



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

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

Citation for published version:

Mikhailova, A, Raposo, A, Della Sala, S & Coco, MI 2021, 'Eye-movements reveal semantic interference effects during the encoding of naturalistic scenes in long-term memory', *Psychonomic Bulletin & Review*, vol. 28, pp. 1601–1614. <https://doi.org/10.3758/s13423-021-01920-1>

Digital Object Identifier (DOI):

[10.3758/s13423-021-01920-1](https://doi.org/10.3758/s13423-021-01920-1)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Psychonomic Bulletin & Review

Publisher Rights Statement:

This is a post-peer-review, pre-copyedit version of an article published in Psychon Bull Rev . The final authenticated version is available online at: <https://doi.org/10.3758/s13423-021-01920-1>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



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

3
4 **^{a,b} Anastasiia Mikhailova (a.mikhailova@psicologia.ulisboa.pt)**

5
6
7 **^aAna Raposo (alraposo@psicologia.ulisboa.pt)**

8
9
10 **^cSergio Della Sala (sergio@ed.ac.uk)**

11
12
13 **^{a,d} Moreno I. Coco* (M.Coco@uel.ac.uk)**

14
15
16
17
18 ^aFaculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013, Lisbon,
19 Portugal

20 ^bInstituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisbon, Portugal

21 ^cHuman Cognitive Neuroscience, Psychology, University of Edinburgh, 7 George Square, EH8 9JZ,
22 Edinburgh, UK

23 ^dSchool of Psychology, University of East London, Water Lane, E15 4LZ, London, UK
24

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; visual saliency; eye-tracking.

Introduction

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 Dewar, Cowan, & Della Sala, 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 memorised (Craig et al., 2013; Underwood, 1945), such as semantically related vs. unrelated word lists (Baddeley & Dale, 1966; McGeoch & McDonald, 1931 and 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, 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, Brady, Alvarez, and Oliva (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- or low-level characteristics or to the type of tasks and the depth

109 of cognitive processing involved (Bainbridge, 2020). The memorability of visual
110 information also relies on patterns of extrinsic responses, e.g., eye-movements, that
111 participants generate when encoding such information in memory (e.g., Bylinskii, Isola,
112 Bainbridge, Torralba, & Oliva, 2015; and see Hannula, 2018, for a review of the topic). For
113 example, a higher number of fixations, or smaller pupil dilations while scenes are studied
114 in preparation for a recognition test are associated with a better memory performance
115 (Kafkas & Montaldi, 2011).

116 Global measures of exploration as obtained from attention maps (Pomplun et al.,
117 1996), in which all fixations on a given image are portrayed along its two-dimensions, are
118 an important predictor of its memorability too. In this context, the more spread out fixations
119 were on an image during encoding, which implied that several regions were attended to,
120 the better this image was later recalled (e.g., Damiano & Walther, 2019; and Lyu et al.,
121 2020, for another application of attention maps in the context of image memorability).

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

127 Another important aspect of oculo-motor control, which has been relatively neglected
128 in the context of memory processes, is the tendency of observers to re-orient their overt
129 attention towards the centre of the display during scene viewing (i.e., centre bias, Tatler,
130 2007). To the best of our knowledge, Lyu et al., (2020) is the only study that has examined
131 the role of centre bias on memory recognition. A centre proximity map was computed to
132 weight low-level saliency maps generated with the Graph-Based Visual Saliency (GBVS)
133 algorithm (Harel et al., 2006) and a single value, representing the probability of salient
134 regions to be positioned in the centre of the display, generated. Their result did not show
135 any significant relationship between centre bias and memorability, which seems to confirm
136 the marginal role played by low-level visual features on scene memorability.

137 Attending and memorising are indeed closely coupled; but high-level semantic
138 mechanisms of interference may influence overt attention as memories get formed. If this
139 supposition is true, we should be able to bridge the expected decrement in recognition
140 memory onto eye-movement responses. Our proposition is that as the fidelity of individual

141 memory representations (e.g., the specific image of a kitchen) degrades under the influence
142 of semantic interference (i.e., a memorised pool of kitchens), oculo-motor compensatory
143 strategies are adopted to cope with the increased complexity of discriminating the memory
144 of each individual instance from a pool of semantically overlapping competitor instances.

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

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

161 On recognition accuracy, we expect to replicate the semantic interference effect
162 observed by Konkle, et al., (2010b), whereby the higher the interference of the semantic
163 category a scene is encoded into, the worse it would be its future recognition. However, if
164 this effect truly relates to memory representations, then it should more strongly manifest in
165 old rather than new images. Moreover, even though images are intrinsically memorable
166 (Bainbridge, 2020), in our paradigm, we expect their memorability to reduce under the
167 influence of semantic interference, and so observe a lower inter-participant correlation than
168 Isola et al. (2014), where the semantic interference between images was not manipulated.

169 On the eye-movement data collected while scenes were viewed for the first time (i.e.,
170 at encoding), we focus on four complementary measures: (a) the amount of visual
171 information that was attended to, by looking at the overall spread of fixations across the
172 scene, (b) the attentional effort to acquire visual information from the scene, by looking at

173 fixation duration, (c) the reliance of participants on low-level visual features of the scene,
174 by looking at the correspondence between fixation positions and visual saliency at such
175 locations and (d) the tendency of participants to re-orient their overt attention towards the
176 centre of the screen, by looking at the correspondence between fixation positions and a
177 centre proximity map (see section *Dependent Variables* for formal definitions of these
178 measures).

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

192 Semantic interference degrades the representational fidelity of individual instances
193 by reinforcing their categorical overlap, and so we expect fixation duration to significantly
194 increase to keep instances discriminable as a response. This prediction will conceptually
195 align with the study by Ryan et al. (2007), showing that the repeated exposure to familiar
196 faces resulted into progressively longer fixation durations; and connect with Loftus et al.
197 (1992) showing that degradation, albeit in the perceptual domain, was associated to an
198 increase in fixation duration in a long-term visual memory task. Moreover, as fixation
199 duration is an index of processing effort (see Coco et al., 2020 for an example in the context
200 of object-scene semantic integration), we expect it to be negatively associated with
201 recognition accuracy, whereby the longer the average fixation duration is, the less likely
202 the scene was efficiently encoded into memory.

203 Moreover, if participants indeed search for potentially diagnostic features in scenes
204 as semantic interference increases, they would rely more on low-level visual features of the

205 scene, i.e., a higher correspondence between fixation position and low-level visual saliency.
206 Building upon Isola et al. (2014), however, we do not expect low-level image features to
207 significantly contribute on whether the scene will be later correctly recognised, or not.

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

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

Insert Figure 1

Method

Participants

Twenty-five native English speakers (17 females, age = 21.95 ± 3.47 SD, range: 20-36) 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 results¹. As the eye-movement data of 2 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.

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 < 0.05, which should minimize the chance of incurring into Type 2 errors.

235 ***Material and Apparatus***

236 We selected 1,488 naturalistic images from SUN database (Xiao et al., 2010) with a
237 minimum of 550 x 550 pixels resolution and which did not include animate objects like
238 humans or animals. All images were cropped and rescaled to 800 x 800 pixels to collect
239 finer-grained eye-movement responses, and were equally drawn from one of twelve
240 categories, six human-made environments (i.e., amusement park, bathroom, gas station,
241 highway, kitchen, library) and six natural environments (i.e., beach, desert, field, forest,
242 mountain, river). Miniatures of all scenes by category are reported in Appendix A and with
243 a greater resolution in the Supplementary Materials² (S1).

244 Images (800 x 800 pixels) were centrally presented on a black background at their
245 resolution³ with a 19" Dell Monitor (16.2-inch x 7.2 inch) screen resolution of 1920 x 1024
246 and set at a viewing distance of ~60 cm. Eye-movement data was recorded binocularly
247 using a Gazepoint GP3 HD eye-tracker sampling 150Hz. The experiment was built on
248 OpenSesame 3.1.9 (Mathôt et al., 2012) and the acquisition of eye-tracking data made
249 possible through the PyGaze Python plug-in (Dalmaijer et al., 2014). Each participant was
250 calibrated on a 9-points, and recalibrated if necessary. The mean degree of visual angle
251 deviation accepted for the calibration was 0.37 degrees on the x-axis (SD = 0.15) and 0.53
252 degrees on the y-axis (SD = 0.29).

253

254 ***Procedure***

255 Participants were administered the WASI-II test of intelligence⁴ (Wechsler, 1999)
256 at the start of each session (Full Scale = 117.79 ± 10.25; 97-140) and then completed a
257 long-term visual memory task which assessed their recognition accuracy using an old/new
258 approach (see Figure 1 for a visualisation of the experimental design). Each participant
259 watched a stream of 372 images, each presented for 3 seconds with 800ms fixation
260 crosshair inter-trial, during encoding. After a short 10 minutes break, she/he was tested on
261 744 images, presented one-by-one. Half of these scenes were the 372 images seen during
262 the encoding phase and the remaining 372 were novel scenes. They were asked to indicate

² The full stimuli dataset 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.

263 whether they remembered or not the image using the keyboard (> yes; < no). The image
264 was visible until a recognition response was made. To implement the semantic interference
265 manipulation, we varied the frequency of images of each semantic category that
266 participants were exposed to. In line with Konkle et al. (2010b), we varied SI in 4 levels
267 (4, 20, 40 or 60 images per category). The total of 1,488 scenes was obtained by selecting
268 124 unique scenes for each of the 12 different categories (124 * 12), which is the number
269 of images needed to cover all 4 levels of semantic interference, distributed across 8
270 randomization lists. Each level of interference was distributed in each list to 3 different
271 categories (4 * 3) and all 4 levels of interference were counterbalanced across semantic
272 categories by rotating such levels onto 4 different lists (e.g., if the kitchen category had an
273 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
274 in list 4). Four additional lists were created by swapping old with novel scenes between the
275 encoding and the recognition phase to ensure that all scenes were seen in both conditions.
276 Images from each semantic category and level of interference were randomly assigned to
277 the lists making sure that they were never repeated within each list. Images in both phases
278 of encoding and recognition were presented in randomized order, i.e., we did not block
279 images sharing the same semantic category to appear contiguously, and equal frequency
280 per semantic category (e.g., for 20 kitchen images in the encoding phase, there were 20 old
281 and 20 novel kitchen images in the testing phase). The experimental session took
282 approximately 2 hours.

283

284

Analyses

285

Data exclusion

286

287

288

289

290

291

292

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 analysed 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

293 recognition response time⁵ (177) below 1% or above 99% of their respective distributions.
294 Thus, the number of encoding trials analysed was 7,942 trials (an average of 345.3 ± 28.17
295 per participant).

296

297 *Independent variables*

298 The key independent variable of this study is *SI*, which was manipulated in the design
299 as frequency of scenes belonging the same semantic category (i.e., 4, 20, 40, 60), and
300 incrementally administered to participants during the entire experimental session (i.e., from
301 1 to 60 during encoding and from 1 to 120 in the recognition phase, which were the
302 maximum number of scenes belonging to the same category that could be seen in either
303 phase). We treated *SI* as a continuous variable⁶ to capture its incremental impact on
304 recognition accuracy and oculo-motor responses on a trial-by-trial basis (refer to
305 Supplementary Material S2 for a visualization of this measure) and standardised it into *z*-
306 scores to minimize convergence issues (e.g., co-linearity) when it was introduced in the
307 regression. *Image novelty* (old and new, set as reference level) was another independent
308 variable that we included in the analyses of recognition accuracy to distinguish between
309 hits and correct rejections. The last independent variable included in the analysis of eye-
310 movement at encoding was the recognition *accuracy* of old scenes, which made possible
311 to differentiate oculo-motor strategies that support effective (vs. ineffective) memory
312 processes. This independent variable was also scaled prior to entering it in the regression
313 to minimize co-linearity with the other predictors.

314

315

316

Insert Figure 2 Here

317

318 *Dependent variables*

319 **Manual responses:** *Recognition accuracy* is a binomial variable which indicates
320 whether a scene was correctly remembered as already seen (old) or rejected as novel (new).
321 Following Konkle et al., (2010a), we also fit a general linear model (binomial link)

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.

322 predicting recognition accuracy as a function of SI independently for each participant, and
323 separately for old and new trials. In this way, we derived the *interference slope* (i.e., the
324 beta coefficient associated with SI), which reflects how much was the recognition accuracy
325 of each participant impacted by increasing interference for scenes she/he already viewed
326 (old) or had never seen (new). Negative coefficients indicate that recognition accuracy
327 decreased when semantic interference increased. In Appendix A, we report additional
328 analyses of *d-prime* and *criterion* to explore how was the signal (hit) discriminated from
329 the noise (false-alarm) by the participants of our task, whether they adopted a conservative
330 or a liberal strategy, and examined the impact of semantic interference on both. Finally,
331 from the recognition accuracy we obtained the *intrinsic memorability* of our images. We
332 used the method by Isola et al., (2014) and computed the Spearman correlation between the
333 recognition accuracy of each individual scene (as hits) in two randomly split sets of
334 participants, and iterated this procedure 50 times to avoid that findings may spuriously
335 relate to a precise random selection of the participants' split.

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

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

348 To compute (b), we first built a fixation probability map of each trial by placing at
349 fixation coordinates, Gaussian kernels with a bandwidth set at 1 degree of visual angle
350 (roughly 27 pixels) to approximate the size of the fovea. The height of the Gaussian was
351 weighted by the proportion of time spent fixating at that location to better integrate
352 differences in the amount of overt attention deployed across the scene. Then, the entropy
353 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

354 the normalized fixation probability at the coordinates of the fixation (x, y) in the scene S
355 (see Castelhana et al., 2009; Coco & Keller, 2014 or Henderson, 2003 for related
356 examples). Thus, the higher the fixation entropy, the more spread out fixations across the
357 scene are. In Figure 2, we visualize four example heatmaps of fixation distributions in low
358 vs. high entropy organized as columns and low vs. high interference organized as rows and
359 report the value of fixation entropy for each map.

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

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

382

383 *Inferential statistics*

384 We used linear mixed-effects models (LMM) and generalized linear mixed-effects
385 models (GLMM) as implemented in the lme4 package in R (Bates et al., 2015) to conduct

386 the statistical analyses of our dependent measures. The fixed effects of our models, i.e., our
387 independent variables, were introduced as main effects as well as in interaction. The
388 random effects were Participant (23) and Scenes (1,488), which were nested into their
389 respective Categories (12) and introduced as intercepts. In the table of results, we reported
390 the beta coefficients, t-values (LMM), z-values (GLMM), and p-values for each model.
391 The level of significance was calculated from an F-test based on the Satterthwaite
392 approximation to the effective degrees of freedom (Satterthwaite, 1946), whereas p-values
393 in GLMMs were based on asymptotic Wald tests. Interference slopes were instead analysed
394 using general linear models because they were obtained from by-participant linear
395 regressions (i.e., we have no random effects, see section *Dependent Variables*). We
396 predicted recognition accuracy, expressed as a probability, as a function of the interference
397 slope separately for old and new trials to examine whether semantic interference was
398 stronger in any of these two conditions. These models will be directly reported in the text.

399
400 -----
401 Insert Figure 3 and Table 1 Here
402 -----

403 404 **Results**

405
406 **Recognition Accuracy and Interference Slope:** We found significant main effects of SI
407 and image novelty. The higher the SI, the worse the recognition performance was. New
408 images were more likely to be correctly rejected as not seen, than old images being
409 correctly recognized as seen. Interestingly, we observed a significant interaction between
410 SI and image novelty, such that the effect of semantic interference was stronger for old than
411 new images (refer to Figure 3A for a visualisation and Table 1 for the model coefficients).
412 This interaction is substantiated by recognition probability being significantly predicted by
413 interference slopes only in old trials [$\beta = .3$, $t(24) = 2.53$, $p = .02$] compared to new trials
414 [$\beta = .15$, $t(24) = 1.67$, $p = .11$], which corroborates that the effect of SI was more prevalent
415 for images previously seen (see Figure 3B for a visualisation). When looking at the intrinsic
416 memorability of our images, we confirmed that there is a significant correlation between
417 the split halves ($r_s = .2$, $p < 0.001$), i.e., there is consistency in the images that are better

418 remembered, but the strength of our correlation was much weaker than the one originally
419 reported ($r_s = .72$, $p < 0.001$; refer to Isola et al., 2014).

420 -----

421 Insert Figure 4 and Table 2 Here

422 -----

423

424 **Fixation entropy**⁷: The spread of fixations, indexed as entropy, was significantly greater
425 for images that were later better recognized. Most importantly, entropy significantly
426 decreased as a function of SI, and the slope of this decrease was steeper for images that
427 were better recognised (refer to Figure 4A, Table 2 for the model coefficients and to
428 Supplementary Material S3 for additional visualisations).

429

430 **Average fixation duration**: Fixations had a significantly shorter duration in images that
431 were subsequently correctly recognized compared to those that were not (Table 2). Again,
432 we observed a significant main effect of SI, whereby the duration of individual fixations
433 increased as SI also increased (refer to Figure 4B, and inferential results in Table 2).

434

435 **NSS (fixation/saliency)**: The correspondence between fixation position and low-level
436 visual salient regions of the scene was also significantly related to the memorability of the
437 scene and it was impacted by semantic interference. NSS was significantly lower for later
438 correctly recognized scenes, and higher for increasing SI (refer to Figure 4C and Table 2).

439

440 **NSS (fixation/centre bias)**: The higher the correspondence between fixation locations and
441 the centre proximity map during the encoding of the scene, the less likely it would be that
442 a scene is later correctly recognized. We did not find any significant main effect of semantic
443 interference, nor this factor interacted with memory recognition (refer to Figure 4D and
444 Table 2)⁸.

445

⁷ We also examined number of fixations, which correlates with fixation entropy ($r = 0.76$) and found a very similar pattern of results.

⁸ We re-used 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 & Henderson, 2020: a greater tendency to inspect the centre of the display was associated with worsen memory recognition ($\beta = -.05$, $SE = .001$, $t = -8.95$, $p < .001$).

446

Discussion

447 The concept of interference has played a pivotal role in the theories of memory since
448 its very beginning (e.g., McGeoch & McDonald, 1931; Müller & Pilzecker, 1900; Skaggs,
449 1933), and helped framing the processes that may hinder, or aid, the formation and access
450 of information in memory (e.g., mental activities, Cowan et al., 2004, or competition
451 between stimuli sharing content, Craig et al., 2013). Similarity-based semantic interference
452 of visual information, operationalised as the frequency of images (or objects) belonging to
453 the same category that participants are asked to memorise, for example, was shown to be
454 detrimental to recognition processes: the higher the semantic interference, the poorer the
455 recognition performance (Konkle et al., 2010a, 2010b). Most importantly, the information
456 we store in memory is acquired through our senses and so, memory for different types of
457 stimuli (e.g., words or images) is known, for example, to be linked to eye-movement
458 responses (see Hannula, 2018 or Ryan & Shen, 2020 for reviews).

459 The core objective of the current study was to provide empirical links between the
460 detrimental effect of similarity-based semantic interference on long-term visual memory
461 and the patterns of overt attention deployed as scenes are studied to be later recalled. On
462 recognition accuracy, we replicated using an old/new paradigm the effect of SI found by
463 Konkle et al., (2010b) on a 2AFC, i.e., the higher the SI, the worse the memory performance
464 is. The use of a different paradigm allowed us to discover that recognition accuracy for
465 images seen during encoding (i.e., old scenes) was lower than accuracy for novel images;
466 and most importantly that the detrimental effect of SI on old scenes was significantly
467 stronger than on new scenes (refer also to the analysis of interference slopes). This result
468 confirms that old/new paradigms probably tap into different recognition mechanisms than
469 2AFC (i.e., recollection more than familiarity, Cunningham, Yassa, & Egeth, 2015), and
470 that interference mostly disrupt existing memory representations. It is important to note
471 that even if participants could successfully discriminate the signal over the noise, they
472 became more conservative in their responses as semantic interference increased, i.e., they
473 required substantial evidence before making an “old” judgment (see Appendix A for
474 additional analyses of d-prime and criterion). This is in line with prior work showing that
475 when distractors are highly similar to targets (Benjamin & Bawa, 2004) or scenes are
476 familiar (Dobbins & Kroll, 2005), as it was the case in our study, a more conservative
477 criterion is used.

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

483 Eye-movement measures demonstrated that four key components of fixation
484 responses (overall spread, average duration, their correspondence with low-level visual
485 saliency and with central bias) during encoding of images, systematically related to
486 memory formation and were impacted by semantic interference. On patterns of global
487 exploration, measured as entropy of fixations' spread across the scene, we observed
488 exploration to become more selective as SI increased. This result parallels the evidence that
489 being exposed to the same scene induces a reduction in the number of visited regions (Ryan
490 et al., 2000). As the fidelity of visual memory representations decreases due to SI, overt
491 attention focuses to local regions of scenes in search for distinctive details that could boost
492 their individual memorability. However, this switch from global to local processing may
493 be an indicator of disrupted memory processes (Macrae & Lewis, 2002). Indeed, in our
494 study, the wider a scene was explored, the more likely it was later successfully recognized
495 (see Damiano and Walther, 2019, for corroborating findings). Of note, the shrinkage of
496 fixations to more local regions due to semantic interference was stronger for correctly
497 recognized scenes (i.e., two-way interaction $SI \times Accuracy$). We argue that as semantic
498 interference deteriorates memory recognition, it pushes fixation entropy of subsequently
499 recognized scenes to approximately the same level of scenes that are later forgotten (see
500 Figure 4A).

501 The average duration of fixations, an index of cognitive effort to acquire visual
502 information, was longer for later forgotten scenes, and increased as semantic interference
503 also increased. This result resembles the finding of increased fixation durations to repeated
504 exposure of the same stimulus (e.g., Ryan et al., 2007), and conceptually links with the
505 evidence of increased fixation duration in perceptually degraded images (Loftus et al.,
506 1992). In practice, as the conceptual overlap between images grows due to semantic
507 interference, i.e., they become more and more similar, a greater allocation of overt attention
508 is required to accrue more information at each fixation that can in turn be used to make
509 each individual image more distinct. Greater attentional effort, however, also implied lower

510 recognition accuracy. Literature on object-scene integration shows that objects violating
511 the contextual fit of the scene (e.g., a toothbrush in a kitchen) require longer fixations as
512 are harder to be integrated (e.g., Coco et al., 2020 for recent behavioural and neural
513 evidences). So, if fixation duration indexes more complex processing, it may also point at
514 encoding difficulties, and hence explain why its increase may be associated with worse
515 recognition accuracy.

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

524 A similar negative impact on recognition memory was observed when examining the
525 tendency of observers to re-orient their gaze towards the centre of the display (e.g., Tatler,
526 2007). We found that a greater focus of overt attention to the centre of display during scene
527 encoding indicated a worsen later recognition. This result corroborated our observation
528 with fixation entropy, whereby a reduced exploration implied worse recognition accuracy
529 and it confirms that scene exploration is key to the successful encoding and later retrieval
530 of visual information from memory.

531 Previous attempts to link the visual saliency of images, or other low-level oculo-
532 motor mechanisms such as the tendency to re-centre gaze, to their memorability had shown
533 a lack of significant association (e.g., Isola et al., 2014; Lyu et al., 2020). A possible
534 explanation of this discrepancy may relate to the fact that these studies have explored the
535 relation between visual saliency, or centre-bias, and memorability without taking directly
536 into account the associated eye-movement responses. In fact, when replicating the analysis
537 by Lyu et al., 2020 of centre-bias, which does not include eye-movements, we confirm it
538 not to be significantly associated with recognition memory. Instead, when we modeled the
539 correspondence between fixation positions and GBVS maps, with (and without) centre bias
540 adjustment, we confirmed a highly significant association between low-level visual
541 saliency and recognition memory (see Supplementary Material S4, for greater details).

542 Thus, we contribute to these previous findings by showing that the role played by low-level
543 features on image memorability may be better accounted for when investigated relative to
544 overt attention. However, as the study by Hayes & Henderson (2020) points out, low-level
545 visual saliency and centre bias are often confounded, and so more accurate predictions of
546 overt attention during scene viewing can only be obtained when the latter is used to adjust
547 the former. We acknowledge that more research is needed to elucidate the patterns of
548 interaction between different oculo-motor responses in face of semantic interference and in
549 relation to memory recognition. One potential approach would be to compare the
550 predictability of memory recognition of different models including a variety of oculo-motor
551 responses (e.g., centre bias, fixation entropy, etc.) and evaluate the contribution of each
552 model parameter to prediction performance (see also Coco & Keller, 2014 for an example
553 application).

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

564 In sibling research, we investigated whether the effect of semantic interference is also
565 observed in a healthy older population, and especially, whether this mechanism may be
566 impacted by neuro-degenerative diseases (Coco et al., 2021). Results showed corroborating
567 effects of semantic interference on recognition accuracy in the healthy older group, which
568 are, however, significantly reduced in people with Mild Cognitive Impairment. We also
569 replicated similar patterns of eye-movement responses, such as the decrease of fixation
570 entropy and the greater reliance on low-level visual saliency for increasing semantic
571 interference at encoding, while also showing subtle oculo-motor compensatory strategies
572 in the MCI group.

573 An outstanding question that germinates from this study regards the interplay
574 between low-level and high-level features of scenes. In fact, even though two images of a
575 kitchen may belong to the same semantic category, they may be very different in terms of
576 their perceptual features or configurational statistics of the objects they are made of. So,
577 future research should aim at developing computational measures, and novel paradigms,
578 that can better disentangle the contribution of these two components in memory
579 interference.

580 In sum, our findings of systematic links between overt attention and memory
581 mechanisms during high-level cognitive processing support the centrality of the oculo-
582 motor system on memory formation (e.g., Chun & Turk-Browne, 2007; Ryan et al., 2020),
583 calling for more integrative research between attention and memory.

584

585 **Acknowledgements**

586 This research was supported by Fundação para a Ciência e Tecnologia under Grant
587 [PTDC/PSI-ESP/30958/2017] to MIC (PI) and AR (Co-PI), PhD scholarship to AM
588 [SFRH/BD/144453/2019]; and the Leverhulme Trust under Grant [ECF-014-205] to MIC.
589 We would also like to thank a group of undergraduate Psychology students at the University
590 of Edinburgh for collecting the data in fulfilling their 3rd year Mini Dissertation essay. They
591 are Maria Costa Pinto, Nicholas Koh, Katrien Kwan, Oli Timmis and Isabelle Sinclair.

592

593 The data and R script to run the analyses and visualise the results of this manuscript
594 are available at https://osf.io/7kj3y/?view_only=772098e3f8964fdbaec5704f558a755b

595

596

597

598

599

600

601

602

603

604

- 606 Althoff, R. R., & Cohen, N. J. (1999). Eye-Movement-Based Memory Effect: A Reprocessing
607 Effect in Face Perception. *Journal of Experimental Psychology. Learning, Memory, and*
608 *Cognition*, 25(4), 997–1010.
- 609 Baddeley, A. D., & Dale, H. C. A. (1966). The effect of semantic similarity on retroactive
610 interference in long- and short-term memory. *Journal of Verbal Learning and Verbal*
611 *Behavior*, 5(5), 417–420. [https://doi.org/10.1016/S0022-5371\(66\)80054-3](https://doi.org/10.1016/S0022-5371(66)80054-3)
- 612 Bainbridge, W. A. (2019). Memorability: How what we see influences what we remember.
613 *Psychology of Learning and Motivation - Advances in Research and Theory*, 70, 1–27.
614 <https://doi.org/10.1016/bs.plm.2019.02.001>
- 615 Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate
616 from attention and priming. *Neuropsychologia*, 141(June 2019), 107408.
617 <https://doi.org/10.1016/j.neuropsychologia.2020.107408>
- 618 Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects
619 Models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
620 <https://doi.org/10.18637/jss.v067.i01>
- 621 Bayley, P. J., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2008). Yes/No Recognition , Forced-
622 choice Recognition, and the Human Hippocampus. *Journal of Cognitive Neuroscience*, 20(3),
623 505–512.
- 624 Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition.
625 *Journal of Memory and Language*, 51(2), 159–172. <https://doi.org/10.1016/j.jml.2004.04.001>
- 626 Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic
627 effects on image memorability. *Vision Research*, 116, 165–178.
628 <https://doi.org/10.1016/j.visres.2015.03.005>
- 629 Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2019). What Do Different Evaluation
630 Metrics Tell Us About Saliency Models? *IEEE Transactions on Pattern Analysis and Machine*
631 *Intelligence*, 41(3), 740–757.
- 632 Castelhana, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye
633 movement control during active scene perception. *Journal of Vision*, 9(6), 1–15.
634 <https://doi.org/10.1167/9.3.6.Introduction>
- 635 Coco, M. I., & Keller, F. (2014). Classification of visual and linguistic tasks using eye-movement
636 features. *Journal of Vision*, 14(3), 11. <https://doi.org/10.1167/14.3.11>
- 637 Coco, M. I., Merendino, G., Zappalà, G., & Della Sala, S. (2021). Semantic interference
638 mechanisms on long-term visual memory and their eye-movement signatures in Mild
639 Cognitive Impairment. *Neuropsychology*, in press.
- 640 Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related brain potentials during
641 semantic integration of object–scene information. *Journal of Cognitive Neuroscience*, 32(4),
642 571–589. https://doi.org/10.1162/jocn_a_01504
- 643 Cowan, N., Beschin, N., & Della Sala, S. (2004). Verbal recall in amnesiacs under conditions of
644 diminished retroactive interference. *Brain*, 127(4), 825–834.
645 <https://doi.org/10.1093/brain/awh107>
- 646 Craig, K. S., Berman, M. G., Jonides, J., & Lustig, C. (2013). Escaping the recent past: Which
647 stimulus dimensions influence proactive interference? *Memory and Cognition*, 41(5), 650–670.
648 <https://doi.org/10.3758/s13421-012-0287-0>
- 649 Cunningham, C. A., Yassa, M. A., & Egeth, H. E. (2015). Massive memory revisited: Limitations
650 on storage capacity for object details in visual long-term memory. *Learning and Memory*,
651 22(11), 563–566. <https://doi.org/10.1101/lm.039404.115>
- 652 Dale, H. C. A. (1964). Retroactive Interference in Short-term Memory. *Nature*, 203(1063), 1964.
- 653 Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: an open-source, cross-platform
654 toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research*
655 *Methods*, 46(4), 913–921. <https://doi.org/10.3758/s13428-013-0422-2>
- 656 Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding

657 and retrieval. *Cognition*, 184(December 2018), 119–129.
658 <https://doi.org/10.1016/j.cognition.2018.12.014>

659 Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief Wakeful Resting Boosts
660 New Memories Over the Long Term. *Psychological Science*, 23(9), 955–960.
661 <https://doi.org/10.1177/0956797612441220>

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

665 Dobbins, I. G., & Kroll, N. E. A. (2005). Distinctiveness and the recognition mirror effect:
666 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>

667
668
669

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

672
673 Hannula, D. E. (2018). Attention and long-term memory: Bidirectional interactions and their effects
674 on behavior. In *Psychology of Learning and Motivation - Advances in Research and Theory*
675 (1st ed., Vol. 69). Elsevier Inc. <https://doi.org/10.1016/bs.plm.2018.09.004>

676 Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth
677 a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, 4, 166. <https://doi.org/10.3389/fnhum.2010.00166>

678
679 Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 545–552.

680
681 Hayes, T. R., & Henderson, J. M. (2020). Center bias outperforms image salience but not semantics
682 in accounting for attention during scene viewing. *Attention, Perception, and Psychophysics*,
683 82(3), 985–994. <https://doi.org/10.3758/s13414-019-01849-7>

684 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>

685
686 Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. C. (2017). Noise-robust fixation
687 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>

688
689 Ishiguro, S., & Saito, S. (2020). The detrimental effect of semantic similarity in short-term memory
690 tasks: A meta-regression approach. *Psychonomic Bulletin and Review*, 1–25.

691 Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph
692 memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1469–
693 1482. <https://doi.org/10.1109/TPAMI.2013.200>

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

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

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

703
704 Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than
705 you think: the role of categories in visual long-term memory. *Psychological Science*, 21(11),
706 1551–1556. <https://doi.org/10.1177/0956797610385359>

707 Loftus, G. R., Kaufman, L., Nishimoto, T., & Ruff, E. R. (1992). *Perceptual Processing, and Long-Term Visual Memory*.

708

- 709 Lyu, M., Choe, K. W., Kardan, O., Kotabe, H., Henderson, J., & Berman, M. (2020). Overt
710 attentional correlates of scene memorability and their relationships to scene semantics. *Journal*
711 *of Vision*, 20(9), 1–17. <https://doi.org/10.31234/osf.io/3e8qm>
- 712 Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical
713 experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324.
714 <https://doi.org/10.3758/s13428-011-0168-7>
- 715 McGeoch, J. A., & McDonald, W. T. (1931). Meaningful Relation and Retroactive Inhibition. *The*
716 *American Journal of Psychology*, 43(4), 579–588.
- 717 Meghanathan, R. N., van Leeuwen, C., & Nikolaev, A. R. (2015). Fixation duration surpasses pupil
718 size as a measure of memory load in free viewing. *Frontiers in Human Neuroscience*, 8(JAN),
719 1–9. <https://doi.org/10.3389/fnhum.2014.01063>
- 720 Müller, G. E., & Pilzecker, A. (1900). *Experimentelle beiträge zur lehre vom gedächtniss* (J. Barth
721 (ed.)).
- 722 Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt
723 visual attention. *Vision Research*, 42(1), 107–123. [https://doi.org/10.1016/S0042-](https://doi.org/10.1016/S0042-6989(01)00250-4)
724 [6989\(01\)00250-4](https://doi.org/10.1016/S0042-6989(01)00250-4)
- 725 Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in
726 natural images. *Vision Research*, 45(18), 2397–2416.
727 <https://doi.org/10.1016/j.visres.2005.03.019>
- 728 Pomplun, M., Ritter, H., & Velichkovsky, B. (1996). Disambiguating complex visual information:
729 Towards communication of personal views of a scene. *Perception*, 25(8), 931–948.
730 <https://doi.org/10.1068/p250931>
- 731 Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a
732 picture word interference task. *Child Develop.*, 46(1), 247–253.
733 <https://doi.org/10.2307/1128859>
- 734 Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational
735 memory. *Psychological Science*, 11(6), 454–461. <https://doi.org/10.1111/1467-9280.00288>
- 736 Ryan, J. D., Hannula, D. E., & Cohen, N. J. (2007). The obligatory effects of memory on eye
737 movements. *Memory*, 15(5), 508–525. <https://doi.org/10.1080/09658210701391022>
- 738 Ryan, J. D., & Shen, K. (2020). The eyes are a window into memory. *Current Opinion in*
739 *Behavioral Sciences*, 32, 1–6. <https://doi.org/10.1016/j.cobeha.2019.12.014>
- 740 Ryan, J. D., Shen, K., & Liu, Z. (2020). The intersection between the oculomotor and hippocampal
741 memory systems : empirical developments and clinical implications. *Annals of the New York*
742 *Academy of Sciences*, 1464(1), 115–141. <https://doi.org/10.1111/nyas.14256>
- 743 Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components.
744 *Biometrics Bulletin*, 2(6), 110–114. <https://doi.org/10.2307/3002019>
- 745 Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin*, 75(6),
746 399–415. <https://doi.org/10.1037/h0031257>
- 747 Skaggs, E. B. (1933). A discussion on the temporal point of interpolation and degree of retroactive
748 inhibition. *Journal of Comparative Psychology*, 16, 411–414.
- 749 Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing
750 position independently of motor biases and image feature distributions. *Journal of Vision*,
751 7(14), 1–17. <https://doi.org/10.1167/7.14.4>
- 752 Tavakoli, H. R., Rahtu, E., & Heikkil, J. (2011). Fast and efficient saliency detection using sparse
753 sampling and kernel density estimation. In *Scandinavian conference on image analysis* (pp.
754 666–675). Springer Berlin Heidelberg.
- 755 Underwood, B. J. (1945). The Effect of Successive Interpolations on Retroactive and Proactive
756 Inhibition. *Psychological Monographs*, 59(273), 1–33.
- 757 Wechsler, D. (1999). Manual for the Wechsler abbreviated scale of intelligence. *San Antonio, TX:*
758 *Psychological Corporation.*
- 759 Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale
760 scene recognition from abbey to zoo. *Proceedings of the IEEE Computer Society Conference*

761 *on Computer Vision and Pattern Recognition*, 3485–3492.
762 <https://doi.org/10.1109/CVPR.2010.5539970>
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793

794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825

Insert Figure 1A and Table 1A Here

Appendix A

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 0, 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 criterion (see Figure 1A and Table 1A for a visualisation and model coefficients).

826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857

Insert Figure 1B

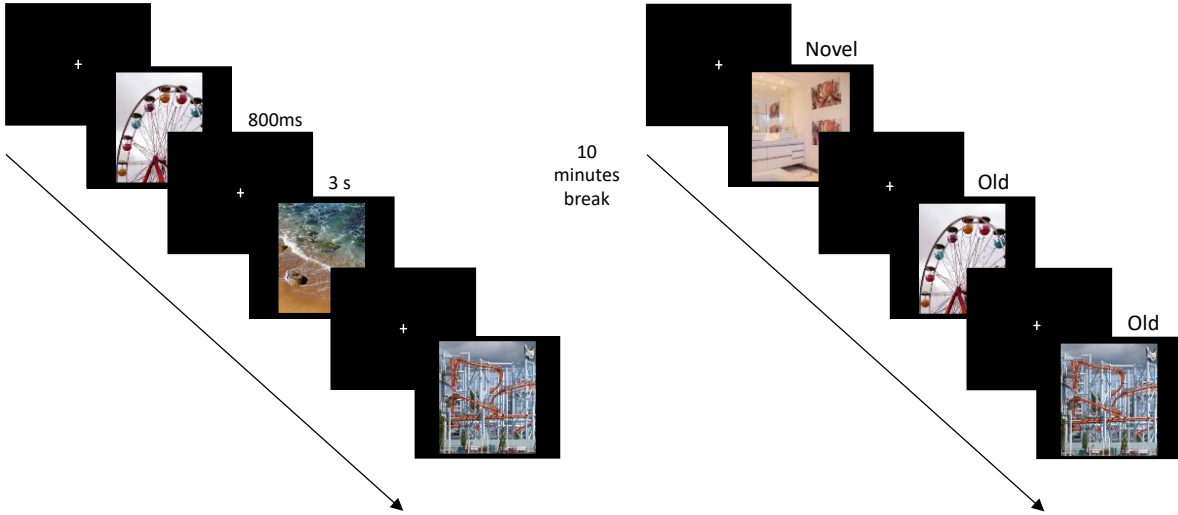
Appendix B

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.

Encoding phase: Memorization task
4, 20, 40 and 60 Exemplars per Category

Testing phase: Recognition task
Have you seen this image before?



859

860 **Figure 1:** Visualisation of the experimental design, procedure and example images used in this study.

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

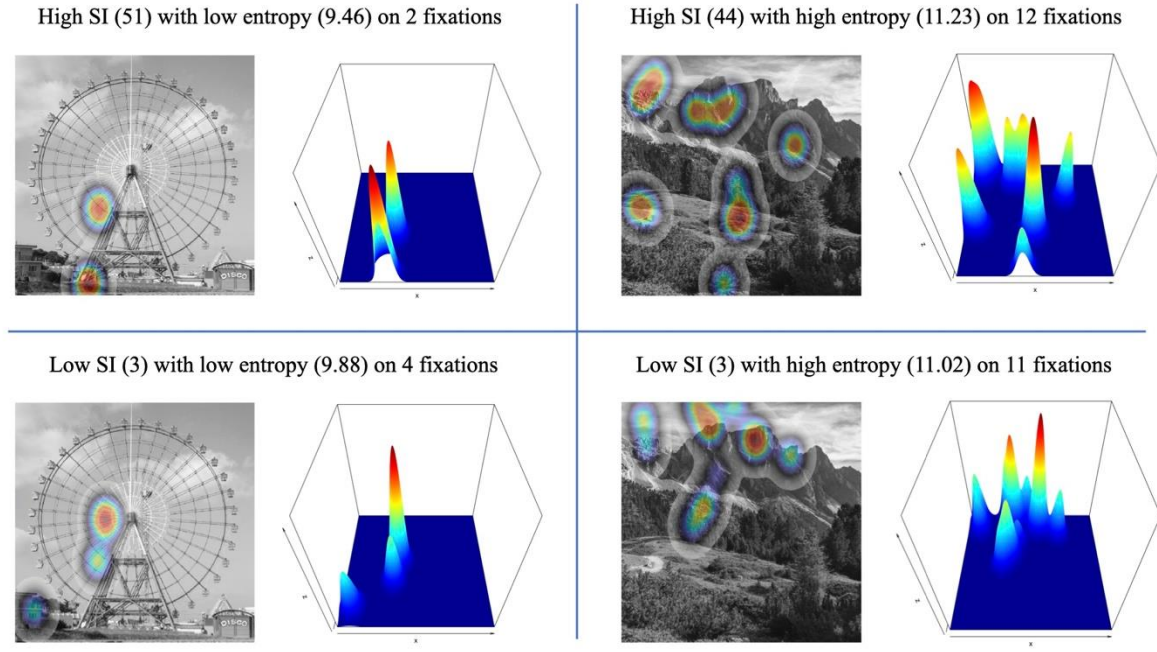
877

878

879

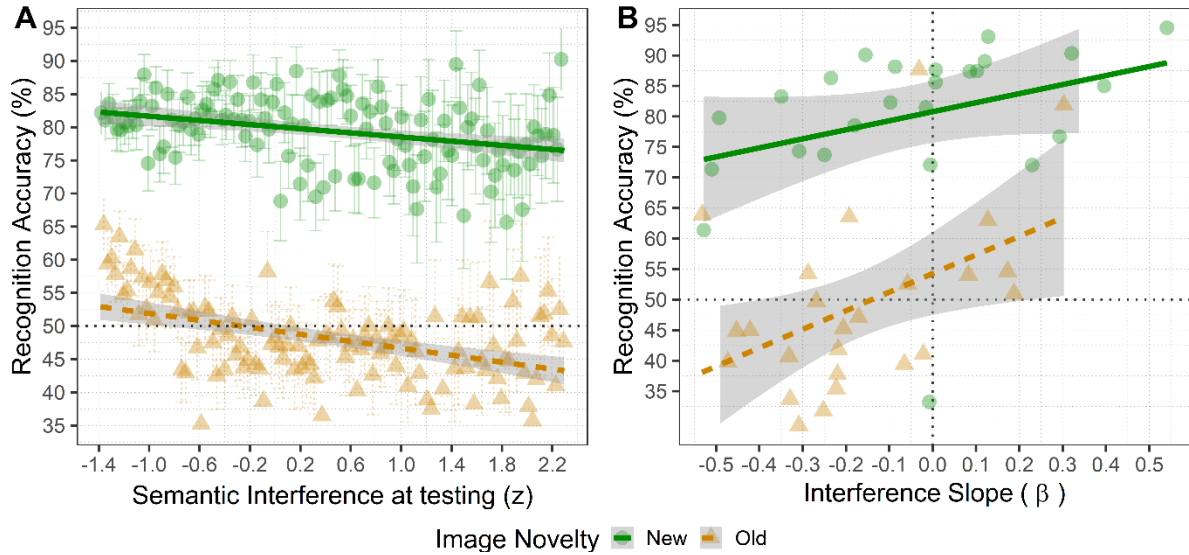
880

881



883
884 **Figure 2.** Examples of attention maps with a high and low fixation entropy (left column, right column) when images
885 were encoded at a high or a low level of semantic interference (top-row, bottom row). On each panel, we present the
886 attention map as an heatmap (left) and as a 3D landscape to better visualize how Gaussians were fit to fixation position
887 and their height scaled by fixation duration. In bracket, we report the fixation entropy obtained from each attention map.

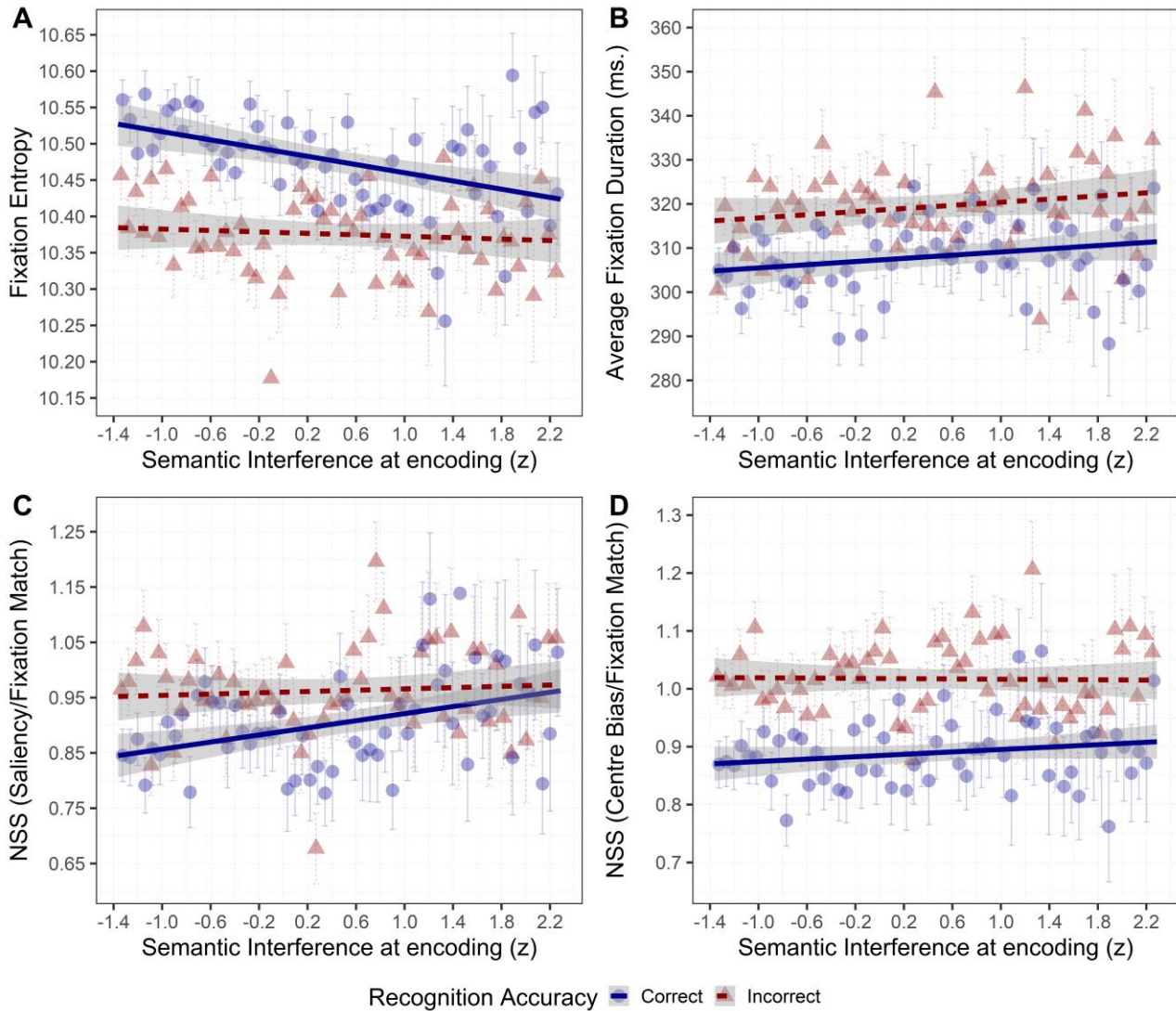
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916



918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957

Figure 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 0, i.e., semantic interference has no effects on recognition accuracy.

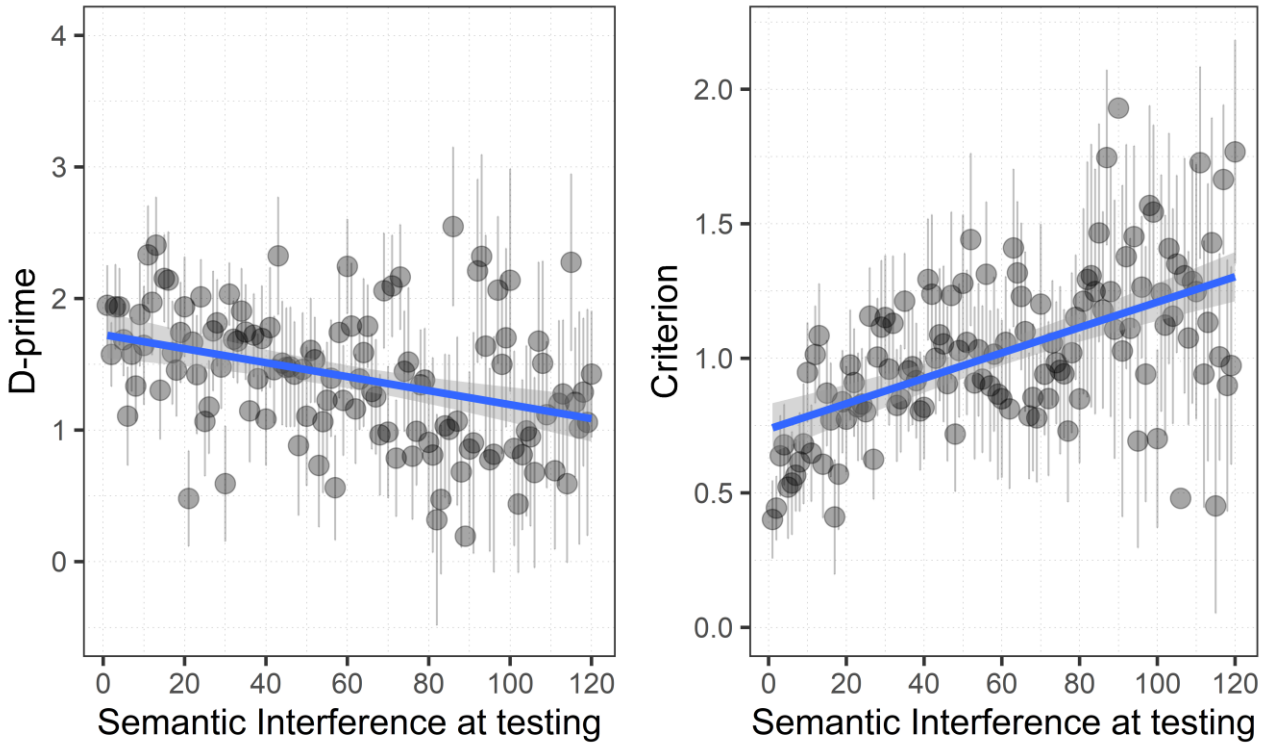
958 Figure 4:
959



960
961 **Figure 4.** Eye-movement measures during the encoding of images as a function of Semantic Interference (a continuous
962 variable ranging from 1 to 60, z-scored). (A) Entropy of the distribution of fixations across the image, (B) Average
963 fixation duration in milliseconds, (C) Normalized Scan-Path Saliency score and (D) Normalized Scan-Path for central
964 bias. Each individual point represents the average of each dependent measure across participants for that level of
965 interference, and distinguishing images that were later correctly recognised (blue circle, solid line) or not (red triangle,
966 dashed line). The lines indicate the fit of linear regression bands with 95% confidence interval represented as shaded
967 bands.

968
969
970
971
972
973
974
975
976
977
978
979
980
981

982 Figure 1A:
983



984
985 **Figure 1A:** D-prime (left-panel) and criterion (right-panel) as a function of semantic interference at testing (a continuous variable
986 from 1 to 120). Each individual point is the average across participants for that level of interference. Lines indicate the estimates
987 from a linear model fit to the data and the shaded bands represent the 95% confidence intervals
988

989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009



1011
1012 **Figure 1B.** Miniatures of all scenes used in this study organised by semantic category.

1015 Table 1:

1016

<i>Predictors</i>	Response Accuracy		
	<i>Estimates</i>	<i>SE</i>	<i>z-value</i>
(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 < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

1017

1018 **Table 1.** Generalised linear-mixed model of recognition accuracy (a binomial variable; 0 = Incorrect, 1 = Correct) as a
1019 function: Semantic Interference (a continuous variable, 1-120, z-scored) and Image Novelty (Old, Novel; with Novel as
1020 the reference level). Random intercepts included in the model are Participant (25) and the Scenes (1,488), which are
1021 nested into their respective Categories (12).

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

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 < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

1059
 1060
 1061
 1062
 1063
 1064
 1065
 1066
 1067
 1068
 1069
 1070
 1071
 1072
 1073
 1074
 1075
 1076
 1077
 1078
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096
 1097
 1098

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

<i>Predictors</i>	D-prime			Criterion		
	<i>Estimates</i>	<i>SE</i>	<i>t-value</i>	<i>Estimates</i>	<i>SE</i>	<i>t-value</i>
(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***

(*) $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1100

1101 **Table 1A:** Linear-mixed model for the d-prime (left column) and criterion (right-column) as a function Semantic
 1102 Interference (a continuous variable, 1-120, z-scored). The random intercept included in the model is Participant (25).
 1103

1104

1105

1106

1107

1108

1109

1110

1111

1112

1113

1114

1115

1116

1117

1118

1119

1120

1121

1122

1123

1124

1125

1126

1127

1128

1129