf/>). Importantly, in all experiments image size affected memory-sensitivity for each category and no interactions between image size and category on memory were found in any experiment apart from the 3°-24° online experiment (see details in <https://osf.io/m59zf/>).

**Image memorability is consistent and influenced by image size**

Memorability consistency analysis of the online experiments (see Figure 3 A and B) revealed that memorability was consistent across sub-groups in each of the experiments ( $1.5^{\circ}$ - $12^{\circ}$ :  $r = 0.51$ ,  $r^2 = 0.27$ ,  $t(158) = 7.54$ ,  $p = 3.42 \cdot 10^{-12}$ ,  $3^{\circ}$ - $24^{\circ}$ :  $r = 0.42$ ,  $r^2 = 0.18$ ,  $t(158) = 5.85$ ,  $p = 2.81 \cdot 10^{-8}$ ,  $3^{\circ}$ - $15.5^{\circ}$  (version 1):  $r = 0.54$ ,  $r^2 = 0.29$ ,  $t(158) = 8.09$ ,  $p = 1.49 \cdot 10^{-13}$ ,  $3^{\circ}$ - $15.5^{\circ}$  (version 2):  $r = 0.61$ ,  $r^2 = 0.38$ ,  $t(158) = 9.81$ ,  $p = 5 \cdot 10^{-18}$ ). These results show that there was a very significant positive relation between memorability scores in each of the study experiments, with 18% and up to 38% of the variance in image memorability scores of one subgroup explained by image memorability scores of the other subgroup. These substantiate the idea that image memorability is a consistent image property (Bainbridge, 2020; Rust and Mehrpour, 2020). The two versions of the  $3^{\circ}$ - $15.5^{\circ}$  experiment (each image appearing in a different size in each version) allowed us to test for each image whether presenting it in a bigger size leads to higher overall memorability across the group relative to when it is presented in a smaller size. Here, we also found that memorability of images presented in one size were significantly predictive of their memorability when presented in another size (see Fig. 3 panels C and D in purple), and this was true for both  $15.5^{\circ}$  and  $3^{\circ}$  ( $r = 0.59$ ,  $r^2 = 0.35$ ,  $t(78) = 6.47$ ,  $p = 7.8 \cdot 10^{-9}$ ; non-directional Pearson correlation) and for  $12^{\circ}$  and  $6^{\circ}$  ( $r = 0.74$ ,  $r^2 = 0.55$ ,  $t(78) = 9.8$ ,  $p = 3.08 \cdot 10^{-15}$ ). These results further show that there is a very significant positive relation between memorability scores even when image size changes, with 35%-55% of the variance in image memorability scores of one image size explained by image memorability scores of another image size. We also replicated our lab-based findings that most images were better remembered when presented as big (above the diagonal equality line) relative to when they were presented as small and this can be seen for the  $15.5^{\circ}$  vs  $3^{\circ}$  (Fig. 3C) and for the  $12^{\circ}$  vs  $6^{\circ}$  (Fig. 3D). The size-related memorability benefit was significant (see Fig. 3E) both when comparing image size presentation of  $15.5^{\circ}$  to  $3^{\circ}$  ( $n = 80$  images, higher memorability of  $15.47\% \pm 1.08\%$  (SE) for  $15.5^{\circ}$  relative to that of  $3^{\circ}$ ,  $t(79) = 14.25$ ,  $p < 2.2 \cdot 10^{-16}$ , paired t-test) and also when comparing image size presentation



**Figure 3. Image memorability is consistent and is influenced by image size during naturalistic encoding (per-image analysis,  $n = 160$  images).** (A) Within experiment image memorability consistency analysis for 1.5°-12° experiment (light gray) and 3°-24° experiment (dark gray) between subgroup A (y axis) and subgroup B (x axis) reveals significant correlations between memorability scores. Each image represented by one point (note that several images may be represented by the same point in the graph due to overlapping memorability scores). (B) Within-version image memorability consistency analysis in the 3°-15.5° experiment (as in (A); version 1 in darker gray, version 2 in lighter gray) again reveals significant correlations between memorability scores. (C) In the 3°-15.5° experiment, each image (indicated by a black dot,  $n = 80$  images) memorability when presented as 15.5° (y axis) is plotted against its memorability when presented as 3° (x axis). Note that each image was presented in one version as big and in the other as small since the sizes of each image were swapped across versions. Image memorability was significantly correlated across image sizes (purple regression line). In addition, most images fell above the equality line (diagonal) indicating that they were better remembered (higher memorability) when presented in the larger size. (D) Same analysis as in C for the memorability of the images ( $n = 80$  images) presented as 12° (y axis) relative to when presented as 6° (x axis). Here too image memorability was significantly correlated across image sizes and most images fell above the equality line (albeit to a lesser magnitude than in C; same conventions as in C). (E) In the 3°-15.5° experiment, for each image, its memorability benefit (the difference in its memorability when presented in a larger size relative to when presented in a smaller size). Data from C on left, data from D on right, mean of each indicated by a blue line. Asterisks denote significant consistency in A, B and in purple in C and D. Black asterisks in C-E denote significant differences in memorability of the larger format relative

to the small format (all  $p$ 's  $< 10^{-5}$ , paired  $t$ -tests). Note that image memorability was significantly correlated within experiments (A, B) and across image sizes (C, D) in line with the findings that image memorability is an image-specific property (Bainbridge, 2020; Rust and Mehrpour, 2020).

of  $12^\circ$  to  $6^\circ$  ( $n = 80$  images, higher memorability of  $4.09\% \pm 0.85\%$  (SE) for  $12^\circ$  relative to that of  $6^\circ$ ,  $t(79) = 4.79$ ,  $p = 7.47 \cdot 10^{-6}$ ; paired  $t$ -test; averages denoted by blue line in Fig. 3E).

## Discussion

Here we examined whether the effect we found in the lab, that image size affects recognition memory accuracy during naturalistic encoding, reflects genuine memory performance or may partially reflect individual response biases. In addition, since in that study memory accuracy was rather low, here we evaluated whether this was above chance or guessing performance especially for smaller images. While reanalyzing the data from the lab and running new online experiments, we found that image size effect on memory was significantly reflected in memory sensitivity measures ( $d'$ ) such that larger images were better remembered in the range of  $3^\circ$ - $12^\circ$ . Across measurements ( $d'$ , AUC) we found that memory performance was above chance indicating that participants had some image knowledge even for the smaller images which allowed them to build image representations. Given that these findings were also evident in more variable and potentially noisier online settings further substantiates the findings that some image knowledge was present even for the smaller images. Additionally, we also found that memory for images of size  $1.5^\circ$  was not worse than that of  $3^\circ$  images, and that memory for images of  $24^\circ$  was lower than that of images of  $15.5^\circ$  which we assume can be attributed to overscan. Taken together, our results indicate that during naturalistic encoding image size affects image memory. Since memory recognition accuracy measures have been suggested to potentially reflect response biases, it is unclear whether the low memory accuracy performance we found in our earlier study truly measures memory. The overall memory accuracy levels we reported earlier (Masarwa, Kreichman and

Gilaie-Dotan, 2022), even for the largest images, may appear not much higher than guessing levels. We followed suggestions and techniques recently presented (Weidemann and Kahana, 2016; Brady *et al.*, 2023) and found that overall memory as measured by  $d'$ prime and by AUC were reliably above chance. These results indicate that the measured memory performance, both in the lab and in the online experiments, even though appearing relatively low, likely surpassed guessing behavior even for the smallest sized images. The knowledge learned or acquired from past experiences (e.g. images viewed) can vary according to multiple factors. An example of such variation has been suggested in the seminal work of Craik and Lockhart (Craik and Lockhart, 1972), who suggested that the encoding task determines the extent of knowledge (i.e. memory) registered from that experience. Specifically, they argued that the level of processing applied to a stimulus or of an experience directly influences the level of knowledge acquired. Support for this framework comes from multiple studies (e.g., Bower and Karlin, 1974; D'Agostino, O'Neill and Paivio, 1977). Our experimental paradigm can be considered as one of the shallowest levels of processing (when participants are passively viewing images and are assumed to have no knowledge that they will be required to report anything about the images at a later stage). Yet, despite the shallow level of processing, some information was registered as evident by both  $d'$ prime and RT-based ROC measures, which were interestingly highly predictive of each other with 79%-85% of the variance explained (Weidemann and Kahana, 2016). Although much more image information can be registered to memory when processing is deeper (Brady *et al.*, 2008; Konkle *et al.*, 2010), our results indicate that during more naturalistic experimental encoding some information from the viewed images is registered even for the smallest images.

While in our earlier study (Masarwa, Kreichman and Gilaie-Dotan, 2022) we found that larger images are better remembered, this finding was based on recognition memory accuracy that has been criticized for potentially reflecting shifts in response biases (e.g. Brady *et al.*, 2023). Indeed, when examining this directly, we found that decision criterion shifted to be more liberal with increasing image size. However,

beyond this decision criterion shift we also found that memory sensitivity increased significantly with larger image sizes, substantiating our earlier findings. Interestingly, even for the smallest images, memory sensitivity was above chance. Since visual processing is significantly driven by retinal image size in all early processing stages from the retina through LGN and in retinotopic cortex (Tootell *et al.*, 1997; Grill-Spector and Malach, 2004; Schneider, Richter and Kastner, 2004), bigger images are represented by significantly more visual system processing resources than smaller images. While higher-order visual areas are assumed to be the ones that are tightly related to visual memory (Felleman and Van Essen, 1991; Golarai *et al.*, 2007; Garrido *et al.*, 2009; Gilaie-Dotan *et al.*, 2015, 2019; Gomez *et al.*, 2015) our results highlight the contribution of image size and early visual processing stages to image memory. We assume that especially when no task-based modulations are present (Hillyard, Vogel and Luck, 1998; Huk and Heeger, 2000; Saalman and Kastner, 2009), these early visual processing contributions increase the quality of the signal (i.e. higher signal to noise ratio (SNR)) transferred downstream to high-order visual cortex and allow more information to be passed onto memory-related areas.

As presentation sizes may have been less consistent across participants in the online experiment relative to consistent presentation sizes across the lab participants, the size effect on memory we investigated here is first and foremost within participant. The precise visual angle an image occupies relies both on its physical size and on the viewing distance. While in the lab both of these factors were precisely controlled for (e.g. Kreichman, Bonneh and Gilaie-Dotan, 2020; Akselevich and Gilaie-Dotan, 2022; Masarwa, Kreichman and Gilaie-Dotan, 2022), in our online experiments we were only able to precisely control for the physical image presentation size (using a screen calibration procedure) while viewing distance may have varied across participants (Rempel *et al.*, 2007; Bilton, 2011; Jaiswal *et al.*, 2019). Previous studies suggest that when using desktops or laptops the viewing distance is typically 2 feet (~55-60 cm) ranging from 43 to 99 cm (Jaschinski, 2002; Rempel *et al.*, 2007). Thus, even if an individual was seated at a different viewing distance than we anticipated (55-60 cm), the different experimental

conditions did vary in their presentation sizes on her/his screen. In the lab we found that the effect of image size on memory occurred for image sized 3°-12°. Given the potential variability in viewing distances the experimental 3°-12° online conditions could have effectively occupied 4.12°-16.44° retinal sizes (for a viewing distance of 40 cm) to 1.65°-6.61° (for a viewing distance of 1 m). In both of these extreme cases of viewing distances the effect of image size on memory was expected to be significant given our prior lab-based results showing that image size affects memory in these image size ranges.

Image memorability is considered an image-specific trait (Bainbridge, 2020; Rust and Mehrpour, 2020) and indeed we found evidence for that across our analyses. For example, in our consistency analysis we found significantly consistent memorability measures across different participant subgroups, and also across varying image sizes. Interestingly, in addition, we also found that physical image size contributes to image memorability during naturalistic encoding (when freely viewing the images and not being aware of any memory or other test that would follow) such that bigger images increase image memorability. While these two results (consistency in memorability and image size influencing memorability) may seem contradictory, they actually show that on top of inherent consistency, image size also shifts memorability score range such that images of lower memorability in one size change their absolute memorability score but remain with relatively lower memorability in the other size, and images with higher memorability in one size change their absolute memorability score but remain with relatively higher memorability in the other size. This is in line with our earlier findings for image size affecting memorability (Masarwa, Kreichman and Gilaie-Dotan, 2022) and is also evident for other physical visual aspects as image contrast (Brook *et al.*, 2024). These results highlight the contribution of physical visual properties to visual memorability in addition to other aspects (as semantic category (Khosla *et al.*, 2015; Kramer *et al.*, 2023)).

## Limitations of the study

The results obtained here of image size affecting memory during naturalistic encoding with young adults (18-35y) as evident by multiple memory measurements are robust but may not generalize to task-based paradigms, to mobile phones or tablets that may impose different constraints on the eyes (e.g., Jaiswal *et al.*, 2019), to potential different setups, or to other age groups. Our results could be attributed to multiple factors that vary with image size as visual system resources, differences in eye scanning patterns, attention, or even different composition of image spatial frequencies. Furthermore, in our study we used images from four main visual categories (with faces or people, and indoor or outdoor images without people). Different visual categories are supported by different high-level visual neural substrates (Gilaie-Dotan, Gelbard-Sagiv and Malach, 2010; Schwarzkopf *et al.*, 2010; Silvano *et al.*, 2010; Gilaie-Dotan *et al.*, 2012; Kreichman and Gilaie-Dotan, 2024) and therefore memory of images from visual categories not tested here may actually be different. In addition, since all the images from all the four categories in a block were randomly ordered, between-category and between-image influences may have occurred (such as for example large face followed by a large outdoor image may feel less natural than a large indoor scene followed by a large outdoor scene). While such potential interactions are likely to have been comparable across image size conditions and also canceled out across participants, it would be interesting to take these world-based relations into account and test them in future studies.

## Conclusions

Image size plays a role in multiple everyday activities such as work, medical communications with practitioners, education or entertainment, and therefore it may possibly influence perception, comprehension, learning and memory. The current results, based on naturalistic encoding design,

support the idea that physical visual properties as retinal size affect visual memory and do not merely reflect shifts in response biases. Our results further substantiate the importance of examining an effect with multiple analytical approaches to provide converging evidence for that effect.

### **Acknowledgements**

This study was funded by ISF individual grants 1485/18 and 1462/23 to SGD and by the Lev Zion Council for Higher Education PhD Scholarship to OK. We thank Yaffa Yeshurun for fruitful suggestions.

### **Author contributions**

SGD conceived the study. OK, SM, and LB performed research and analyzed data. SGD and OK wrote the paper. All authors approved the final version of the manuscript.

### **Conflicts of interest**

The authors declare no competing interests.

### **Open practices statement**

The data generated and analyzed during the current study are available in the OSF repository at <https://osf.io/m59zf/>.

## References

- Akselevich, V. and Gilaie-Dotan, S. (2022) 'Positive and negative facial valence perception are modulated differently by eccentricity in the parafovea', *Scientific Reports*, 12(1), p. 21693. Available at: <https://doi.org/10.1038/s41598-022-24919-7>.
- Bainbridge, W.A. (2017) 'The memorability of people: Intrinsic memorability across transformations of a person's face.', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(5), pp. 706–716. Available at: <https://doi.org/10.1037/xlm0000339>.
- Bilton, N. (2011) *I Live in the Future & Here's How It Works: Why Your World, Work & Brain Are Being Creatively Disrupted*. Currency.
- Bönstrup, M. et al. (2020) 'Mechanisms of offline motor learning at a microscale of seconds in large-scale crowdsourced data', *npj Science of Learning*, 5(1), p. 7. Available at: <https://doi.org/10.1038/s41539-020-0066-9>.
- Bower, G.H. and Karlin, M.B. (1974) 'Depth of processing pictures of faces and recognition memory', *Journal of Experimental Psychology*, 103(4), pp. 751–757. Available at: <https://doi.org/10.1037/h0037190>.
- Brady, T.F. et al. (2008) 'Visual long-term memory has a massive storage capacity for object details', *Proceedings of the National Academy of Sciences*, 105(38), pp. 14325–14329. Available at: <https://doi.org/10.1073/pnas.0803390105>.
- Brady, T.F. et al. (2023) 'Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research', *Psychonomic Bulletin & Review*, 30(2), pp. 421–449. Available at: <https://doi.org/10.3758/s13423-022-02179-w>.
- Brady, T.F. and Alvarez, G.A. (2015) 'No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects.', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), pp. 921–929. Available at: <https://doi.org/10.1037/xlm0000075>.
- Brook, L. et al. (2024) 'Higher-contrast images are better remembered during naturalistic encoding', *Scientific Reports*, 14(1), p. 13445. Available at: <https://doi.org/10.1038/s41598-024-63953-5>.
- Brown, H.R. et al. (2014) 'Crowdsourcing for Cognitive Science – The Utility of Smartphones', *PLoS ONE*. Edited by A. D'Ausilio, 9(7), p. e100662. Available at: <https://doi.org/10.1371/journal.pone.0100662>.
- Craik, F.I.M. and Lockhart, R.S. (1972) 'Levels of processing: A framework for memory research', *Journal of Verbal Learning and Verbal Behavior*, 11(6), pp. 671–684. Available at: [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X).
- D'Agostino, P.R., O'Neill, B.J. and Paivio, A. (1977) 'Memory for pictures and words as a function of level of processing: Depth or dual coding?', *Memory & Cognition*, 5(2), pp. 252–256. Available at: <https://doi.org/10.3758/BF03197370>.

DisplaySize (no date) *Display size, Display size*. Available at: [https://en.wikipedia.org/wiki/Display\\_size](https://en.wikipedia.org/wiki/Display_size) (Accessed: 9 February 2023).

Felleman, D.J. and Van Essen, D.C. (1991) 'Distributed Hierarchical Processing in the Primate Cerebral Cortex', *Cerebral Cortex*, 1(1), pp. 1–47. Available at: <https://doi.org/10.1093/cercor/1.1.1>.

Fiser, J. and Biederman, I. (2001) 'Invariance of long-term visual priming to scale, reflection, translation, and hemisphere', *Vision Research*, 41(2), pp. 221–234. Available at: [https://doi.org/10.1016/S0042-6989\(00\)00234-0](https://doi.org/10.1016/S0042-6989(00)00234-0).

Gallant, J. and Libben, G. (2019) 'No lab, no problem: Designing lexical comprehension and production experiments using PsychoPy3', *The Mental Lexicon*, 14(1), pp. 152–168. Available at: <https://doi.org/10.1075/ml.00002.gal>.

Garrido, L. *et al.* (2009) 'Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics', *Brain*, 132(12), pp. 3443–3455. Available at: <https://doi.org/10.1093/brain/awp271>.

Gilaie-Dotan, S. *et al.* (2012) 'Neuroanatomical correlates of visual car expertise', *NeuroImage*, 62(1), pp. 147–153. Available at: <https://doi.org/10.1016/j.neuroimage.2012.05.017>.

Gilaie-Dotan, S. *et al.* (2015) 'Ventral aspect of the visual form pathway is not critical for the perception of biological motion', *Proceedings of the National Academy of Sciences*, 112(4), pp. E361–E370. Available at: <https://doi.org/10.1073/pnas.1414974112>.

Gilaie-Dotan, S. *et al.* (2019) 'Impairment in facial expression perception but normal biological motion perception in a patient with a lesion to right posterior STS', *Journal of Vision*, 19(10), p. 22a. Available at: <https://doi.org/10.1167/19.10.22a>.

Gilaie-Dotan, S., Gelbard-Sagiv, H. and Malach, R. (2010) 'Perceptual shape sensitivity to upright and inverted faces is reflected in neuronal adaptation', *NeuroImage*, 50(2), pp. 383–395. Available at: <https://doi.org/10.1016/j.neuroimage.2009.12.077>.

Gilaie-Dotan, S., Nir, Y. and Malach, R. (2008) 'Regionally-specific adaptation dynamics in human object areas', *NeuroImage*, 39(4), pp. 1926–1937. Available at: <https://doi.org/10.1016/j.neuroimage.2007.10.010>.

Golarai, G. *et al.* (2007) 'Differential development of high-level visual cortex correlates with category-specific recognition memory', *Nature Neuroscience*, 10(4), pp. 512–522. Available at: <https://doi.org/10.1038/nn1865>.

Gomez, J. *et al.* (2015) 'Functionally Defined White Matter Reveals Segregated Pathways in Human Ventral Temporal Cortex Associated with Category-Specific Processing', *Neuron*, 85(1), pp. 216–227. Available at: <https://doi.org/10.1016/j.neuron.2014.12.027>.

Grill-Spector, K. and Malach, R. (2004) 'The human visual cortex', *Annual Review of Neuroscience*, 27(1), pp. 649–677. Available at: <https://doi.org/10.1146/annurev.neuro.27.070203.144220>.

Hillyard, S.A., Vogel, E.K. and Luck, S.J. (1998) 'Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence', *Philosophical Transactions of the*

*Royal Society of London. Series B: Biological Sciences*. Edited by G.W. Humphreys, J. Duncan, and A. Treisman, 353(1373), pp. 1257–1270. Available at: <https://doi.org/10.1098/rstb.1998.0281>.

Huk, A.C. and Heeger, D.J. (2000) ‘Task-Related Modulation of Visual Cortex’, *Journal of Neurophysiology*, 83(6), pp. 3525–3536. Available at: <https://doi.org/10.1152/jn.2000.83.6.3525>.

Ioannidis, J.P.A. (2005) ‘Why Most Published Research Findings Are False’, *PLoS Medicine*, 2(8), p. e124. Available at: <https://doi.org/10.1371/journal.pmed.0020124>.

Isola, P. et al. (2014) ‘What makes a photograph memorable?’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), pp. 1469–1482. Available at: <https://doi.org/10.1109/TPAMI.2013.200>.

Jaiswal, S. et al. (2019) ‘Ocular and visual discomfort associated with smartphones, tablets and computers: what we do and do not know’, *Clinical and Experimental Optometry*, 102(5), pp. 463–477. Available at: <https://doi.org/10.1111/cxo.12851>.

Jaschinski, W. (2002) ‘The Proximity-Fixation-Disparity Curve and the Preferred Viewing Distance at a Visual Display as an Indicator of Near Vision Fatigue’, *Optometry and Vision Science*, 79(3), pp. 158–169. Available at: <https://doi.org/10.1097/00006324-200203000-00010>.

Kardosh, R. et al. (2022) ‘Minority salience and the overestimation of individuals from minority groups in perception and memory’, *Proceedings of the National Academy of Sciences*, 119(12), p. e2116884119. Available at: <https://doi.org/10.1073/pnas.2116884119>.

Khosla, A. et al. (2015) ‘Understanding and Predicting Image Memorability at a Large Scale’, in *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, pp. 2390–2398. Available at: <https://doi.org/10.1109/ICCV.2015.275>.

Konkle, T. et al. (2010a) ‘Conceptual distinctiveness supports detailed visual long-term memory for real-world objects.’, *Journal of Experimental Psychology: General*, 139(3), pp. 558–578. Available at: <https://doi.org/10.1037/a0019165>.

Konkle, T. et al. (2010b) ‘Scene Memory Is More Detailed Than You Think’, *Psychological Science*, 21(11), pp. 1551–1556. Available at: <https://doi.org/10.1177/0956797610385359>.

Kramer, M.A. et al. (2023) ‘The features underlying the memorability of objects’, *Science Advances*, 9(17). Available at: <https://doi.org/10.1126/sciadv.add2981>.

Kreichman, O., Bonneh, Y.S. and Gilaie-Dotan, S. (2020) ‘Investigating face and house discrimination at foveal to parafoveal locations reveals category-specific characteristics’, *Scientific Reports*, 10(1), p. 8306. Available at: <https://doi.org/10.1038/s41598-020-65239-y>.

Kreichman, O. and Gilaie-Dotan, S. (2024) ‘Parafoveal vision reveals qualitative differences between fusiform face area and parahippocampal place area’, *Human Brain Mapping*, 45(3). Available at: <https://doi.org/10.1002/hbm.26616>.

Levy, I. et al. (2001) ‘Center–periphery organization of human object areas’, *Nature Neuroscience*, 4(5), pp. 533–539. Available at: <https://doi.org/10.1038/87490>.

Li, X., Bainbridge, W.A. and Bakkour, A. (2022) ‘Item memorability has no influence on value-based decisions’, *Scientific Reports*, 12(1), p. 22056. Available at: <https://doi.org/10.1038/s41598-022-26333-5>.

Marr, David. (2010) *Vision : a computational investigation into the human representation and processing of visual information*. MIT Press.

Masarwa, S., Kreichman, O. and Gilaie-Dotan, S. (2022) 'Larger images are better remembered during naturalistic encoding', *Proceedings of the National Academy of Sciences*, 119(4), p. e2119614119. Available at: <https://doi.org/10.1073/pnas.2119614119>.

McGill, R., Tukey, J.W. and Larsen, W.A. (1978) 'Variations of Box Plots', in *The American Statistician*, p. 12. Available at: <https://doi.org/10.2307/2683468>.

Peirce, J. et al. (2019) 'PsychoPy2: Experiments in behavior made easy', *Behavior Research Methods*, 51(1), pp. 195–203. Available at: <https://doi.org/10.3758/s13428-018-01193-y>.

Rempel, D. et al. (2007) 'The Effects of Visual Display Distance on Eye Accommodation, Head Posture, and Vision and Neck Symptoms', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(5), pp. 830–838. Available at: <https://doi.org/10.1518/001872007X230208>.

RStudio Team (2020) 'RStudio: Integrated Development for R'. Boston, MA: RStudio, PBC, Boston, MA.

Saalman, Y.B. and Kastner, S. (2009) 'Gain control in the visual thalamus during perception and cognition', *Current Opinion in Neurobiology*, 19(4), pp. 408–414. Available at: <https://doi.org/10.1016/j.conb.2009.05.007>.

Sahar, T. and Yeshurun, Y. (2024) 'Temporal crowding with central vision reveals the fragility of visual representations.', *Journal of Experimental Psychology: General*, 153(2), pp. 339–351. Available at: <https://doi.org/10.1037/xge0001496>.

Sato, W. and Yoshikawa, S. (2013) 'Recognition Memory for Faces and Scenes', *The Journal of General Psychology*, 140(1), pp. 1–15. Available at: <https://doi.org/10.1080/00221309.2012.710275>.

Sauter, M., Draschkow, D. and Mack, W. (2020) 'Building, Hosting and Recruiting: A Brief Introduction to Running Behavioral Experiments Online', *Brain Sciences*, 10(4), p. 251. Available at: <https://doi.org/10.3390/brainsci10040251>.

Schneider, K.A., Richter, M.C. and Kastner, S. (2004) 'Retinotopic Organization and Functional Subdivisions of the Human Lateral Geniculate Nucleus: A High-Resolution Functional Magnetic Resonance Imaging Study', *The Journal of Neuroscience*, 24(41), pp. 8975–8985. Available at: <https://doi.org/10.1523/JNEUROSCI.2413-04.2004>.

Schwarzkopf, D.S. et al. (2010) 'Investigating object representations during change detection in human extrastriate cortex', *European Journal of Neuroscience*, 32(10), pp. 1780–1787. Available at: <https://doi.org/10.1111/j.1460-9568.2010.07443.x>.

Silvanto, J. et al. (2010) 'Differing causal roles for lateral occipital cortex and occipital face area in invariant shape recognition', *European Journal of Neuroscience*, 32(1), pp. 165–171. Available at: <https://doi.org/10.1111/j.1460-9568.2010.07278.x>.

Standing, L. (1973) 'Learning 10000 pictures', *Quarterly Journal of Experimental Psychology*, 25(2), pp. 207–222. Available at: <https://doi.org/10.1080/14640747308400340>.

Tootell, R.B.H. *et al.* (1997) 'Functional Analysis of V3A and Related Areas in Human Visual Cortex', *The Journal of Neuroscience*, 17(18), pp. 7060–7078. Available at: <https://doi.org/10.1523/JNEUROSCI.17-18-07060.1997>.

Tukey, J. (1977) 'Exploratory Data Analysis', in *Exploratory Data Analysis*. London, UK: Pearson.

Wandell, B.A. (1995) *Foundations of vision*. Sunderland, MA, US: Sinauer Associates.

Weidemann, C.T. and Kahana, M.J. (2016) 'Assessing recognition memory using confidence ratings and response times', *Royal Society Open Science*, 3(4), p. 150670. Available at: <https://doi.org/10.1098/rsos.150670>.

Wolfe, J.M. and Kuzmova, Y.I. (2011) 'How many pixels make a memory? Picture memory for small pictures', *Psychonomic Bulletin & Review*, 18(3), pp. 469–475. Available at: <https://doi.org/10.3758/s13423-011-0075-z>.

## Tables

	Exp.	1.5°	3°	6°	12°	15.5°	21°	24°	New	Overall
Accuracy	Lab3°-21°		45.6 ± 2.4				61.0 ± 2.25		81.2 ± 2.31	
	Lab3°-24°		33.1 ± 2.25	42.9 ± 2.46	52.0 ± 2.91			56.6 ± 2.47	81.6 ± 1.68	
	3°-15.5°		42.46 ± 1.92	52.92 ± 1.84	56.93 ± 1.78	57.92 ± 1.71			74.95 ± 2.01	
	1.5°-12°	34.57 ± 4.07	39.22 ± 3.26	48.62 ± 3.06	52.76 ± 3.58				76.38 ± 3.48	
	3°-24°		41.25 ± 3.13	47.50 ± 3.06	58.84 ± 3.22			50.09 ± 3.18	73.50 ± 3.79	
Dprime	Lab3°-21°		0.91 ± 0.08				1.33 ± 0.10			1.12 [0.94, 1.29]
	Lab3°-24°		0.52 ± 0.72	0.81 ± 0.07	1.07 ± 0.88			1.21 ± 0.08		1.09 [0.95, 1.23]
	3°-15.5°		0.60 ± 0.06	0.90 ± 0.07	1.00 ± 0.07	1.04 ± 0.07				0.88 [0.75, 1.00]
	1.5°-12°	0.37 ± 0.10	0.54 ± 0.09	0.80 ± 0.13	0.94 ± 0.14					0.68 [0.48, 0.87]
	3°-24°		0.53 ± 0.09	0.70 ± 0.14	1.02 ± 0.17			0.78 ± 0.14		0.75 [0.50, 1.00]
Criterion	Lab3°-21°		0.58 ± 0.07				0.37 ± 0.06			0.47 [0.35, 0.59]
	Lab3°-24°		0.75 ± 0.06	0.60 ± 0.59	0.47 ± 0.06			0.40 ± 0.06		0.45 [0.37, 0.55]
	3°-15.5°		0.51 ± 0.06	0.36 ± 0.05	0.31 ± 0.05	0.29 ± 0.05				0.37 [0.27, 0.47]
	1.5°-12°	0.67 ± 0.12	0.58 ± 0.11	0.45 ± 0.09	0.39 ± 0.10					0.52 [0.32, 0.70]
	3°-24°		0.51 ± 0.11	0.42 ± 0.09	0.26 ± 0.08			0.38 ± 0.09		0.40 [0.22, 0.57]
AUC	3°-15.5°									0.64 [0.62, 0.66]
	1.5°-12°									0.60 [0.57, 0.64]
	3°-24°									0.62 [0.58, 0.66]

RT (ms)	3°-15.5°		882 [797.08, 1029.90]	873.15 [798.97, 1022.28]	872 [777.65, 1007.34]	860 [793.6 3, 1023.6 3]			882.31 [799.88, 1023.93]	
	1.5°-12°	864.85 [821.95, 985.25]	866.5 [805.65, 1026.95]	860 [826.2, 935.45]	861.4 [824.65, 1012.9]				861.95 [809.7, 994.55]	
	3°-24°		893.44 [816.56, 1014.83]	873.20 [804.85, 994.21]	879.16 [822.29, 968.05]			877.13 [817.91, 982.30]	904.99 [831.73, 999.31]	

**Table 1. Test phase summary statistics by experiment, condition and measurement type.** Experimental condition (columns) for each measurement type with each row representing data from one of the experiments (from top to bottom: memory accuracy, memory sensitivity (dprime), decision criterion, AUC, and RTs). Gray cells indicate non-available data. Accuracy (% correct), dprime and criterion by mean  $\pm$  SE, RTs by median RT [Q1, Q3] (in ms). Overall column for dprime, criterion and AUC represents mean and 95% confidence interval range. Lab3°-21° (n=50), Lab3°-24° (n=51), 1.5°-12° experiment (n = 29), 3°-15.5° experiment (n = 119), and 3°-24° experiment (n = 28). See Methods for more details.

Exp.		Factors	Main effect	Post-hoc (Bonferroni/Dunn)
Lab 3°-24° (n = 51)	Accuracy	Size 3°, 6° 6°, 12° 12°, 24°	$F(3,150) = 57.31$ $p = 9.59*10^{-25}$ $\eta_G^2 = .202$	$p = 4.1*10^{-5}$ $p = 0.0001$ $p = 0.068$
	Dprime	Size 3°, 6° 6°, 12° 12°, 24°	$F(3,150) = 55.85$ $p = 2.68*10^{-24}$ $\eta_G^2 = .184$	$p = 4.4*10^{-5}$ $p = 0.0002$ $p = 0.055$
	Criterion	Size 3°, 6° 6°, 12° 12°, 24°	$F(3,150) = 55.85$ $p = 2.68*10^{-24}$ $\eta_G^2 = .086$	$p = 4.4*10^{-5}$ $p = 0.0002$ $p = 0.055$
3°-15.5° (n = 119)	Accuracy	Size 3°, 6° 6°, 12° 12°, 15.5°	$F(2.73,322.02) = 46.80$ $p = 1.98*10^{-23}$ $\eta_G^2 = .088$	$p = 2.3*10^{-10}$ $p = 0.049$ $p = 1$
	Dprime	Size 3°, 6° 6°, 12° 12°, 15.5°	$F(2.72,320.92) = 44.79$ $p = 1.67*10^{-22}$ $\eta_G^2 = .052$	$p = 2.6*10^{-10}$ $p = 0.09$ $p = 1$
	Criterion	Size 3°, 6° 6°, 12° 12°, 15.5°	$F(2.72,320.92) = 44.79$ $p = 1.67*10^{-22}$ $\eta_G^2 = .022$	$p = 2.6*10^{-10}$ $p = 0.09$ $p = 1$
1.5°-12° (n = 29)	Accuracy	Size 1.5°, 3° 3°, 6° 6°, 12°	$F(3,84) = 11.20$ $p = 2.94*10^{-6}$ $\eta_G^2 = .132$	$p = 0.88$ $p = 0.023$ $p = 1$
	Dprime	Size 1.5°, 3° 3°, 6° 6°, 12°	$F(2.3,64.42) = 11.63$ $p = 2.1*10^{-5}$ $\eta_G^2 = .113$	$p = 0.62$ $p = 0.024$ $p = 0.74$
	Criterion	Size 1.5°, 3° 3°, 6° 6°, 12°	$F(2.3,64.42) = 11.63$ $p = 2.1*10^{-5}$ $\eta_G^2 = .037$	$p = 0.62$ $p = 0.024$ $p = 0.74$
3°-24° (n = 28)	Accuracy	Size 3°, 6° 6°, 12° 12°, 24°	$F(3,81) = 10.23$ $p = 8.7*10^{-6}$ $\eta_G^2 = .13$	$p = 0.28$ $p = 0.021$ $p = 0.037$
	Dprime	Size 3°, 6° 6°, 12° 12°, 24°	$F(3,81) = 10.50$ $p = 6.6*10^{-6}$ $\eta_G^2 = .057$	$p = 0.31$ $p = 0.015$ $p = 0.04$
	Criterion	Size 3°, 6° 6°, 12° 12°, 24°	$F(3,81) = 10.50$ $p = 6.6*10^{-6}$ $\eta_G^2 = .032$	$p = 0.31$ $p = 0.015$ $p = 0.04$

**Table 2. Repeated-measures ANOVA statistical analyses on test-phase results.** One-way repeated-measures ANOVA results by experiment (from top to bottom: Lab3°-24°, 3°-15.5°, 1.5°-12° and 3°-24°) for size effect on performance for accuracy, dprime and criterion. Bonferroni/Dunn post-hoc results in rightmost column. Values in bold represent significant results. Accuracy analyses based only on performance for the older images.

Exp.	Test	Variables	Results
<b>Lab3°-21°</b>	Paired, two tailed t-test	Accuracy of 21° vs 3°	$t(49) = 8.31, p = 6.4 \cdot 10^{-11}$
	Paired, two tailed t-test	Dprime of 21° vs 3°	$t(49) = 8.47, p = 3.6 \cdot 10^{-11}$
	Paired, two tailed t-test	Criterion of 21° vs 3°	$t(49) = -8.47, p = 3.6 \cdot 10^{-11}$
	One sample, two tailed t-test	Overall dprime vs mean = 0	$t(49) = 12.59, p < 2.2 \cdot 10^{-16}$
	One sample, two tailed t-test	Dprime 3° vs mean = 0	$t(49) = 11.23, p = 3.7 \cdot 10^{-15}$
<b>Lab3°-24°</b>	One sample, two tailed t-test	Overall dprime vs mean = 0	$t(50) = 15.49, p < 2.2 \cdot 10^{-16}$
	One sample, two tailed t-test	Dprime 3° vs mean = 0	$t(50) = 7.24, p = 2.6 \cdot 10^{-9}$
<b>3°-15.5°</b>	Paired, two tailed t-test	Accuracy of 15.5° vs 3°	$t(118) = 10.22, p < 2.2 \cdot 10^{-16}$
	One sample, two tailed t-test	Overall dprime vs mean = 0	$t(118) = 14.01, p < 2.2 \cdot 10^{-16}$
	One sample, two tailed t-test	Dprime 3° vs mean = 0	$t(118) = 10.19, p < 2.2 \cdot 10^{-16}$
	One sample, two tailed t-test	AUC vs mean = 0.5	$t(118) = 13.90, p < 2.2 \cdot 10^{-16}$
	Non-directional Pearson correlation	AUC, overall dprime	$r = 0.922$ $t(117) = 25.8, p = 4.5 \cdot 10^{-50}$
	Friedman rank sum test	Effect of image size on reaction time	$\chi^2(3) = 3.38, p = 0.34$
<b>1.5°-12°</b>	One sample, two tailed t-test	Overall dprime vs mean = 0	$t(28) = 6.80, p = 2.2 \cdot 10^{-7}$
	One sample, two tailed t-test	Dprime 1.5° vs mean = 0	$t(28) = 3.59, p = 0.0012$
	One sample, two tailed t-test	Dprime 3° vs mean = 0	$t(28) = 6.05, p = 1.6 \cdot 10^{-6}$
	One sample, two tailed t-test	AUC vs mean = 0.5	$t(28) = 6.53, p = 4.4 \cdot 10^{-7}$
	Non-directional Pearson correlation	AUC, overall dprime	$r = 0.893$ $t(27) = 10.3, p = 7.8 \cdot 10^{-11}$
	Friedman rank sum test	Effect of image size on reaction time	$\chi^2(3) = 0.32, p = 0.95$
<b>3°-24°</b>	One sample, two tailed t-test	Overall dprime vs mean = 0	$t(27) = 5.90, p = 2.7 \cdot 10^{-6}$
	One sample, two tailed t-test	Dprime 3° vs mean = 0	$t(27) = 5.72, p = 4.4 \cdot 10^{-6}$
	One sample, two tailed t-test	AUC vs mean = 0.5	$t(27) = 6.10, p = 1.6 \cdot 10^{-6}$
	Non-directional Pearson correlation	AUC, overall dprime	$r = 0.89$ $t(26) = 9.9, p = 2.4 \cdot 10^{-10}$
	Friedman rank sum test	Effect of image size on reaction time	$\chi^2(3) = 2.23, p = 0.53$

**Table 3. Additional statistical analyses on test-phase results.** Results by experiment (on the left) detailed by the statistical test (2<sup>nd</sup> column from left), the relevant test variables (3<sup>rd</sup> column from left) and the results (rightmost column). Values in bold represent significant results.