BRIEF REPORT

Eye-movements reveal semantic interference effects during the encoding of naturalistic scenes in long-term memory

Anastasiia Mikhailova[1,2] · Ana Raposo[1] · Sergio Della Sala[3] · Moreno I. Coco[1,4]

Abstract

Similarity-based semantic interference (SI) hinders memory recognition. Within long-term visual memory paradigms, the more scenes (or objects) from the same semantic category are viewed, the harder it is to recognize each individual instance. A growing body of evidence shows that overt attention is intimately linked to memory. However, it is yet to be understood whether SI mediates overt attention during scene encoding, and so explain its detrimental impact on recognition memory. In the current experiment, participants watched 372 photographs belonging to different semantic categories (e.g., a kitchen) with different frequency (4, 20, 40 or 60 images), while being eye-tracked. After 10 minutes, they were presented with the same 372 photographs plus 372 new photographs and asked whether they recognized (or not) each photo (i.e., old/new paradigm). We found that the more the SI, the poorer the recognition performance, especially for old scenes of which memory representations existed. Scenes more widely explored were better recognized, but for increasing SI, participants focused on more local regions of the scene in search for its potentially distinctive details. Attending to the centre of the display, or to scene regions rich in low-level saliency was detrimental to recognition accuracy, and as SI increased participants were more likely to rely on visual saliency. The complexity of maintaining faithful memory representations for increasing SI also manifested in longer fixation durations; in fact, a more successful encoding was also associated with shorter fixations. Our study highlights the interdependence between attention and memory during high-level processing of semantic information.

Keywords Long-term visual memory · Semantic interference · Eye movements · Visual saliency · Eye-tracking

✉ Moreno I. Coco
  M.Coco@uel.ac.uk; moreno.cocoi@gmail.com

  Anastasiia Mikhailova
  a.mikhailova@psicologia.ulisboa.pt

  Ana Raposo
  alraposo@psicologia.ulisboa.pt

  Sergio Della Sala
  sergio@ed.ac.uk

1. Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisbon, Portugal

2. Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal

3. Human Cognitive Neuroscience, Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

4. School of Psychology, University of East London, Water Lane, E15 4LZ, London, UK

When recalling the memory of a certain episode, other episodes sharing a similar context may interfere with it. For example, when trying to remember the specific image of a kitchen, memories of images from the same semantic category (i.e., other kitchens) may also get activated, and so interfere with the recognition of this exemplar. This cognitive phenomenon, identified for the first time by Müller and Pilzecker (1900), has been ever since at the heart of memory research (see M. T. Dewar et al., 2007, for a review) and attributed to either mental activities intervening between the encoding of a stimulus and its retrieval, such as comparing periods of wakeful rest versus cognitive engagement (Cowan et al., 2004; M. Dewar et al., 2012), or to response competition due to the content similarity of the stimuli that are memorized (Craig et al., 2013; Underwood, 1945), such as semantically related versus unrelated word lists (Baddeley & Dale, 1966; McGeoch & McDonald, 1931; see Ishiguro & Saito, 2020, for a recent review).

Memory interference has historically been investigated using verbal recall or picture-word associations tasks (Dale, 1964; Rosinski et al., 1975; Shulman, 1971). In recent years,

Ⓐ Springer

interest grew around the impact of similarity-based semantic interference on long-term memory for visual information, which also constitutes the focus of the present research.

In a series of studies, Konkle et al. (2010a, 2010b) demonstrated that the fidelity of memory representations for arrays of standalone objects, or naturalistic scenes, critically depends on the semantic interference occurring between stimuli that have been encoded in memory: an increase in the frequency of scenes (or objects) per semantic category was associated with a systematic decrement in the recognition of each individual exemplar encoded in that category.

Beside their semantic content, visual images also convey low-level information (e.g., colour or luminosity), which can be computationally quantified in a synthetic measure known as visual saliency (see Itti et al., 1998, for a well-known model). When looking at the impact of low-level information on memorability, however, research seems to indicate no significant correlations between the two (see Isola et al., 2014, for natural scenes, Dubey et al., 2015, for objects, and refer to Bainbridge, 2019, for a review on the topic of memorability). Instead, visual images own an intrinsic memorability strength that is independent of their high-level or low-level characteristics or to the type of tasks and the depth of cognitive processing involved (Bainbridge, 2020). The memorability of visual information also relies on patterns of extrinsic responses (e.g., eye-movements, that participants generate when encoding such information in memory; e.g., Bylinskii et al., 2015; and see Hannula, 2018, for a review of the topic). For example, a higher number of fixations, or smaller pupil dilations while scenes are studied in preparation for a recognition test are associated with a better memory performance (Kafkas & Montaldi, 2011).

Global measures of exploration as obtained from attention maps (Pomplun et al., 1996), in which all fixations on a given image are portrayed along its two-dimensions, are an important predictor of its memorability, too. In this context, the more spread out fixations were on an image during encoding, which implied that several regions were attended to, the better this image was later recalled (e.g., Damiano & Walther, 2019; and Lyu et al., 2020, for another application of attention maps in the context of image memorability).

The duration of individual fixations can also express ongoing memory processes. Meghanathan et al. (2015), for example, showed that fixation duration linearly increases as the number of target distractors present in the context also increases in a change detection task, or Loftus et al. (1992), where increasing fixation durations were associated with the amount of degradation in low-level features of images to be later remembered.

Another important aspect of oculo-motor control, which has been relatively neglected in the context of memory processes, is the tendency of observers to re-orient their overt attention towards the centre of the display during scene viewing (i.e., centre bias; Tatler, 2007). To the best of our knowledge, Lyu et al. (2020) is the only study that has examined the role of centre bias on memory recognition. A centre proximity map was computed to weight low-level saliency maps generated with the Graph-Based Visual Saliency (GBVS) algorithm (Harel et al., 2006) and a single value, representing the probability of salient regions to be positioned in the centre of the display, generated. Their result did not show any significant relationship between central bias and memorability, which seems to confirm the marginal role played by low-level visual features on scene memorability.

Attending and memorizing are indeed closely coupled, but high-level semantic mechanisms of interference may influence overt attention as memories get formed. If this supposition is true, we should be able to bridge the expected decrement in recognition memory onto eye-movement responses. Our proposition is that as the fidelity of individual memory representations (e.g., the specific image of a kitchen) degrades under the influence of semantic interference (i.e., a memorized pool of kitchens), oculo-motor compensatory strategies are adopted to cope with the increased complexity of discriminating the memory of each individual instance from a pool of semantically overlapping competitor instances.

Thus, the current study aims at demonstrating that semantic interference on long-term visual memory directly mediates overt attention at encoding of visual information. Most importantly, our goal is to gauge the oculo-motor dynamics that underlie the successful formation and later access of memory representations as they degrade due to semantic interference.

We manipulated semantic interference of naturalistic images following the procedure by Konkle et al. (2010b), but tested recognition memory on an old/new paradigm rather than a two-alternative forced choice (2AFC), which elicits recollection more than familiarity mechanisms (for a direct comparison of these two paradigms, see Bayley et al., 2008, and Cunningham et al., 2015). Eye-tracking was included in the procedure to examine oculo-motor patterns associated with the encoding of visual information in memory. Departing from previous work, we examined the impact of semantic interference as a continuous, rather than as a categorical, variable. This approach allowed us to estimate the incremental (trial-by-trial) impact of semantic interference on recognition accuracy and how this is accommodated by changes in eye-movement responses.

On recognition accuracy, we expect to replicate the semantic interference effect observed by Konkle, et al. (2010b), whereby the higher the interference of the semantic category a scene is encoded into, the worst it would be its future recognition. However, if this effect truly relates to memory representations, then it should more strongly manifest in old rather than new images. Moreover, even though images are intrinsically memorable (Bainbridge, 2020), in our paradigm, we expect their memorability to reduce under the influence of semantic interference, and so observe a lower inter-

participant correlation than Isola et al. (2014), where the semantic interference between images was not manipulated.

On the eye-movement data collected while scenes were viewed for the first time (i.e., at encoding), we focus on four complementary measures: (a) the amount of visual information that was attended to, by looking at the overall spread of fixations across the scene; (b) the attentional effort to acquire visual information from the scene, by looking at fixation duration; (c) the reliance of participants on low-level visual features of the scene, by looking at the correspondence between fixation positions and visual saliency at such locations; and (d) the tendency of participants to re-orient their overt attention towards the centre of the screen, by looking at the correspondence between fixation positions and a centre proximity map (see section Dependent Variables for formal definitions of these measures).

In line with Damiano and Walther (2019), a high spread of the fixation distribution across a scene during its encoding, which indicates that it was widely inspected, should reflect a later better recognition. However, as the exposure to scenes from the same category increases (i.e., semantic interference), the representational fidelity of each individual scene decreases, and so we expect participants to attend more local regions in search for its potentially distinctive features. This suggestion would theoretically corroborate that the repeated exposure to the same visual scene is associated with a systematic reduction in the number of regions explored (see Althoff & Cohen, 1999, and Ryan et al., 2000, for an example using naturalistic scenes). This strategy may support the successful encoding of an image up to a certain level of semantic interference though. As fixation entropy is expected to drop due to semantic interference, it may reach the same level for scenes that will and scenes that will not be later correctly recognized, and so lose discriminative power.

Semantic interference degrades the representational fidelity of individual instances by reinforcing their categorical overlap, and so we expect fixation duration to significantly increase to keep instances discriminable as a response. This prediction will conceptually align with the study by Ryan et al. (2007), showing that the repeated exposure to familiar faces resulted into progressively longer fixation durations, and connect with Loftus et al. (1992), showing that degradation, albeit in the perceptual domain, was associated with an increase in fixation duration in a long-term visual memory task. Moreover, as fixation duration is an index of processing effort (see Coco et al., 2020, for an example in the context of object-scene semantic integration), we expect it to be negatively associated with recognition accuracy, whereby the longer the average fixation duration is, the less likely the scene was efficiently encoded into memory.

Moreover, if participants indeed search for potentially diagnostic features in scenes as semantic interference increases, they would rely more on low-level visual features of the scene (i.e., a higher correspondence between fixation position and low-level visual saliency). Building upon Isola et al. (2014), however, we do not expect low-level image features to significantly contribute to whether the scene will be later correctly recognized, or not.

Finally, an increased tendency to re-centre gaze during the encoding of the scene may indicate that it was not exhaustively explored, and so, along with our prediction about fixation entropy (Damiano & Walther, 2019), we would expect it to be negatively associated with memory recognition (i.e., the greater the centre bias the poorer memory recognition). This may especially be true if fixation responses and centre-bias are analyzed in tandem. Instead, if centre-bias is considered as independent from eye-movement responses, in line with Lyu et al. (2020), we would instead predict a lack of its association with recognition memory.

Method

Participants

Twenty-five native English speakers (17 females, age = 21.95 ± 3.47 SD, range: 20–36 years) with normal or corrected to normal vision took voluntarily part in the study. Participants' sample size and number of trials were based on Konkle et al. (2010b), as our aim was to replicate as close as possible the original design, even if with a different memory paradigm, and so draw sounded comparisons between the original study and our results.[1] As the eye-movement data of two participants were not correctly acquired, they were excluded from these analyses (i.e., $N = 23$) and kept in for the analyses of their manual recognition responses. The Psychology and Research Ethics Committee of the University of Edinburgh approved the study before data collection, and all participants gave their written consent at the start of the experimental session.

Material and apparatus

We selected 1,488 naturalistic images from SUN database (Xiao et al., 2010) with a minimum of 550 × 550 pixels resolution and which did not include animate objects like humans or animals. All images were cropped and rescaled to 800 × 800 pixels to collect finer-grained eye-movement responses, and were equally drawn from one of 12 categories, six human-made environments (i.e., amusement park, bathroom, gas station, highway, kitchen, library) and six natural environments (i.e., beach, desert, field, forest, mountain, river). Miniatures

---

[1] The power and p value in 100,000 simulated experiments based on the same number of conditions and participants of our study show that it is possible to detect a significant effect with a power above .3 assuming a p value <.05, which should minimize the chance of incurring into Type 2 errors.
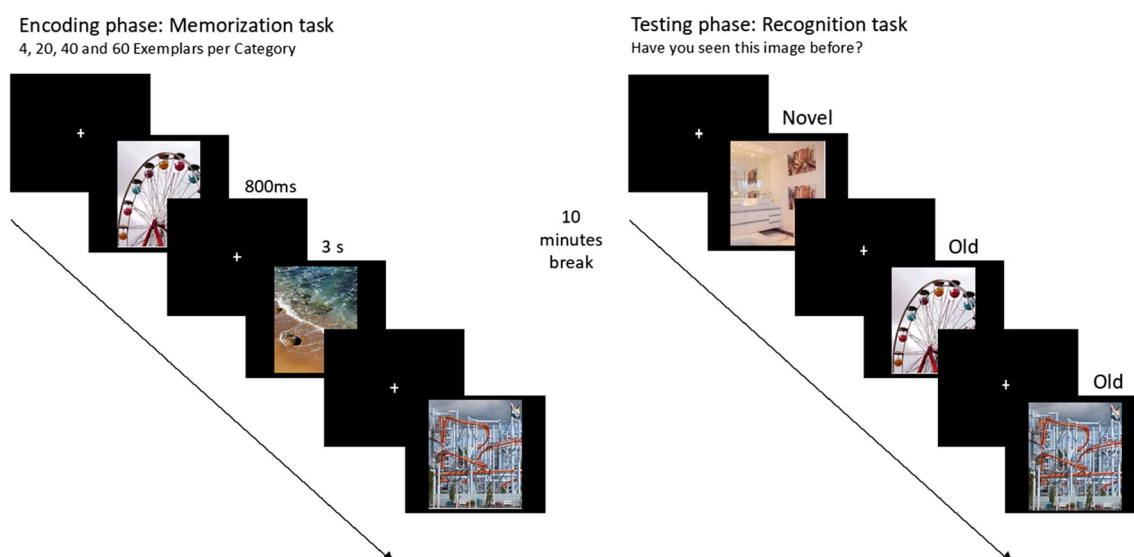
Fig. 1 Visualisation of the experimental design, procedure, and example images used in this study

of all scenes by category are reported in Appendices (*Visualization of the miniatures of all scenes used in this study organized by semantic category*) and with a greater resolution in the Supplementary Materials[2] (S1).

Images (800 × 800 pixels) were centrally presented on a black background at their resolution[3] with a 19-in. Dell Monitor (16.2 inch × 7.2 inch) screen resolution of 1,920 × 1,024 and set at a viewing distance of ~60 cm. Eye-movement data was recorded binocularly using a Gazepoint GP3 HD eye-tracker, sampling at 150 Hz. The experiment was built on OpenSesame 3.1.9 (Mathôt et al., 2012) and the acquisition of eye-tracking data made possible through the PyGaze Python plug-in (Dalmaijer et al., 2014). Each participant was calibrated on 9 points, and recalibrated if necessary. The mean degree of visual angle deviation accepted for the calibration was 0.37 degrees on the *x*-axis (*SD* = 0.15) and 0.53 degrees on the *y*-axis (*SD* = 0.29).

Procedure

Participants were administered the WASI-II test of intelligence[4] (Wechsler, 1999) at the start of each session (Full Scale = 117.79 ± 10.25; 97–140) and then completed a long-term visual memory task, which assessed their recognition accuracy using an old/new approach (see Fig. 1 for a visualization of the experimental design). Each participant watched a stream of 372 images, each presented for 3 seconds with 800-ms fixation crosshair intertrial, during encoding. After a short 10 minutes break, the participant was tested on
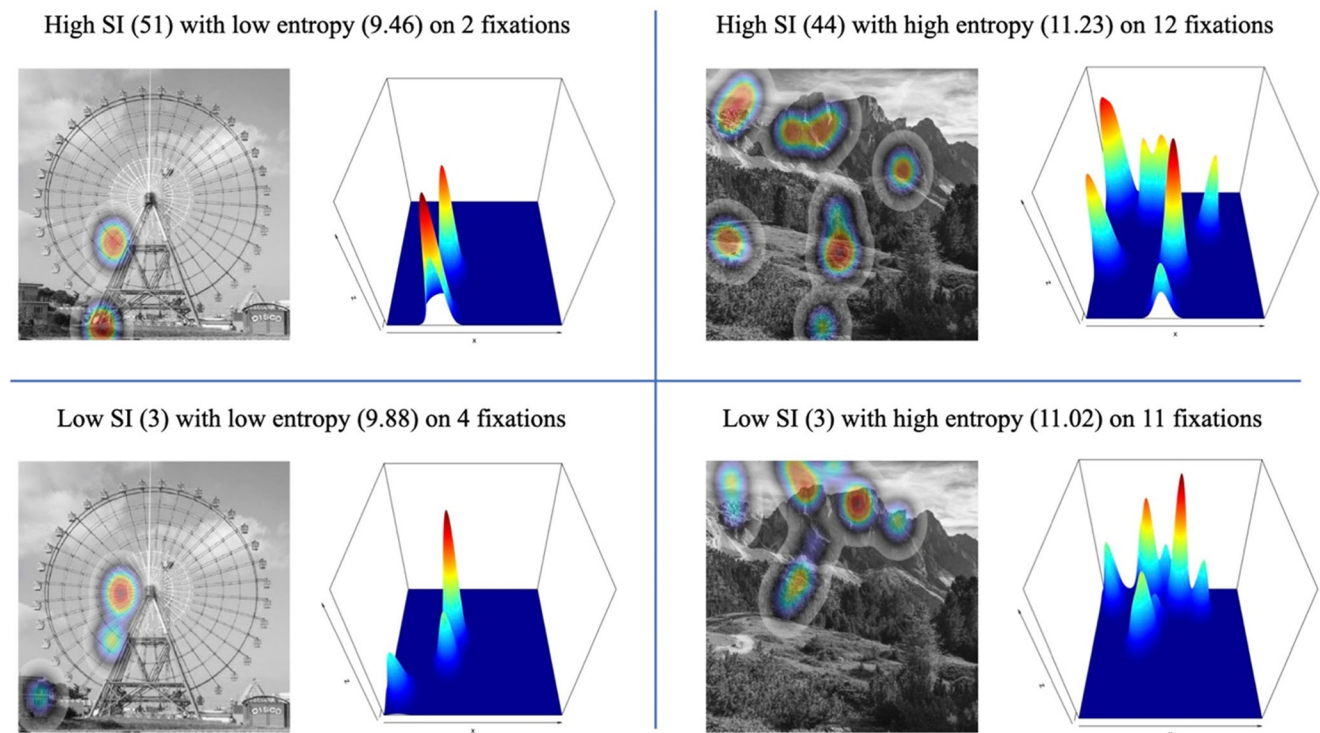
744 images, presented one by one. Half of these scenes were the 372 images seen during the encoding phase and the remaining 372 were novel scenes. They were asked to indicate whether they remembered or not the image using the keyboard (> yes; < no). The image was visible until a recognition response was made. To implement the semantic interference manipulation, we varied the frequency of images of each semantic category that participants were exposed to. In line with Konkle et al. (2010b), we varied SI in four levels (4, 20, 40, or 60 images per category). The total of 1,488 scenes was obtained by selecting 124 unique scenes for each of the 12 different categories (124 × 12), which is the number of images needed to cover all four levels of semantic interference, distributed across eight randomization lists. Each level of interference was distributed in each list to three different categories (4 × 3) and all four levels of interference were counterbalanced across semantic categories by rotating such levels onto four different lists (e.g., if the kitchen category had an SI of 4 in List 1, the same category had an SI of 20 in List 2, an SI of 40 in List 3, and of 60 in List 4). Four additional lists were created by swapping old with novel scenes between the encoding and the recognition phase to ensure that all scenes were seen in both conditions. Images from each semantic category and level of interference were randomly assigned to the lists making sure that they were never repeated within each list. Images in both phases of encoding and recognition were presented in randomized order—that is, we did not block images sharing the same semantic category to appear contiguously, and equal frequency per semantic category (e.g., for 20 kitchen images in the encoding phase, there were 20 old and 20 novel kitchen images in the testing phase). The experimental session took approximately 2 hours.

---

[2] The full stimuli data set will be made available upon request.
[3] Images were not scaled to fit the display dimensions.
[4] WASI-II was administered as pedagogical training for the undergraduate students who helped us with the data collection. As this test did not show any significant link with the long-term visual memory study, we only reported the full-scale score for completeness.

High SI (51) with low entropy (9.46) on 2 fixations

High SI (44) with high entropy (11.23) on 12 fixations

Low SI (3) with low entropy (9.88) on 4 fixations

Low SI (3) with high entropy (11.02) on 11 fixations

Fig. 2 Examples of attention maps with a high and low fixation entropy (left column, right column) when images were encoded at a high or a low level of semantic interference (top-row, bottom row). On each panel, we present the attention map as an heatmap (left) and as a 3D landscape to better visualize how Gaussians were fit to fixation position and their height scaled by fixation duration. In bracket, we report the fixation entropy obtained from each attention map

Analyses

Data exclusion

Of the 18,600 recognition trials (25 participants × 744 recognition trials), we excluded 393 trials (2.11%) with a response time either faster than 1% or slower than 99% of all trials as separately computed independently per participant. The number of recognition trials analyzed was 18,207 (an average of 728.28 ± 1.4 per participant). On 8,556 encoding trials (23 participants × 372 encoding trials), we excluded 255 trials (2.98%) because most fixations were out of range (i.e., bad eye-tracking), and a further 389 trials (4.54%) which had an average fixation duration (164), total number of fixations (48), or a subsequent recognition response time[5] (177) below 1% or above 99% of their respective distributions. Thus, the number of encoding trials analyzed was 7,942 trials (an average of 345.3 ± 28.17 per participant).

Independent variables

The key independent variable of this study is *SI*, which was manipulated in the design as frequency of scenes belonging the same semantic category (i.e., 4, 20, 40, 60), and incrementally administered to participants during the entire experimental session (i.e., from 1 to 60 during encoding and from 1 to 120 in the recognition phase, which were the maximum number of scenes belonging to the same category that could be seen in either phase). We treated SI as a continuous variable[6] to capture its incremental impact on recognition accuracy and oculo-motor responses on a trial-by-trial basis (refer to Supplementary Material S2 for a visualization of this measure) and standardized it into *z*-scores to minimize convergence issues (e.g., co-linearity) when it was introduced in the regression. *Image novelty* (old and new, set as reference level) was another independent variable that we included in the analyses of recognition accuracy to distinguish between hits and correct rejections. The last independent variable included in the analysis of eye-movement at encoding was the recognition *accuracy* of old scenes, which made possible to differentiate oculo-motor strategies that support effective (vs. ineffective) memory processes. This independent variable was

---

[5] Recognition accuracy was used as predictor in all eye-movements models, and so if the response time were unrealistic, then the associated response accuracy would also be unreliable, hence motivating the exclusion of these trials in these analyses.

[6] The post hoc analysis showed corroborating results when SI was introduced in the analysis as a categorical variable, although its effect became weaker.

also scaled prior to entering it in the regression to minimize co-linearity with the other predictors.

Dependent variables

Manual responses *Recognition accuracy* is a binomial variable that indicates whether a scene was correctly remembered as already seen (old) or rejected as novel (new). Following Konkle et al. (2010a), we also fit a general linear model (binomial link) predicting recognition accuracy as a function of SI independently for each participant, and separately for old and new trials. In this way, we derived the *interference slope* (i.e., the beta coefficient associated with SI), which reflects how much was the recognition accuracy of each participant impacted by increasing interference for scenes she/he already viewed (old) or had never seen (new). Negative coefficients indicate that recognition accuracy decreased when semantic interference increased. In Appendices (D-prime and criterion), we report additional analyses of *d*-prime and criterion to explore how was the signal (hit) discriminated from the noise (false-alarm) by the participants of our task, whether they adopted a conservative or a liberal strategy, and examined the impact of semantic interference on both. Finally, from the recognition accuracy we obtained the *intrinsic memorability* of our images. We used the method by Isola et al. (2014) and computed the Spearman correlation between the recognition accuracy of each individual scene (as hits) in two randomly split sets of participants, and iterated this procedure 50 times to avoid that findings may spuriously relate to a precise random selection of the participants' split.

Eye-movement responses As we were mainly interested in how the initial patterns of scene exploration related to memory formation, we only considered eye-movement data of the encoding phase in this analysis. Raw eye-movement sample were parsed into fixations and saccades using the I2MC algorithm by Hessels et al. (2017), implemented in MATLAB, which is suited to low-resolution data.

From fixation events, we computed four dependent measures: (a) the *average fixation duration* of all fixations in a trial to index processing effort, (b) the *entropy* of the spatial spread of fixations to get at global patterns of scene exploration, (c) the *Normalized Scanpath Saliency* (NSS; Peters et al., 2005) to tap into the attentional guidance provided by low-level visual features of the scene, and (d) the NSS score between a centre proximity map and fixation positions to examine the tendency of our participants to re-centre their gaze.

To compute (b), we first built a fixation probability map of each trial by placing at fixation coordinates, Gaussian kernels with a bandwidth set at 1 degree of visual angle (roughly 27 pixels) to approximate the size of the fovea. The height of the Gaussian was weighted by the proportion of time spent fixating at that location to better integrate differences in the amount of overt attention deployed across the scene. Then, the entropy of the resulting fixation map was calculated as $-\sum_{x, y} p(S_{x, y}) \log_2 p(S_{x, y})$, where $p(S_{x, y})$ is the normalized fixation probability at the coordinates of the fixation (x, y) in the scene S (see Castelhano et al., 2009; Coco & Keller, 2014; or Henderson, 2003, for related examples). Thus, the higher the fixation entropy, the more spread out fixations across the scene are. In Fig. 2, we visualize four example heatmaps of fixation distributions in low versus high entropy organized as columns and low vs. high interference organized as rows and report the value of fixation entropy for each map.

The NSS score (c) was instead obtained by first computing a visual saliency map of each scene using the Fast and Efficient Saliency model (FES; Tavakoli et al., 2011), where saliency is estimated from contrasts of local features (centre-surround) in a Bayesian framework and centre bias in eye movement responses taken into account by using an average fixation map. Then, saliency maps were normalized to have zero mean and unit in standard deviation, and saliency values at fixation positions of each trial extracted, and averaged to the NSS score (refer to Bylinskii et al., 2019, for the implementation we used). NSS is invariant for linear transformations and positive scores indicate above chance correspondence between fixation positions and visual saliency of the image.

Finally, for (d), we created a centre proximity matrix (800 × 800 pixels) by first calculating the Euclidian distance of each pixel with respect to the centre pixel, then normalizing this distance map to range between 0 and 1 and inverting it (see Hayes & Henderson, 2020). As a second step, we computed the NSS score between fixation positions and the centre proximity map for each scene and each participant. Note, this approach differs from Lyu et al. (2020) and Hayes and Henderson (2020), as we only used the centre proximity map to isolate the tendency of viewers to re-centre their gaze, independently of any other low-level features of the scene. See also Supplementary Material S4, replicating the approach by Lyu et al. (2020), and confirming: (1) a lack of association between centre bias and memorability when eye-movement responses are not taken into account but (2) a clear effect of visual saliency on memorability with, and without, weighting the saliency maps by centre bias when eye-movement responses are instead integrated in the analysis.

Inferential statistics

We used linear mixed-effects models (LMM) and generalized linear mixed-effects models (GLMM) as implemented in the lme4 package in R (Bates et al., 2015) to conduct the statistical analyses of our dependent measures. The fixed effects of our models (i.e., our independent variables) were introduced as main effects as well as in interaction. The random effects were participant (23) and scenes (1,488), which were nested into their respective categories (12) and introduced as intercepts. In the table of results, we reported the beta coefficients, *t* values (LMM), *z*
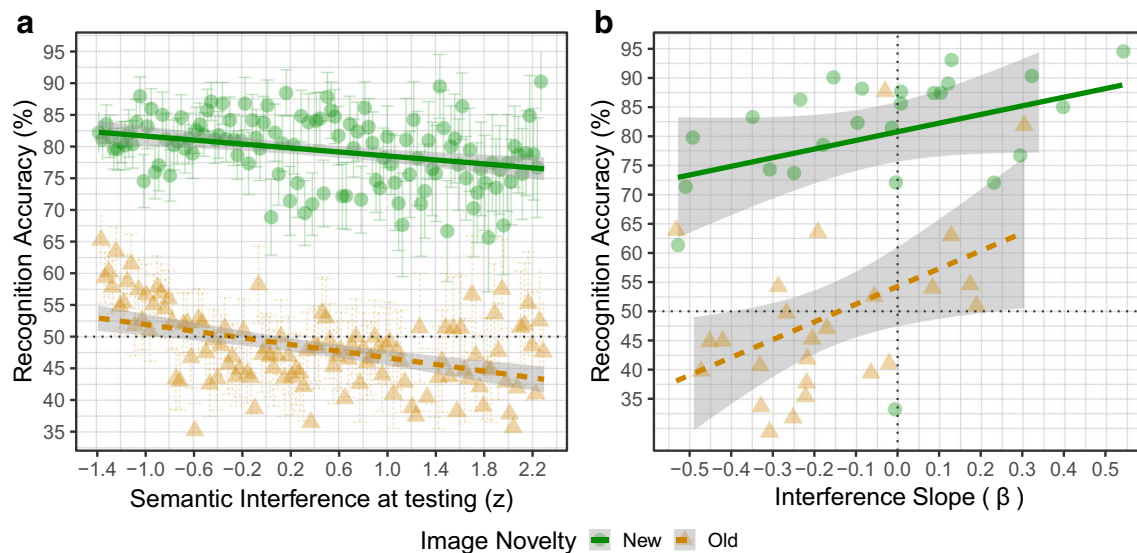
Fig. 3 Recognition accuracy. **a** Percentage recognition accuracy (*y*-axis) as a function of Semantic Interference (a continuous variable ranging from 2 to 120, *z*-scored) grouped by the image novelty (new scenes, green circles; old scenes, yellow triangles). Each individual point represents the average across participants and trials for that level of interference. Lines indicate the estimates of a linear model fit to the data and the shaded bands represent the 95% confidence intervals. We mark 50% recognition accuracy in the plot using a dotted line. **b** Percentage recognition accuracy (*y*-axis) as a function of the interference slope (*x*-axis), calculated by fitting a general linear model of recognition accuracy (binomial link) as a function of semantic interference at testing (*z*-scored) independently for each participant. Each point in the plot represents an individual participant for the two levels of image novelty (new scenes, green circles; old scene, yellow triangles). We mark with dotted lines the 50% recognition accuracy and when interference slope is zero (i.e., semantic interference has no effects on recognition accuracy)

values (GLMM), and *p* values for each model. The level of significance was calculated from an *F* test based on the Satterthwaite approximation to the effective degrees of freedom (Satterthwaite, 1946), whereas *p* values in GLMMs were based on asymptotic Wald tests. Interference slopes were instead analyzed using general linear models because they were obtained from by-participant linear regressions (i.e., we have no random effects; see section Dependent Variables). We predicted recognition accuracy, expressed as a probability, as a function of the interference slope separately for old and new trials to examine

whether semantic interference was stronger in any of these two conditions. These models will be directly reported in the text.

Results

### Recognition accuracy and interference slope

We found significant main effects of SI and image novelty. The higher the SI, the worse the recognition performance was. New images were more likely to be correctly rejected as not seen, than old images being correctly recognized as seen. Interestingly, we observed a significant interaction between SI and image novelty, such that the effect of semantic interference was stronger for old than new images (refer to Fig. 3a for a visualization and Table 1 for the model coefficients). This interaction is substantiated by recognition probability being significantly predicted by interference slopes only in old trials, $\beta = .3$, $t(24) = 2.53$, $p = .02$, compared with new trials, $\beta = .15$, $t(24) = 1.67$, $p = .11$, which corroborates that the effect of SI was more prevalent for images previously seen (see Fig. 3b for a visualization). When looking at the intrinsic memorability of our images, we confirmed that there is a significant correlation between the split halves ($r_s = .2$, $p < .001$)—that is, there is consistency in the images that are better remembered, but the strength of our correlation was much weaker than the one originally reported ($r_s = .72$, $p < .001$; refer to Isola et al., 2014).

Table 1 Generalized linear-mixed model of recognition accuracy (a binomial variable; 0 = incorrect, 1 = correct) as a function: Semantic interference (a continuous variable, 1–120, *z*-scored) and image novelty (old, novel; with novel as the reference level)

Predictors	Response accuracy		
	Estimates	*SE*	*z*-value
(Intercept)	1.47	0.11	13.97***
SI	−0.08	0.03	−2.96**
Image novelty	−1.51	0.04	−42.14***
SI × Image Novelty	−0.09	0.04	−2.68**
(*) $p < .10$, *$p < 0.05$, **$p < .01$, ***$p < .001$			

Note. Random intercepts included in the model are Participant (25) and the Scenes (1,488), which are nested into their respective Categories (12).
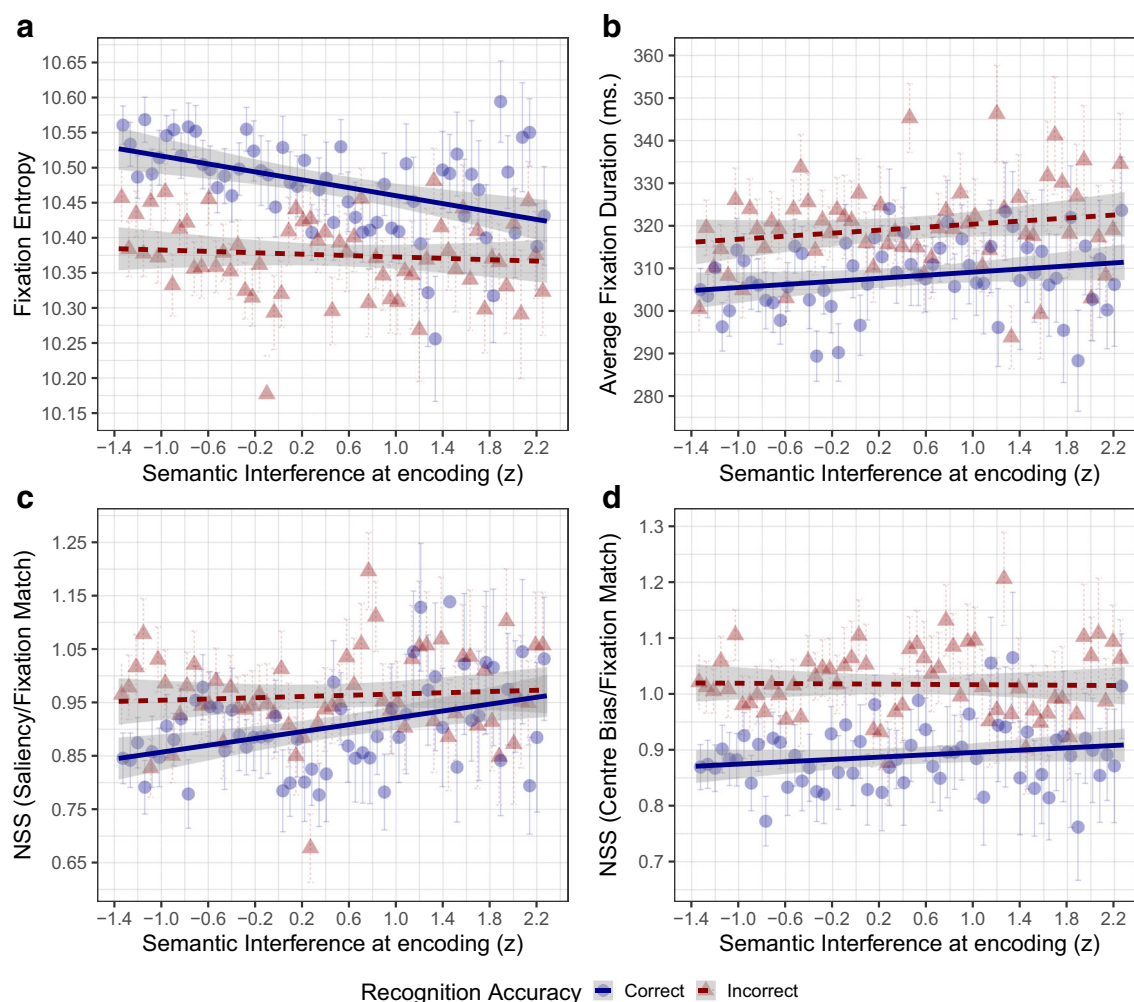
Fig. 4 Eye-movement measures during the encoding of images as a function of semantic interference (a continuous variable ranging from 1 to 60, z-scored). **a** Entropy of the distribution of fixations across the image. **b** Average fixation duration in milliseconds. **c** Normalized scan-path saliency score. **d** Normalized scan-path for central bias. Each individual point represents the average of each dependent measure across participants for that level of interference, and distinguishing images that were later correctly recognized (blue circle, solid line) or not (red triangle, dashed line). The lines indicate the fit of linear regression models with 95% confidence interval represented as shaded bands

Table 2 Linear-mixed model outputs for fixation entropy, average fixation duration, normalized scan-path saliency and normalized scan-path for centre bias as a function: Semantic interference (a continuous variable, 1–60, z-scored) and recognition accuracy scaled to reduce collinearity (incorrect = −1, correct = 1)

Predictors	Fixation entropy			Fixation duration (ms)			NSS (fixation/saliency)			NSS (fixation/centre bias)		
	Estimates	SE	t-value	Estimates	SE	t-value	Estimates	SE	t-value	Estimates	SE	t-value
(Intercept)	10.43	0.03	356.59***	313.32	4.12	76.09***	0.93	0.05	18.77***	0.95	0.05	28.93***
SI	−0.03	0.004	−6.49***	2.43	0.7	3.54***	0.01	0.007	2.08*	0.01	0.005	1.52
Accuracy	0.04	0.004	10.73***	−3.86	0.73	−5.29***	−0.03	0.006	−5.05***	−0.05	0.005	−10.38***
SI:Accuracy	−0.01	0.004	−2.11*	−0.06	0.7	−0.08	0.008	0.006	1.4	0.01	0.005	0.94

(*) $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note. Random intercepts included in the model are Participant (23), and the Scenes (1,488), which are nested into their respective Categories (12).

Fixation entropy[7]

The spread of fixations, indexed as entropy, was significantly greater for images that were later better recognized. Most importantly, entropy significantly decreased as a function of SI, and the slope of this decrease was steeper for images that were better recognized (refer to Fig. 4a, Table 2, for the model coefficients and to Supplementary Material S3 for additional visualizations).

Average fixation duration

Fixations had a significantly shorter duration in images that were subsequently correctly recognized compared to those that were not (Table 2). Again, we observed a significant main effect of SI, whereby the duration of individual fixations increased as SI also increased (refer to Fig. 4b, and inferential results in Table 2).

NSS (fixation/saliency)

The correspondence between fixation position and low-level visual salient regions of the scene was also significantly related to the memorability of the scene and it was impacted by semantic interference. NSS was significantly lower for later correctly recognized scenes, and higher for increasing SI (refer to Fig. 4c and Table 2).

NSS (fixation/centre bias)

The higher the correspondence between fixation locations and the centre proximity map during the encoding of the scene, the less likely it would be that a scene is later correctly recognized. We did not find any significant main effect of semantic interference, nor this factor interacted with memory recognition (refer to Fig. 4d and Table 2).[8]

Discussion

The concept of interference has played a pivotal role in the theories of memory since its very beginning (e.g., McGeoch & McDonald, 1931; Müller & Pilzecker, 1900; Skaggs, 1933), and helped framing the processes that may hinder, or aid, the formation and access of information in memory (e.g., mental activities, Cowan et al., 2004, or competition between stimuli sharing content, Craig et al., 2013). Similarity-based semantic interference of visual information, operationalized as the frequency of images (or objects) belonging to the same category that

---

[8] We reused the centre proximity map weighting FES maps to compute the NSS correspondence between such a map and fixation positions. We corroborated the same result shown in the main text using the weight matrix by Hayes and Henderson (2020): a greater tendency to inspect the centre of the display was associated with worsen memory recognition ($\beta = -.05$, $SE = .001$, $t = -8.95$, $p < .001$).

participants are asked to memorize, for example, was shown to be detrimental to recognition processes: the higher the semantic interference, the poorer the recognition performance (Konkle et al., 2010a, 2010b). Most importantly, the information we store in memory is acquired through our senses and so, memory for different types of stimuli (e.g., words or images) is known, for example, to be linked to eye-movement responses (see Hannula, 2018, or Ryan & Shen, 2020, for reviews).

The core objective of the current study was to provide empirical links between the detrimental effect of similarity-based semantic interference on long-term visual memory and the patterns of overt attention deployed as scenes are studied to be later recalled. On recognition accuracy, we replicated using an old/new paradigm the effect of SI found by Konkle et al. (2010b) on a 2AFC (i.e., the higher the SI, the worse the memory performance is). The use of a different paradigm allowed us to discover that recognition accuracy for images seen during encoding (i.e., old scenes) was lower than accuracy for novel images; and most importantly that the detrimental effect of SI on old scenes was significantly stronger than on new scenes (refer also to the analysis of interference slopes). This result confirms that old/new paradigms probably tap into different recognition mechanisms than 2AFC (i.e., recollection more than familiarity; Cunningham et al., 2015), and that interference mostly disrupt existing memory representations. It is important to note that even if participants could successfully discriminate the signal over the noise, they became more conservative in their responses as semantic interference increased (i.e., they required substantial evidence before making an "old" judgment; see Appendix A for additional analyses of d-prime and criterion). This is in line with prior work showing that when distractors are highly similar to targets (Benjamin & Bawa, 2004) or scenes are familiar (Dobbins & Kroll, 2005), as it was the case in our study, a more conservative criterion is used.

When looking at the intrinsic memorability of images in our task (Bainbridge, 2020), we find it to hold, even though weaker than originally reported (Isola et al., 2014). We qualitatively interpret this comparison as indicating that, despite scenes being intrinsically memorable, the effect of semantic interference in our design reduced their individual discriminability.

Eye-movement measures demonstrated that four key components of fixation responses (overall spread, average duration, their correspondence with low-level visual saliency and with central bias) during encoding of images, systematically related to memory formation and were impacted by semantic interference. On patterns of global exploration, measured as entropy of fixations' spread across the scene, we observed exploration to become more selective as SI increased. This result parallels the evidence that being exposed to the same scene induces a reduction in the number of visited regions (Ryan et al., 2000). As the fidelity of visual memory representations decreases due to SI, overt attention focuses to local regions of scenes in search for

distinctive details that could boost their individual memorability. However, this switch from global to local processing may be an indicator of disrupted memory processes (Macrae & Lewis, 2002). Indeed, in our study, the wider a scene was explored, the more likely it was later successfully recognized (see Damiano & Walther, 2019, for corroborating findings). Of note, the shrinkage of fixations to more local regions due to semantic interference was stronger for correctly recognized scenes (i.e., two-way interaction, SI × Accuracy). We argue that as semantic interference deteriorates memory recognition, it pushes fixation entropy of subsequently recognized scenes to approximately the same level of scenes that are later forgotten (see Fig. 4a).

The average duration of fixations, an index of cognitive effort to acquire visual information, was longer for later forgotten scenes, and increased as semantic interference also increased. This result resembles the finding of increased fixation durations to repeated exposure of the same stimulus (e.g., Ryan et al., 2007), and conceptually links with the evidence of increased fixation duration in perceptually degraded images (Loftus et al., 1992). In practice, as the conceptual overlap between images grows due to semantic interference (i.e., they become more and more similar), a greater allocation of overt attention is required to accrue more information at each fixation that can in turn be used to make each individual image more distinct. Greater attentional effort, however, also implied lower recognition accuracy. Literature on object-scene integration shows that objects violating the contextual fit of the scene (e.g., a toothbrush in a kitchen) require longer fixations and are harder to be integrated (e.g., Coco et al., 2020, for recent behavioural and neural evidences). So, if fixation duration indexes more complex processing, it may also point at encoding difficulties, and hence explain why its increase may be associated with worse recognition accuracy.

We also examined the reliance of participants to low-level features of images as evidence of strategic compensation to increasing semantic interference. Here, we found that indeed overt attention was allocated more frequently to regions of the images that were rich in low-level features as semantic interference increased. This result is intriguing because it points at a reduction in top-down control due to the increase in content overlap of the images, and a shift towards bottom-up stimulus-driven control, as is usually observed in free-viewing tasks (Parkhurst et al., 2002). However, attending to low-level features of the image, in general, was detrimental to its later recognition (see main effect of accuracy).

A similar negative impact on recognition memory was observed when examining the tendency of observers to re-orient their gaze towards the centre of the display (e.g., Tatler, 2007). We found that a greater focus of overt attention to the centre of display during scene encoding indicated a worsen later recognition. This result corroborated our observation with fixation entropy, whereby a reduced exploration implied worse recognition accuracy and it confirms that scene exploration is key to the successful encoding and later retrieval of visual information from memory.

Previous attempts to link the visual saliency of images, or other low-level oculo-motor mechanisms such as the tendency to re-centre gaze, to their memorability had shown a lack of significant association (e.g., Isola et al., 2014; Lyu et al., 2020). A possible explanation of this discrepancy may relate to the fact that these studies have explored the relation between visual saliency, or centre-bias, and memorability without taking directly into account the associated eye-movement responses. In fact, when replicating the analysis by Lyu et al. (2020) of centre-bias, which does not include eye-movements, we confirm it not to be significantly associated with recognition memory. Instead, when we modeled the correspondence between fixation positions and GBVS maps, with (and without) centre bias adjustment, we confirmed a highly significant association between low-level visual saliency and recognition memory (see Supplementary Material S4, for greater details). Thus, we contribute to these previous findings by showing that the role played by low-level features on image memorability may be better accounted for when investigated relative to overt attention. However, as the study by Hayes and Henderson (2020) points out, low-level visual saliency and centre bias are often confounded, and so more accurate predictions of overt attention during scene viewing can only be obtained when the latter is used to adjust the former. We acknowledge that more research is needed to elucidate the patterns of interaction between different oculo-motor responses in face of semantic interference and in relation to memory recognition. One potential approach would be to compare the predictability of memory recognition of different models including a variety of oculo-motor responses (e.g., centre bias, fixation entropy) and evaluate the contribution of each model parameter to prediction performance (see also Coco & Keller, 2014, for an example application).

Another point of caution in the results of the current study is that recognition accuracy for old scenes was rather low, even at low-level of interference, which may cast doubts on how informative eye-movement measures really are about memory processes that were inherently weak. It is important to note that we only considered eye-movement responses collected during encoding, and in this phase, the mechanisms of explicit memory recognition were not yet at work. Moreover, a significant main effect of semantic interference on eye-movement responses was observed regardless of whether participants successfully recalled, or not, the scenes. Thus, even though, memory for old images was surprisingly poor, we doubt that this may have had any important repercussion to the effects of semantic interference on the oculo-motor responses reported here.

In sibling research, we investigated whether the effect of semantic interference is also observed in a healthy older population, and especially, whether this mechanism may be impacted by neuro-degenerative diseases (Coco et al., in press).
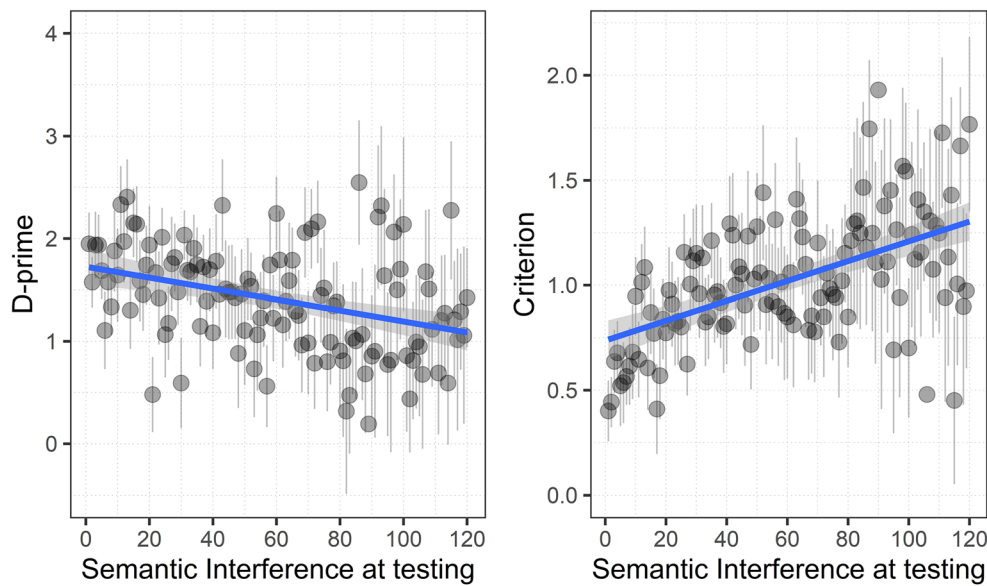
Fig. 5 *D*-prime (left-panel) and criterion (right-panel) as a function of semantic interference at testing (a continuous variable from 1 to 120). Each individual point is the average across participants for that level of interference. Lines indicate the estimates from a linear model fit to the data and the shaded bands represent the 95% confidence intervals

Results showed corroborating effects of semantic interference on recognition accuracy in the healthy older group, which are, however, significantly reduced in people with mild cognitive impairment. We also replicated similar patterns of eye-movement responses, such as the decrease of fixation entropy and the greater reliance on low-level visual saliency for increasing semantic interference at encoding, while also showing subtle oculo-motor compensatory strategies in the MCI group.

An outstanding question that germinates from this study regards the interplay between low-level and high-level features of scenes. In fact, even though two images of a kitchen may belong to the same semantic category, they may be very different in terms of their perceptual features or configurational statistics of the objects they are made of. So, future research should aim at developing computational measures, and novel paradigms, that can better disentangle the contribution of these two components in memory interference.
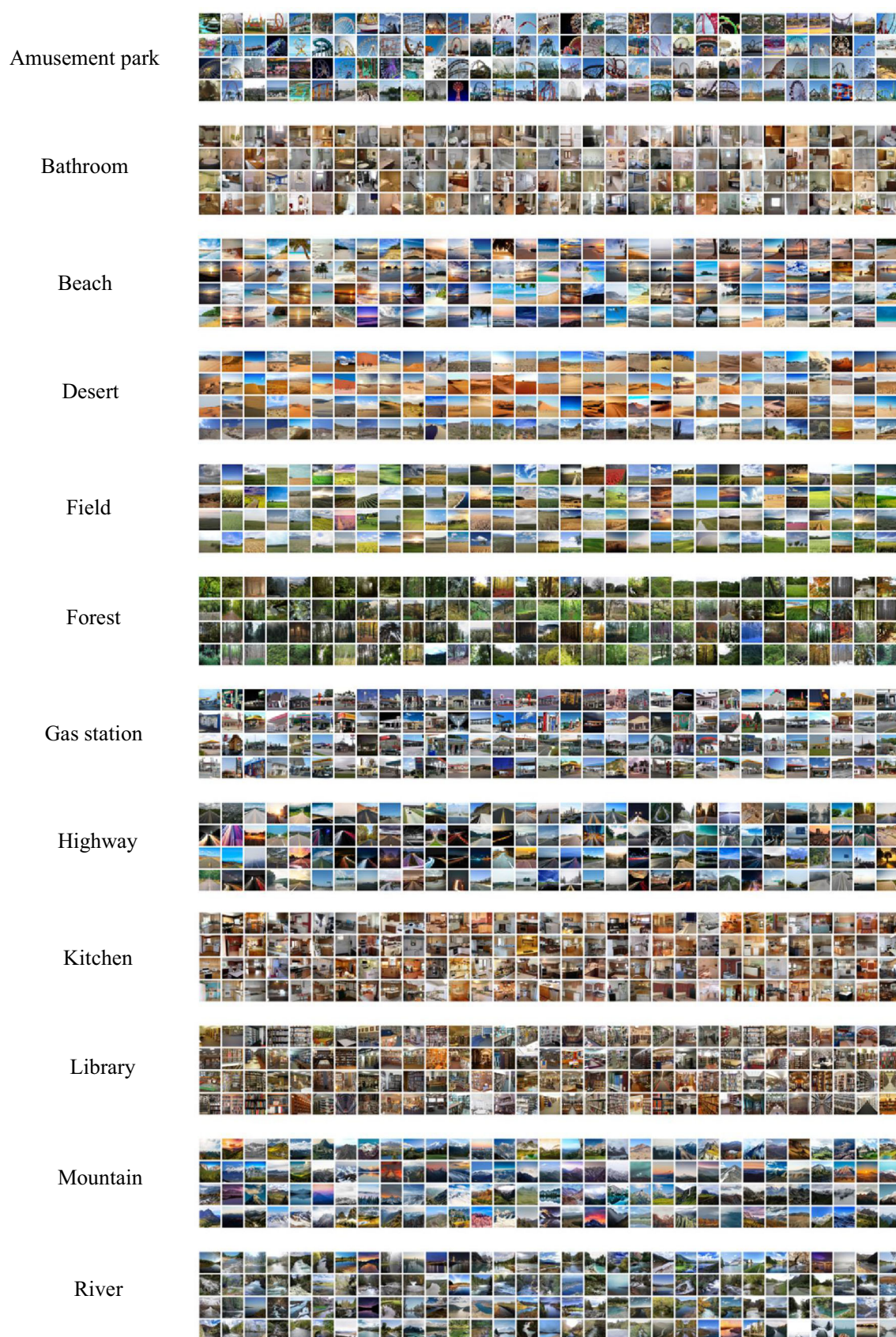
In sum, our findings of systematic links between overt attention and memory mechanisms during high-level cognitive processing support the centrality of the oculo-motor system on memory formation (e.g., Chun & Turk-Browne, 2007; Ryan et al., 2020), calling for more integrative research between attention and memory.

Appendices

D-prime and criterion

In this analysis, we tested whether participants were able to discriminate the signal from the noise using *d*-prime, while using criterion to determine the direction of participants' choices in case of uncertainty. We also examined whether the effect of semantic interference was confirmed by these two measures. We found a *d*-prime significantly above zero, which indicates that participants' performance was not at random. The criterion showed that participants were conservative (i.e., greater tendency to respond "no" rather than "yes") explaining the higher rate of correct rejections (i.e., accurate responses for the novel scene) than hits (i.e., accurate responses for old scenes). On both measures, we found a significant effect of semantic interference whereby the higher the SI, the smaller the *d*-prime and the higher the

Table 3 Linear-mixed models for the *d*-prime (left column) and criterion (right-column) as a function semantic interference (a continuous variable, 1–120, *z*-scored)

Predictors	*D*-prime			Criterion		
	Estimates	*SE*	*t*-value	Estimates	*SE*	*t*-value
(Intercept)	1.4	0.11	12.44***	1.01	0.14	7.05***
SI	−0.21	0.04	−4.99***	0.17	0.02	7.48***

(*) *p* < .10, * *p* < .05, ** *p* < .01, *** *p* < .001

Note. The random intercept included in the model is participant (25).

Fig. 6 Miniatures of all scenes used in this study organized by semantic category

criterion (see Fig. 5 and Table 3 for a visualization and model coefficients).

Visualization of the miniatures of all scenes used in this study organized by semantic category.

We refer the reader to Supplementary Material S1 for another version of this visualization with scenes displayed at a higher resolution Fig. 6.

References

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(4), 997–1010. https://doi.org/10.1037/0278-7393.25.4.997

Baddeley, A. D., & Dale, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *5*(5), 417–420. https://doi.org/10.1016/S0022-5371(66)80054-3

Bainbridge, W. A. (2019). Memorability: How what we see influences what we remember. *Psychology of Learning and Motivation—Advances in Research and Theory*, *70*, 1–27. https://doi.org/10.1016/bs.plm.2019.02.001

Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia*, *141*(June 2019), 107408. https://doi.org/10.1016/j.neuropsychologia.2020.107408

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. 10.18637/jss.v067.i01

Bayley, P. J., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2008). Yes/no recognition, forced-choice recognition, and the human hippocampus. *Journal of Cognitive Neuroscience*, *20*(3), 505–512. https://doi.org/10.1162/jocn.2008.20038

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language*, *51*(2), 159–172. https://doi.org/10.1016/j.jml.2004.04.001

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*, 165–178. https://doi.org/10.1016/j.visres.2015.03.005

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2019). What Do Different Evaluation Metrics Tell Us About Saliency Models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(3), 740–757.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(6), 1–15. https://doi.org/10.1167/9.3.6.Introduction

Chun, M. M. and Turk-Browne, N. B. (2007) 'Interactions between attention and memory', *Current Opinion in Neurobiology, 17*(2), pp. 177–184. https://doi.org/10.1016/j.conb.2007.03.005.

Coco, M. I., & Keller, F. (2014). Classification of visual and linguistic tasks using eye-movement features. *Journal of Vision*, *14*(3), 11. https://doi.org/10.1167/14.3.11

Coco, M. I., Merendino, G., Zappalà, G., & Della Sala, S. (in press). Semantic interference mechanisms on long-term visual memory and their eye-movement signatures in mild cognitive impairment. *Neuropsychology.* https://doi.org/10.1037/neu0000734

Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related brain potentials during semantic integration of object–scene information. *Journal of Cognitive Neuroscience*, *32*(4), 571–589. https://doi.org/10.1162/jocn_a_01504

Cowan, N., Beschin, N., & Della Sala, S. (2004). Verbal recall in amnesiacs under conditions of diminished retroactive interference. *Brain*, *127*(4), 825–834. https://doi.org/10.1093/brain/awh107

Craig, K. S., Berman, M. G., Jonides, J., & Lustig, C. (2013). Escaping the recent past: Which stimulus dimensions influence proactive interference? *Memory & Cognition*, *41*(5), 650–670. https://doi.org/10.3758/s13421-012-0287-0

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. https://doi.org/10.1101/lm.039404.115

Dale, H. C. A. (1964). Retroactive interference in short-term memory. *Nature*, *203*(1063), 1964. https://doi.org/10.1038/2031408a0

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://doi.org/10.3758/s13428-013-0422-2

Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and retrieval. *Cognition*, *184*(December 2018), 119–129. https://doi.org/10.1016/j.cognition.2018.12.014

Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief wakeful resting boosts new memories over the long term. *Psychological Science*, *23*(9), 955–960. https://doi.org/10.1177/0956797612441220

Dewar, M. T., Cowan, N., & Sala, S. Della. (2007). Forgetting due to retroactive interference: A fusion of Müller and Pilzecker's (1990) early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex*, *43*(5), 616–634. https://doi.org/10.1016/S0010-9452(08)70492-1

Dobbins, I. G., & Kroll, N. E. A. (2005). Distinctiveness and the recognition mirror effect: Evidence for an item-based criterion placement heuristic. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(6), 1186–1198. https://doi.org/10.1037/0278-7393.31.6.1186

Dubey, R., Peterson, J., Khosla, A., Yang, M. H., & Ghanem, B. (2015). What makes an object memorable? *Proceedings of the IEEE International Conference on Computer Vision*, *2015 Inter*, 1089–1097. https://doi.org/10.1109/ICCV.2015.130

Hannula, D. E. (2018). Attention and long-term memory: Bidirectional interactions and their effects on behavior. In K. Federmeier (Ed.0, *Psychology of learning and motivation—Advances in research and theory* (Vol. 69). Elsevier. https://doi.org/10.1016/bs.plm.2018.09.004

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, *19*, 545–552.

Hayes, T. R., & Henderson, J. M. (2020). Center bias outperforms image salience but not semantics in accounting for attention during scene

viewing. *Attention, Perception, & Psychophysics*, *82*(3), 985–994. https://doi.org/10.3758/s13414-019-01849-7

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. C. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, *49*(5), 1802–1823. https://doi.org/10.3758/s13428-016-0822-1

Ishiguro, S., & Saito, S. (2020). The detrimental effect of semantic similarity in short-term memory tasks: A meta-regression approach. *Psychonomic Bulletin & Review*, 1–25. Advance online publication. https://doi.org/10.3758/s13423-020-01815-7

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. https://doi.org/10.1109/TPAMI.2013.200

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transaction on Pattern Analysis and Machine Learning*, *20*(11), 1254–1259. https://doi.org/10.1109/TPAMI.2012.125

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology*, *64*(10), 1971–1989. https://doi.org/10.1080/17470218.2011.588335

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. https://doi.org/10.1037/a0019165

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. https://doi.org/10.1177/0956797610385359

Loftus, G. R., Kaufman, L., Nishimoto, T., & Ruthruff, E. (1992). Effects of visual degradation on eye-fixation duration, perceptual processing, and longterm visual memory. In Eye movements and visual cognition (pp. 203–226). Springer, New York, NY.

Lyu, M., Choe, K. W., Kardan, O., Kotabe, H., Henderson, J., & Berman, M. (2020). Overt attentional correlates of scene memorability and their relationships to scene semantics. *Journal of Vision*, *20*(9), 1–17. 10.31234/osf.io/3e8qm

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science, 13*(2), 194–196.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McGeoch, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *The American Journal of Psychology*, *43*(4), 579–588.

Meghanathan, R. N., van Leeuwen, C., & Nikolaev, A. R. (2015). Fixation duration surpasses pupil size as a measure of memory load in free viewing. *Frontiers in Human Neuroscience*, *8*(JAN), 1–9. https://doi.org/10.3389/fnhum.2014.01063

Müller, G. E., & Pilzecker, A. (1900). *Experimentelle beiträge zur lehre vom gedächtniss* (J. Barth (ed.)).

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123. https://doi.org/10.1016/S0042-6989(01)00250-4

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*(18), 2397–2416. https://doi.org/10.1016/j.visres.2005.03.019

Pomplun, M., Ritter, H., & Velichkovsky, B. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception*, *25*(8), 931–948. https://doi.org/10.1068/p250931

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture word interference task. *Child Develop.*, *46*(1), 247–253. https://doi.org/10.2307/1128859

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science*, *11*(6), 454–461. https://doi.org/10.1111/1467-9280.00288

Ryan, J. D., Hannula, D. E., & Cohen, N. J. (2007). The obligatory effects of memory on eye movements. *Memory*, *15*(5), 508–525. https://doi.org/10.1080/09658210701391022

Ryan, J. D., & Shen, K. (2020). The eyes are a window into memory. *Current Opinion in Behavioral Sciences*, *32*, 1–6. https://doi.org/10.1016/j.cobeha.2019.12.014

Ryan, J. D., Shen, K., & Liu, Z. (2020). The intersection between the oculomotor and hippocampal memory systems : empirical developments and clinical implications. *Annals of the New York Academy of Sciences*, *1464*(1), 115–141. https://doi.org/10.1111/nyas.14256

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110–114. https://doi.org/10.2307/3002019

Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin*, *75*(6), 399–415. https://doi.org/10.1037/h0031257

Skaggs, E. B. (1933). A discussion on the temporal point of interpolation and degree of retroactive inhibition. *Journal of Comparative Psychology*, *16*, 411–414. https://doi.org/10.1037/h0074460

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 1–17. https://doi.org/10.1167/7.14.4

Tavakoli, H. R., Rahtu, E., & Heikkil, J. (2011). Fast and efficient saliency detection using sparse sampling and kernel density estimation. In A. Heyden & F. Kahl (Eds.), *Scandinavian conference on image analysis* (pp. 666–675). Springer. https://doi.org/10.1007/978-3-642-21227-7_62

Underwood, B. J. (1945). The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs*, *59*(273), 1–33. https://doi.org/10.1037/h0093547

Wechsler, D. (1999). *Manual for the Wechsler Abbreviated Scale of Intelligence*. Psychological Corporation.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970