



City Research Online

City St George's, University of London

Citation: Doidy, F., Desauay, P., Rebillard, C., Clochon, P., Lambrechts, A., Wantzen, P., Guérolé, F., Baleyte, J. M., Eustache, F., Bowler, D. M., et al (2023). How scene encoding affects memory discrimination: Analysing eye movements data using data driven methods. *Visual Cognition*, 31(1), pp. 1-17. doi: 10.1080/13506285.2023.2188335

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/33130/>

Link to published version: <https://doi.org/10.1080/13506285.2023.2188335>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

How scene encoding affects memory discrimination: analyzing eye movements data using data driven methods

Doidy F1, Desaunay P1,2, Rebillard C1, Clochon P1, Lambrechts A3, Wantzen P1, Guérolé F1,2, Baleyte J.M1,4, Eustache F1, Bowler D.M3, Lebreton K1, Guillery-Girard B1*

1 Normandie Univ, UNICAEN, PSL Université Paris, EPHE, INSERM, U1077, CHU de Caen, GIP Cyceron, Neuropsychologie et Imagerie de la Mémoire Humaine, 14000 Caen, France

2 Service de Psychiatrie de l'enfant et de l'adolescent, CHU de Caen, Caen, France

3 Autism Research Group, Department of Psychology, City, University of London, London, United Kingdom

4 Service de Psychiatrie de l'enfant et de l'adolescent, Centre Hospitalier Interuniversitaire de Créteil, Créteil, France

* Correspondence:

Bérengère Guillery-Girard; berengere.guillery-girard@ephe.psl.eu

Unité INSERM – EPHE – Université de Caen – U1077, Pôle des Formations et de Recherche en Santé (PFRS), "Neuropsychologie et Imagerie de la mémoire humaine", 2, rue des Rochambelles, F-14032 Caen Cedex CS 14032

ABSTRACT

The encoding of visual scenes remains an under-explored field due to methodological limitations. In this study, we evaluated the relationship between memory accuracy for a visual scene and eye movements at encoding. First, we used data-driven methods, a fixation density map (using iMap4) and a saliency map (using GBVS), to analyze the visual attention for items. Second, and in a more novel way, we conducted scanpath analyses without *a priori*, using the ScanMatch toolbox. Scene memory accuracy was assessed by asking participants to discriminate identical scenes (targets) among rearranged scenes sharing some items with targets (distractors) and new scenes. Results showed that shorter fixation duration in regions of interest (ROIs) at encoding was associated with a better rejection of distractors. However, there was no significant difference in the relative fixation time in ROIs at encoding, between subsequent hits and misses at test. Hence, density of eye fixations in data-driven ROIs seems to be a marker of subsequent memory discrimination and pattern separation. Interestingly, we also identified a negative correlation between average MultiDimensional Scaling (MDS) distance scanpaths and the correct rejection of distractors, which supports the idea that scanpath consistency significantly affects the ability to discriminate distractors from targets. Taken together, these data suggest that visual exploration at encoding participates in discrimination processes at test. Eye-tracking analyses methods without *a priori* are particularly relevant to study these processes. Future studies could further explore the notion of scanpath consistency to further delineate its characteristics.

Key words: eye tracking, fixation, scanpath, saliency, memory.

1 **INTRODUCTION**

2 The interest in using eye tracking to study memory processes is growing. The majority of
3 studies analyzed the link between encoding and the subsequent recognition without going further into
4 the process of discrimination (Schurgin, 2018). Research on visual scenes adds a layer of difficulty,
5 because of their complexity and diversity. Scenes are consensually defined as a human-scaled view of
6 an environment, associating items arranged in specific locations with larger surfaces and elements that
7 constitute the background, which together form a coherent semantic concept (see Williams &
8 Castelhana, 2019, for a review). During visual exploration, eye movements bring the image of a
9 stimulus to the fovea, and each eye movement to a position is preceded by a shift of attention to that
10 position (Hoffman, 1998). Eye-tracking methodology provides the opportunity to capture these eye
11 movements and therefore to explore overt attention for visual stimuli such as scenes, focusing on
12 fixations, defined as the stationary states of the eyes during which gaze is held upon a specific location
13 in the image, and saccades, that correspond to rapid eye movements between fixations, but provides
14 only limited insights on covert attention (see Vidal et al., 2012, for a review).

15 Encoding of natural scenes depends on both the visual properties of the scene, i.e., saliency,
16 semantic integrity (i.e., semantically congruent vs. incongruent objects within a scene context), spatial
17 associations, and the task set, i.e., search, memorization, or free viewing (see Castelhana & Krzyś,
18 2020, and Williams & Castelhana, 2019, for reviews). Eye-tracking studies on visual scene encoding
19 have so far assessed the link between eye movements and subsequent recognition performance (using
20 old/new paradigms) or awareness (using remember/know paradigms). Some studies identified that
21 the number of fixation points made during scene encoding is associated with greater subsequent
22 recognition performance (Choe et al., 2017), or memory awareness (Kafkas & Montaldi, 2012). In
23 addition, the spatial distribution of clustered eye fixations during encoding has been found to predict
24 subsequent remembering awareness of visual scenes relative to knowing, due to the recognition of
25 distinct details (Sharot et al., 2008). With regards to the duration of fixations, shorter fixation durations
26 during face encoding (considering the whole face as the area of interest) have been found to predict
27 subsequent hits over misses (Parag & Vakil, 2018). Eye fixation can also be guided by visual saliency,
28 which encompasses low-level scene features. The *bottom-up saliency map hypothesis* (Itti & Koch,
29 2001) points to the role of salient low-level features in capturing attention during free-viewing or
30 explicit memory tasks (Underwood & Foulsham, 2006), but not in a search task (Foulsham &
31 Underwood, 2011). Henderson & Hayes (2018) goes beyond this model by showing that both meaning
32 and image saliency predicted the distribution of attention during explicit encoding.

33 To date, no eye-tracking study on scenes encoding has assessed memory discrimination, which
34 differs from memory recognition, on methodological, behavioral, and neural levels. Memory
35 discrimination tasks require participants to recognize previously encoded stimuli among highly similar
36 lures and new stimuli (i.e., old/lure discrimination), and rely on detailed memory representations of
37 the old stimuli, while memory recognition tasks (i.e., old/new discrimination) do not need such
38 detailed representations, and can be achieved at least with gist-based representations of the old
39 stimuli (Lioatile & Courtney, 2015). Memory discrimination and memory recognition tasks explore
40 respectively pattern separation at encoding, and pattern completion at retrieval, which are episodic
41 memory processes involving different hippocampal subfields (Liu et al., 2016). Pattern separation is
42 based on dentate gyrus activity, and allows to assign different neural codes to events with overlapping
43 content, which results in distinct representations stored in memory, and supports subsequent old/lure
44 discrimination (Hainmueller & Bartos, 2020; Rolls, 2016). Pattern completion is based on the activity
45 of the cornus ammonis (CA3), and triggers from partial cues, the retrieval of the whole representation
46 of the encoded stimuli (Liu et al., 2016). Only one eye-tracking study has been conducted on memory
47 discrimination, by Molitor et al. (2014), in which participants at test had to recognize old items (i.e.,
48 pictures of everyday objects) among similar distractors and new different items. Compared to hits,
49 false alarms were associated with fewer fixations to the target at encoding, suggesting that errors were
50 driven by diminished encoding (the *poor encoding hypothesis*).

51 The methodology of more recent research on eye fixations in relation to recognition rates and
52 awareness has been based on fixation maps, which correspond to duration-weighted fixation density.
53 Damiano & Walther (2019) used a paradigm in which observers were asked either to fixate or to
54 explore scenes. During the study phase, exploration, relative to fixation, led to higher subsequent
55 recognition rates, while at test exploration, relative to fixation, led to same hit rates but a higher
56 rejection rate of new scenes. This demonstrates that fixations do not completely reflect visual
57 attention, and what is encoded/recognized. In contrast to these results, Wynn, Ryan, and Buchsbaum's
58 (2020) evaluated the level of similarity between fixation maps at encoding and retrieval (i.e., gaze
59 reinstatement) during a pattern completion paradigm, and found similar gaze reinstatement for hit
60 and false alarm responses at test. Other studies have replicated previous findings on spatial
61 distribution of fixations, finding that a broader exploration during encoding leads to higher recognition
62 performance (Damiano & Walther, 2019), and stronger familiarity awareness (Ramey, Henderson, &
63 Yonelinas, 2020). Looking at the inter-observer congruency of fixation maps, Lyu et al., (2019)
64 identified that scene memory is related to the consistency of fixation maps across viewers at encoding,
65 which is itself related to the signal-to-noise fixation ratio between preferentially-viewed regions of the
66 scene and other regions.

67 Research focusing on scanpaths, i.e., the spatial distributions of eye gaze, were initially guided
68 by *the scanpath theory*, which postulates that scanpath similarity between encoding and retrieval is
69 predictive of memory performance (Noton & Stark, 1971). Subsequent studies have shown that
70 scanpaths are idiosyncratic, showing a greater individual consistency between encoding and retrieval,
71 than between different observers viewing the same image (e.g., Foulsham et al., 2012; French, Glady,
72 & Thibaut, 2017). Similarly to research using fixation maps, more recent research on scanpaths tends
73 to identify specific patterns of eye movement associated with memory formation, and distinct roles
74 for eye movements during encoding and retrieval. Analyzing scanpath rehearsal during study phase,
75 Meghanathan et al. (2019) described different types of refixations (the repetition in time of fixation
76 sequences), which supported memory-encoding strategies in free viewing. Comparing scanpath for
77 novel and repeated scenes during a search task, Wynn et al. (2016) evidenced a scanpath repetition of
78 initial and final but not middle fixations during repeated scene; early scanpath similarity increased
79 search efficiency by reducing search time at test. These results have therefore challenged the *scanpath*
80 *theory*. Arizpe et al. (2019) used faces to propose an alternative model, suggesting that an increasing
81 number of fixations during encoding enables the gradual integration of disparate information into a
82 coherent representation, that can be activated during recognition within a small number of fixations.
83 Scanpath analysis allows to distinguish a population with typical development from participants with
84 neuropsychiatric disorders (see Armstrong & Olatunji, 2012, and Toh, Rossell, & Castle, 2011, for
85 reviews). Notably, during scene encoding and recognition, Shakespeare et al. (2015) have evidenced a
86 greater *consistency* of scanpaths between healthy participants compared to participants with
87 neurological condition, suggesting that healthy participants looked more appropriately at task-
88 relevant regions of the scene.

89 To summarize, both fixation maps and scanpaths eye-tracking studies using scenes suggest
90 that eye-movements during encoding are predictive of recognition rates and awareness at test, but so
91 far no study has explored memory discrimination via pattern separation processes. Hence, the present
92 methodological study aimed to determine what patterns of eye movements during scene encoding are
93 related to memory discrimination for visual scenes, focusing on fixations and scanpaths that predict
94 correct recognition and false alarms. We designed an *old/rearranged/new* paradigm suitable to test
95 memory accuracy and pattern separation processes, including an incidental encoding phase using
96 drawings of scenes, followed by a surprise recognition phase in which participants had to recognize
97 “target” scenes (same items and background), among “distractor” scenes (same items but new
98 background) and new scenes (items and background were new). Incidental encoding of scenes benefits
99 more to memory for visual features than for the whole scene, with images containing many features
100 being richly encoded when compared to other kind of images (Evans & Baddeley, 2018). Drawings and

101 photographs of scenes share large similarities; however, drawings allow a better handling of the
102 content including many details and are more attractive compared to photographs (Park et al., 2019).
103 First, we used data-driven methods, a fixation density map (using iMap4), to evaluate the visual
104 attention for items. We proposed to extend the Molitor et al's. (2014) results to fixation durations and
105 predicted an association between increased false alarm rate and longer fixation durations on details
106 shared between targets and distractors. Such an association would support the presence of an
107 attentional bias at encoding, that would increase interference during the recognition phase. We also
108 generated a saliency map (using GBVS) to ensure that our data-driven fixation maps were not related
109 to low-level features, mainly saliency. Second, and in a more novel way, we conducted scanpath
110 analyses without semantic *a priori* expectations in the definition of ROIs. As memory discrimination
111 relies on detailed memory representations, and presumably on extensive exploration of visual scenes
112 at encoding, without *a priori* analyses, which are based on a high number of similar ROIs, each
113 containing a variable amount of semantic information, seem more suitable than a priori analyses which
114 focus on specific items, for assessing participants' exploration of the whole visual scene. We
115 considered that inter-observer scanpath consistency would be a valuable marker of the quality of the
116 memory trace and predicted greater scanpath consistency at encoding would be associated with
117 higher discrimination at test.

118

119 **METHOD**

120 **1. Participants**

121 Forty-nine healthy young adults were recruited for this study (26 males, mean age 21.3 ± 2.41
122 years, mean level education 13.81 ± 1.86 years). Of these 49 participants, 42 were right-handed
123 (assessed by the De Agostini & Dellatolas checklist, 1988), and 41 had right ocular dominance. All
124 participants had normal or corrected-to-normal vision. None reported recent use of alcohol or illicit
125 drugs, current or past mental disorder, neurological disorder (including history of head trauma with
126 loss of consciousness, or seizures), or current medical condition. An intellectual impairment was ruled
127 out, using the Wechsler Adult Intelligence Scale-IV (Verbal Comprehension Index= 111.60 ± 17 ,
128 Perceptual Reasoning Index= 101.62 ± 12.60). Forty-two participants were retained for analysis (1
129 participant was excluded because he did not look sufficiently at some pictures, i.e., total visit duration
130 $< 3\%$ of total scene duration for one picture, and $< 43\%$ for two pictures) and 6 participants were
131 excluded due to eye tracker dysfunction including calibration. When using Molitor's results as criteria,
132 the required sample size was about 35 participants for $\text{Alpha} = 0.05$ (power goal= 0.90). We decided to

133 increase the number of inclusions to have a minimum of 35 participants for analyses: 42 participants
134 were retained for analyses, we aimed for this large number of participants to compensate for the small
135 number of stimuli.

136 This study was conducted in accordance with the Code of Ethics of the World Medical
137 Association (Declaration of Helsinki 2008), and was approved by the local ethics committee before it
138 started (CPP Nord Ouest III, N° ID RCB : 2013-A01800-45); all participants signed for consent.

139 **2. Material**

140 A total of 43 colored pictures (1024 x 717 pixels) created by a professional cartoonist were
141 used (figure 1): 7 were used for the training phase before study phase, 6 were used to avoid primacy
142 and recency effects at study phase, and the remaining 30 were used in the data analyses. These 30
143 item-background pictures belonged to one of 10 different semantic categories (rural, farm, forest,
144 garden, mountain, snow, car park, swimming pool, beach, and city). Three pictures were created for
145 each semantic category: a “target” picture identical at study and at test phase, a “distractor” picture,
146 and a “new” picture. The “distractor” picture contained the same items in the foreground as the
147 “target” picture combined with a new background from the same semantic category. The foreground
148 items, common to both the target and distractor, were two or three per picture (3 for “forest” and
149 “beach” categories, and 2 for other categories). The “new” picture combined new foreground items
150 and a new background from the same semantic category as the “target”.

151 **3. Procedure**

152 E-Prime software (PST, Pittsburgh, PA) was used to control stimulus presentation and to record
153 participants’ responses. Participants were sitting comfortably 79 cm from the screen in a dimly lit room
154 during the whole experiment. Eye positions and gaze durations were measured at 60 Hz with a remote
155 eye tracker (Tobii X120 A, Tobii Technology AB). A box with infrared sources and a camera were set
156 below a 22-inch TFT flat-screen monitor (HP, x22LED). Stimuli were presented in full screen at a
157 resolution of 1024 x 768 pixels. The size of the projection screen was 47.75 x 26.92 cm², subtended
158 33.6° x 19.4° in visual angle. Before each recording, a 9-points eye-tracker calibration and validation
159 procedure was performed.

160 The incidental memory task included a study phase followed by a surprise recognition phase
161 (figure 2). At study phase, a trial started with a dynamic fixation central white cross presented on a
162 black background for 1280ms. A picture then appeared on the screen for 5000ms, followed by a black

163 screen for 500ms. The 10 “target” pictures were presented in pseudo-random order. Participants were
164 asked to freely and carefully explore each picture for the full duration of the presentation. The test
165 phase was provided after a 20-minute break during which participants performed unrelated tests (i.e.,
166 Rey's Figure test, and a verbal fluency test, not affecting participants' ability to complete the test
167 phase; no fatigue was observed). The test phase started with four familiarization trials comprising of 4
168 pictures (not presented in the study phase), followed by the 30 pictures of interest presented in
169 pseudo-random (10 “target”, 10 “distractor”, and 10 “new” pictures). Participants were instructed to
170 identify “target” pictures and reject other kinds of pictures, i.e., “distractor” and “new” pictures. In
171 both cases, participants responded by pressing one of two keys on a computer mouse. There was no
172 time limit to respond. A trial started by means of a fixation cross, then pictures were presented on the
173 screen until the participant responded, followed by a 500ms black screen as in the study phase.

174 **4. Eye-tracking analyses**

175 Eye tracking recording and analyses were carried out separately for each picture, using data
176 from the study phase only (figure 3).

177 **4.1. Data preprocessing**

178 An I-VT filter (Velocity-Threshold Identification filter: classifier: 30°/s; velocity calculator
179 window length: 20 ms) was applied to the eye-tracking data prior to exporting as recommended by the
180 eye tracker manufacturer. The output was based on the average of both eyes. The data was then
181 exported (Identifier, Scene Name, Gaze Event Duration, Fixation Point X, Fixation Point Y) into Matlab.
182 All data processing and analyses were performed using Matlab (R2015a) software. The data was
183 preprocessed before statistical analyses.

184 As the picture sizes were smaller than the screen size, fixations on the screen but outside of
185 the pictures were removed from all analyses. For the same reason, the gaze point coordinates were
186 recalculated to reflect this shift.

187 For each trial, the first fixation was discarded if it was central, as it was likely to be a reflection
188 (or *remanence*) of the fixation cross presented just before the scene, and therefore was not
189 informative (Hayes & Henderson, 2017; Ramey, Yonelinas, & Henderson, 2020; Wang et al., 2015). As
190 a result, 37 first fixations (8.8 % of all first fixations) which were not to the picture center and thus
191 were informative, were kept for analyses. All the remaining fixations were used for the analyses (see
192 Table 1).

193 **4.2. Fixation analysis**

194 Gaze fixation location data were analyzed using iMap4 (Caldara & Mielle, 2011; Lao et al.,
195 2017; and see also Nicholls et al., 2019, with visual scenes), a freely available MATLAB open source
196 toolbox for the statistical fixation mapping of eye movement data. iMap4 requires a two-step process:
197 first, iMap4 generates fixation duration maps for each participant and every picture, creating individual
198 fixation maps which are smoothed by convoluting them with a two-dimensional Gaussian Kernel
199 function. These individual maps can be averaged to produce a group fixation map (heatmap). Second,
200 these 3D fixation maps (dimensions are x, y, and fixation duration) are compared to produce
201 statistically significant duration-weighted difference maps. The main advantage of this method is that
202 it is data-driven, inspired by methods used in functional Magnetic Resonance Imaging, and no prior
203 segmentation of pictures into ROIs is required.

204 Toolbox parameters were adapted for the scene stimuli. In order to clearly identify the items
205 embedded in the scenes, the smoothing parameter was set to a 2.5-degree visual angle. To reduce
206 memory usage and computational time, picture size was rescaled. We used the default scaler
207 parameter of 0.24 (updated X size = 246 pixels, updated Y size = 173). An average “duration map” was
208 computed for each picture. A one-tailed t-test was then performed against the “baseline activation”
209 (“the mean fixation intensity within the iMap mask”). We used an alpha level of .01 Bonferroni-
210 corrected for multiple comparisons. The output provided, for each scene, a map showing the
211 significant above-chance fixation duration regions. From these maps, we extracted statistical iMap
212 Regions of Interest (data-driven ROIs), where each data-driven ROI corresponded to statistically
213 significant cluster of voxels. These data-driven ROIs were binarized. To simulate the central foveal
214 vision, binary ROIs were dilated by 25 pixels, which corresponded to the size, on the image, of half of
215 a viewing angle of 1.5°.

216 Using these parameters, we calculated the relative fixation duration spent in data-driven ROIs
217 at study phase for 4 subsequent response categories at test: hits (correctly recognized “targets”),
218 misses (unrecognized “targets”), correct rejections (correctly rejected “distractors”), and false alarms
219 (“distractors” incorrectly identified as “targets”). The relative fixation duration in ROIs was equal to
220 the sum of fixation durations spent in ROI divided by the sum of all fixation durations spent in the
221 picture. A Mann Whitney test was used to compare relative time spent in ROIs at study phase for each
222 category.

223 **4.3. Visual saliency analysis**

224 Visual saliency map

225 For each picture, a saliency map was created using the GBVS Matlab toolbox (Graph-Based
226 Visual Saliency, Harel, Koch, & Perona, 2007). GBVS was selected over other saliency toolboxes because
227 it offers the highest prediction level for saliency-based ocular fixations, especially for scenes (Borji,
228 Sihite, & Itti, 2013). The GBVS process is divided into two stages: activation maps are first created using
229 specific feature channels, which are then normalizing in a way that highlights conspicuity and admits
230 combination with other maps, to generate a final saliency map. We used the default parameters of
231 GBVS.

232 Fixation duration map

233 For each participant and for each image, an individual fixation duration map was constructed,
234 by taking the set of locations where the eyes were fixated and the duration of the gaze point. A fixation
235 duration map was computed for each image by adding up the fixation durations of all participants
236 (O'Connell & Walther, 2015). To simulate the central foveal vision, the result was then convolved with
237 a Gaussian kernel. The full width at half maximum of the Gaussian kernel was set to 1.5° of visual angle.

238 Correlation between saliency map and fixation duration map

239 For each picture, a Spearman's rank correlation was used to measure the similarity between
240 saliency map and fixation duration map (Riche et al., 2013).

241 Correlation between visual saliency and fixation ROI

242 A Spearman's rank correlation was used to test the relation between saliency and data-driven
243 ROI fixation duration found using iMap. In the saliency interaction analysis, pixel-level saliency for each
244 ROI was selected as the maximum value of the object region in order to minimize the object size effect.
245 This was because big objects tend to include uniform texture regions and thus have much smaller
246 average pixel-level saliency, while fixations were normally attracted to the most salient region of an
247 object. Thus, maximum saliency rather than average saliency was more representative of pixel-level
248 saliency of an object (Wang et al., 2015).

249 **4.4. Scanpath analysis**

250 The similarity between participants' sequences of eye movements was analyzed using the
251 ScanMatch open source Matlab toolbox (Cristino et al., 2010). This choice was guided by Anderson et

252 al. (2015) study which compared scanpaths methods and concluded that ScanMatch is the most
253 adapted tool for analyzing fixation sequences, since it can take into account spatial location, temporal
254 duration, and sequential similarity between scanpaths.

255 For the spatial binning, each picture was divided with a grid composed of 96 (12×8)
256 rectangular ROIs of 85 x 89 pixels. This grid size was determined without a priori: we used the same
257 size bin as Cristino et al. (2010). Each rectangular ROI was labeled by a combination of two letters. Each
258 fixation within the ROI was tagged with its name in the string sequence: a sequence contains a list of
259 visited ROIs, respecting the order of fixations for a specific picture. For the temporal binning, the letters
260 corresponding to a ROI were repeated in the sequence proportionally to the fixation duration: string
261 sequences were divided into 100ms bins.

262 Then, the ScanMatch string-edit distance methodology was used to find the best alignment
263 over the whole string of two sequences by maximizing its score. This methodology was based on the
264 Needleman-Wunsch algorithm which has been implemented to compare DNA sequence. To this end,
265 the sequences were aligned based on a substitution matrix which provides a score for every alignment,
266 based on the spatial relationship between ROIs. A similarity score of 1 corresponds to sequences being
267 identical, while a similarity score of 0 indicates that there is no similarity.

268 We used a gap penalty of 0 which means that adding gaps decreased similarity scores (Frame,
269 Warren, & Maresca, 2019). The “threshold value” was equal to 2 standard deviations of all the saccade
270 amplitudes. This means that the alignment algorithm only aimed to align regions which were within
271 the variability of the saccade amplitudes (Cristino et al., 2010).

272 The Scanmatch algorithm was used to calculate the similarity score between participant pairs,
273 resulting in between-participant ScanMatch score matrix (matrix size: 42 x 42 participants).

274 ScanMatch statistical analysis

275 This matrix was transformed with Matlab function “mdscale”, allowing to perform
276 MultiDimensional Scaling (MDS) (Kruskal, 1964) and to project each participant in a n-dimensional
277 space. MDS then computed the participant coordinates in this n-dimensional space.

278 Beforehand, it was necessary to determine the number of dimensions of this space. The
279 goodness-of-fit of such space was quantified using a residual sum of squares called the stress of the
280 map. Stress values are positive and small values are better, e.g., a stress value of 10 % indicates a fair
281 fit (Kruskal, 1964). The stress value depends on the space dimension: the greater the n-dimensional

282 space, the smaller the stress value. We chose the size of the n-dimensional space so that the stress
283 value of all the scenes was strictly less than 7.5 %, which guaranteed a correct fit.

284 At the end of this procedure, MDS provided relative locations for all participants. Participants
285 with similar temporo-spatial gaze patterns were plotted close together near the center of the space
286 while those with atypical gaze patterns were plotted towards the periphery. The median location of
287 the group, i.e., the MDS median point, was defined as the center of the participants' group and
288 represented the standard temporo-spatial gaze pattern. The relative location of participants was then
289 computed by calculating the Euclidean distance (i.e., L2-norm) between the respective MDS
290 coordinates participants and MDS median point. The mean MDS-distance of the 42 participants was
291 calculated for each picture based on eye gaze patterns at study phase; higher MDS-distance values
292 reflected dissimilar temporo-spatial gaze patterns.

293 To test the relation between dissimilarity at study phase and the subsequent recognition
294 performance at test, we conducted Spearman correlations between the mean MDS-distance of 42
295 participants and accuracy scores, i.e., the number of targets correctly recognized and distractors
296 correctly rejected.

297

298 **RESULTS**

299 **1. Behavioral performance**

300 During the test phase, participants correctly recognized 79.52 ±13.96 % of target scenes,
301 correctly rejected 70.24 ±15.85 % of distractor scenes and correctly rejected 97.86 ±4.15 % of new
302 scenes (figure 4). A Friedman's test (3 conditions: target, distractor, new) revealed a main effect of
303 condition ($\chi^2_F(2) = 64.78, p < 0.001$). A Wilcoxon signed-rank test indicated that all of these results were
304 significantly different from each other: Target (*Mdn*= 80%) vs Distractor (*Mdn*= 70%): $T = 129, z = 2.71,$
305 $p < 0.01$; Target vs New (*Mdn*= 100%): $T = 5.5, z = 5.07, p < 0.001$; Distractor vs New: $T = 0, z = 5.65, p <$
306 0.001 .

307 **2. Fixation duration in iMap data-driven ROIs**

308 A total of 20 data-driven ROIs were identified, at study phase, in the 10 target scenes (table 2,
309 figure 5).

310 There was no significant difference in the relative fixation time in data-driven ROIs at study
311 phase, between subsequent correctly recognized targets (Hits) ($Mdn= 30.22\%$) and missed targets
312 (Misses) at test ($Mdn= 31.55\%$), ($U(N_{Hits}= 334, N_{Misses}= 86)= 14099, z= 0.40, p= 0.80, \text{Mann Whitney}$) (figure
313 6). By contrast, we found a significant difference in the relative fixation time in data-driven ROIs at
314 study phase, between false recognitions (false alarms) and correct rejection of distractors, with longer
315 fixation durations for the former ($Mdn= 38.46\%$) compared to correct rejections ($Mdn= 29.98\%$),
316 ($U(N_{Correct Rejections}= 295, N_{False Alarms}= 125)= 14191, z= 3.73, p< 0.001, \text{Mann Whitney}$) (figure 6). This means that
317 the more time participants spent looking in data-driven ROIs in the study phase, the less successful
318 they were at rejecting “distractor” pictures in the test phase.

319 **3. Visual saliency**

320 We found a significant positive correlation between visual saliency and fixation duration for
321 whole pictures at study phase (Spearman rho range: [0.526; 0.773], $p< 0.001, N= 1024 \times 717$ pixels)
322 (figure 7). By contrast, there was no significant correlation between the visual saliency of data-driven
323 ROIs and the relative fixation duration time in these ROIs (Spearman rho= 0.332, $N= 840$), which
324 confirmed that fixations were not driven by low-level scene features, i.e., saliency.

325 **4. ScanMatch results**

326 We chose a size of 8 for the n-dimensional space, for which all the stress values of images were
327 strictly less than 7.5% (figure 8). The stress values, found and used to determine the goodness of fit of
328 an MDS solution, are in accordance with the literature. For instance, in an eye-tracking study consisting
329 of sentence analysis in 44 students (von der Malsburg & Vasishth, 2011), the stress value was found to
330 be 22 % for a 2-dimensional space and decreased to 8.1 % in a 7-dimensional space.

331 Using an 8-dimensional space, we determined the mean MDS-distance of 42 subjects for each
332 picture seen during the study phase (L2-norm mean, table 3), and calculated the Spearman correlation
333 coefficients between recognition performances (Hits for “targets” and Correct Rejection of
334 “distractors”) and this average distance.

335 There was no significant correlation between the proportion of correctly recognized targets
336 and average MDS-distance (Spearman rho= -0.092, $p= 0.80, N= 10$) (figure 9). By contrast, we identified
337 a significant negative correlation between the proportion of correctly rejected distractors and average
338 MDS-distance (Spearman rho= -0.659, $p= 0.04, N= 10$) (figure 9). This result indicates that the more

339 similar the picture exploration was between subjects in the study phase, the more correctly they
340 rejected “distractor” pictures at test.

341

342 **DISCUSSION**

343 This study aimed to provide a better understanding of how scene exploration strategies during
344 free viewing at study phase, is associated with subsequent memory discrimination between targets,
345 distractors, and new scenes, at test phase. Eye movement data were analyzed at study phase using
346 two data-driven methods, namely a fixation density map (using iMap4) associated with a saliency map
347 (using GBVS), and scanpath analyses without *a priori* (using the ScanMatch toolbox). First, we identified
348 longer fixation durations in data-driven ROIs for subsequent false alarms over rejection of distractors,
349 while there was no significant difference in the relative fixation time in data-driven ROIs between
350 subsequent hits and misses, indicating that fixation maps were more an indicator of memory
351 discrimination accuracy than of recognition performance. Second, we identified a negative correlation
352 between average MDS-distance scanpaths and the correct rejection of distractors, and no significant
353 correlation between average MDS-distance and target recognition performance, which suggests that
354 scanpath consistency (or inter-observers congruency of scanpaths) at study phase was a factor of
355 subsequent memory discrimination abilities, rather than of memory performance. Taken together,
356 eye-tracking can provide insights into the pattern separation process, suggesting that a broad and
357 consistent exploration during encoding increases subsequent memory discrimination.

358 First, longer fixation durations in data-driven ROIs (using iMap4) were associated with
359 subsequent false alarms over rejection of distractors. As fixation locations are a marker of visual
360 attention (Vidal et al., 2012), this result implies that the more the participants focused their attention
361 on these ROIs at encoding, the less they explored the rest of the visual scene, to be able subsequently
362 to reject distractors. Using single items, Molitor et al.'s (2014) have previously suggested that low
363 abilities in memory discrimination – i.e., false alarms – results from insufficient number of fixations at
364 study – i.e., the *poor encoding hypothesis*. The current results go further, by showing that the relative
365 fixation duration in data-driven ROIs is a marker of subsequent memory discrimination. A high relative
366 fixation duration in some ROIs implies a low relative fixation duration outside these ROIs that accounts
367 for a reduced global exploration and is predictive of false alarms. Memory discrimination requires high-
368 resolution mnemonic representations of studied items supported by the *pattern separation* process,
369 and the intra-hippocampal binding of features that constitute an event (Hunsaker & Kesner, 2013).
370 This suggests that encoding details with high specificity to be subsequently discriminated from other

371 similar lures requires a broad exploration of scenes during encoding. Interestingly, the relative fixation
372 time in data-driven ROIs at study phase did not differ for subsequent hits and misses, indicating that
373 fixation time is not coupled with recognition performance. In line with this finding, several studies have
374 shown that recognition performance is better when fixation times are shorter (see Choe et al., 2017,
375 in a search task, and Parag & Vakil, 2018, with faces), or that they are independent parameters
376 (Schomaker & Wittmann, 2017).

377 Second, the positive correlation between the visual saliency and fixation duration for the full
378 picture highlights the role of visual saliency on full scene exploration, as previously demonstrated in
379 landscape photographs (Dupont et al., 2016). Interestingly, the absence of significant correlations
380 between the visual saliency of data-driven ROIs and the fixation duration time in these ROIs confirmed
381 that data-driven fixation maps were not related to low-level scene features, i.e., saliency.

382 Third, scanpath analyses revealed no significant correlation between the average MDS-
383 distance and target recognition performance which appeared to be independent factors. This result is
384 consistent with other studies focusing on scanpaths and scene recognition, which showed that
385 performance is related to scanpath idiosyncrasy (i.e., greater within- than between-participants
386 similarity when compared at encoding and at retrieval, Foulsham et al., 2012), or that recognition
387 performance is relatively dissociable from scanpath analyses (Foulsham & Kingstone, 2013). By
388 contrast, we identified a negative correlation between the average MDS-distance and the correct
389 rejection of distractors, which implies that scanpath consistency across observers during scene
390 exploration at study phase is a significant factor in the ability to discriminate distractors from targets
391 at test. This hypothesis is congruent with Lyu et al.'s (2019) study which identified that the *consistency*
392 *of fixation maps* across viewers was predictive of scene memory. Few studies have highlighted
393 scanpath consistency, using various paradigms. Shakespeare et al. (2015) observed higher scanpath
394 consistency between healthy participants compared to neurological patients in search and memory
395 tasks with scenes, although not considering the order or duration of fixations in their analyses, and
396 suggested a *disorganized approach* when patients viewed scenes. Using a Deep Neural Networks, Wei
397 et al. (2017) developed a method to predict which features will capture the most attention in a visual
398 scene (i.e., *gaze agreement*), and the most consistent scanpath across viewers (i.e., *scanpath*
399 *agreement*). More recently, Frame, Warren, & Maresca (2019), using dynamic surveillance videos and
400 a guided search task, showed that an effective search strategy was associated with consistent
401 scanpaths across observers. Together, our results suggest that there may be an optimal scene
402 exploration strategy during free viewing at study phase to encode foreground and background scene
403 features with high specificity, leading to a better memory discrimination at test phase.

404 Although this study provides interesting results, there are some limitations that open up
405 opportunities for further research. First, to further study visual exploration strategies during scene
406 encoding, we need to evaluate how the visualization of items of different sizes may influence fixation
407 density maps and behavioral performance. Because participants viewed large items at different points
408 in space, iMap4 did not detect their fixations which led us to use a 2.5° smoothing. In addition,
409 statistical data-driven ROIs comparisons are more difficult to realize, and their interpretation is mainly
410 visual. Second, we used a small number of images, which may limit the statistical power to assess
411 memory discrimination. In addition, using same items in both target and distractor supported a certain
412 degree of visual similarity necessary to assess memory discrimination, but did not allow to control
413 whether the proportion of eye movements in these items versus the background influenced the
414 rejection of the distractors. Third, we suggested that the consistency of scanpath at scene encoding
415 affects the subsequent identification of “distractors” at test. The next step would be to identify scene
416 characteristics that may facilitate scanpath consistency and increase memorization. This would open
417 up new possibilities in terms of care and support for people with memory difficulties.

418

419 **CONCLUSIONS**

420 To conclude, eye-tracking analyses methods without *a priori* are particularly suitable to study encoding
421 in memory for visual scenes, given their size and the amount of information they contain. In particular,
422 we showed that iMap4 and ScanMatch toolboxes are valuable methods to study gaze movements
423 associated with scene memory discrimination. The current study contributes to our growing
424 understanding of memory strategies at encoding. Most importantly, we identified that scene memory
425 discrimination is associated with a wide gaze exploration at study phase, and a consistent scanpath.
426 Future studies may further explore the inter-observer scanpath consistency and their relationship to
427 memory and cognitive parameters.

428

ACKNOWLEDGEMENTS:

429 Authors are grateful to L. Raoult, C. Piet and L. Legentil for their help with recruitment, cognitive testing
430 and eye-tracking examination. We thank M. Lalevée and J. Spiess for the creation of drawings. We also
431 thank F. Wallois, C. Mille and J.M. Guilé as well as the clinical team who participated into inclusions at
432 the Amiens hospital.

433

DISCLOSURE STATEMENT

434 No potential conflict of interest was reported by the author(s).

435 **FUNDING**

436 This study was supported by the Normandy Regional Council and the French Clinical Research Hospital
437 Program.

438

439 **REFERENCES**

- 440 Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath
441 comparison methods. *Behavior Research Methods*, *47*(4), 1377–1392.
442 <https://doi.org/10.3758/s13428-014-0550-3>
- 443 Arizpe, J. M., Noles, D. L., Tsao, J. W., & Chan, A. W. Y. (2019). Eye Movement Dynamics Differ
444 between Encoding and Recognition of Faces. *Vision*, *3*(1), 9.
445 <https://doi.org/10.3390/vision3010009>
- 446 Armstrong, Thomas and Olatunji, B. O. (2012). Analytic Review and Synthesis. In *Clin Psychol Rev* (Vol.
447 32). <https://doi.org/10.1016/j.cpr.2012.09.004>.Eye
- 448 Borji, A., Sihite, D. N., & Itti, L. (2013). What stands out in a scene? A study of human explicit saliency
449 judgment. *Vision Research*, *91*, 62–77. <https://doi.org/10.1016/j.visres.2013.07.016>
- 450 Caldara, R., & Miellet, S. (2011). iMap: a novel method for statistical fixation mapping of eye
451 movement data. *Behavior Research Methods*, *43*(3), 864–878. [https://doi.org/10.3758/s13428-](https://doi.org/10.3758/s13428-011-0092-x)
452 [011-0092-x](https://doi.org/10.3758/s13428-011-0092-x)
- 453 Castelhana, M. S., & Krzyś, K. (2020). Rethinking Space: A Review of Perception, Attention, and
454 Memory in Scene Processing. *Annual Review of Vision Science*, *6*(1), 563–586.
455 <https://doi.org/10.1146/annurev-vision-121219-081745>
- 456 Choe, K. W., Kardan, O., Kotabe, H. P., Henderson, J. M., & Berman, M. G. (2017). To search or to like:
457 Mapping fixations to differentiate two forms of incidental scene memory. *Journal of Vision*,
458 *17*(12), 8. <https://doi.org/10.1167/17.12.8>
- 459 Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for
460 comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692–700.
461 <https://doi.org/10.3758/BRM.42.3.692>
- 462 Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and
463 retrieval. *Cognition*, *184*(December 2018), 119–129.
464 <https://doi.org/10.1016/j.cognition.2018.12.014>
- 465 De Agostini, M., & Dellatolas, G. (1988). Une épreuve simple pour évaluer la préférence manuelle
466 chez l'enfant à partir de 3 ans. *Enfance*, *41*(3), 139–147.
467 <https://doi.org/10.3406/enfan.1988.2161>
- 468 Declaration of Helsinki. (2008, octobre 22). (n.d.). *WMA Declaration of Helsinki—Ethical principles*
469 *for medical research involving human subjects*. [https://doi.org/https://www.wma.net/what-we-](https://doi.org/https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/)
470 [do/medical-ethics/declaration-of-helsinki/](https://doi.org/https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/)
- 471 Dupont, L., Ooms, K., Antrop, M., & Van Eetvelde, V. (2016). Comparing saliency maps and eye-
472 tracking focus maps: The potential use in visual impact assessment based on landscape
473 photographs. *Landscape and Urban Planning*, *148*, 17–26.
474 <https://doi.org/10.1016/j.landurbplan.2015.12.007>
- 475 Evans, K. K., & Baddeley, A. (2018). Intention, attention and long-term memory for visual scenes: It all
476 depends on the scenes. *Cognition*, *180*(February), 24–37.
477 <https://doi.org/10.1016/j.cognition.2018.06.022>
- 478 Foulsham, T., Dewhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., & Holmqvist, K.
479 (2012). Comparing scanpaths during scene encoding and recognition: A multi-dimensional
480 approach. *Journal of Eye Movement Research*, *5*(4), 3. <https://doi.org/10.16910/jemr.5.4.3>
- 481 Foulsham, T., & Kingstone, A. (2013). Fixation-dependent memory for natural scenes: An
482 experimental test of scanpath theory. *Journal of Experimental Psychology: General*, *142*(1), 41–
483 56. <https://doi.org/10.1037/a0028227>

- 484 Foulsham, T., & Underwood, G. (2011). If Visual Saliency Predicts Search, Then Why? Evidence from
485 Normal and Gaze-Contingent Search Tasks in Natural Scenes. *Cognitive Computation*, 3(1), 48–
486 63. <https://doi.org/10.1007/s12559-010-9069-9>
- 487 Frame, M. E., Warren, R., & Maresca, A. M. (2019). Scanpath comparisons for complex visual search
488 in a naturalistic environment. *Behavior Research Methods*, 51(3), 1454–1470.
489 <https://doi.org/10.3758/s13428-018-1154-0>
- 490 French, R. M., Glady, Y., & Thibaut, J.-P. (2017). An evaluation of scanpath-comparison and machine-
491 learning classification algorithms used to study the dynamics of analogy making. *Behavior*
492 *Research Methods*, 49(4), 1291–1302. <https://doi.org/10.3758/s13428-016-0788-z>
- 493 Hainmueller, T., & Bartos, M. (2020). Dentate gyrus circuits for encoding, retrieval and discrimination
494 of episodic memories. *Nature Reviews Neuroscience*, 21(3), 153–168.
495 <https://doi.org/10.1038/s41583-019-0260-z>
- 496 Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information*
497 *Processing Systems*, 545–552.
- 498 Hayes, T. R., & Henderson, J. M. (2017). Scan patterns during real-world scene viewing predict
499 individual differences in cognitive capacity. *Journal of Vision*, 17(5), 23.
500 <https://doi.org/10.1167/17.5.23>
- 501 Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images:
502 Evidence from eye movements and meaning maps. *Journal of Vision*, 18(6), 10.
503 <https://doi.org/10.1167/18.6.10>
- 504 Hoffman, J. (1998). Visual attention and eye movements. In H. Pashler (Ed.), *Attention* (pp. 119–153).
505 Retrieved from <https://psycnet.apa.org/record/1998-07791-003>
- 506 Hunsaker, M. R., & Kesner, R. P. (2013). The operation of pattern separation and pattern completion
507 processes associated with different attributes or domains of memory. *Neuroscience &*
508 *Biobehavioral Reviews*, 37(1), 36–58. <https://doi.org/10.1016/j.neubiorev.2012.09.014>
- 509 Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*,
510 2(3), 194–203. <https://doi.org/10.1038/35058500>
- 511 Kafkas, A., & Montaldi, D. (2012). Familiarity and recollection produce distinct eye movement, pupil
512 and medial temporal lobe responses when memory strength is matched. *Neuropsychologia*,
513 50(13), 3080–3093. <https://doi.org/10.1016/j.neuropsychologia.2012.08.001>
- 514 Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric
515 hypothesis. *Psychometrika*, 29(1), 1–27. <https://doi.org/10.1007/BF02289565>
- 516 Lao, J., Mielle, S., Pernet, C., Sokhn, N., & Caldara, R. (2017). iMap4: An open source toolbox for the
517 statistical fixation mapping of eye movement data with linear mixed modeling. *Behavior*
518 *Research Methods*, 49(2), 559–575. <https://doi.org/10.3758/s13428-016-0737-x>
- 519 Liu, K. Y., Gould, R. L., Coulson, M. C., Ward, E. V., & Howard, R. J. (2016). Tests of pattern separation
520 and pattern completion in humans—A systematic review. *Hippocampus*, 26(6), 705–717.
521 <https://doi.org/10.1002/hipo.22561>
- 522 Loiotile, R. E., & Courtney, S. M. (2015). A signal detection theory analysis of behavioral pattern
523 separation paradigms. *Learning & Memory*, 22(8), 364–369.
524 <https://doi.org/10.1101/lm.038141.115>
- 525 Lyu, M., Choe, K. W., Kardan, O., Kotabe, H. P., Henderson, J. M., & Berman, M. G. (2019). Scenes that
526 produce more consistent fixation maps are more memorable. *PsyArXiv Preprints*. Retrieved
527 from <https://pdfs.semanticscholar.org/50b7/6aa18ed9ba66e0cca3d5f630307981b2512d.pdf>
- 528 Meghanathan, R. N., Nikolaev, A. R., & van Leeuwen, C. (2019). Refixation patterns reveal memory-
529 encoding strategies in free viewing. *Attention, Perception, & Psychophysics*, 81(7), 2499–2516.

- 530 <https://doi.org/10.3758/s13414-019-01735-2>
- 531 Molitor, R. J., Ko, P. C., Hussey, E. P., & Ally, B. A. (2014). Memory-related eye movements challenge
532 behavioral measures of pattern completion and pattern separation. *Hippocampus*, *24*(6), 666–
533 672. <https://doi.org/10.1002/hipo.22256>
- 534 Nicholls, V. I., Jean-Charles, G., Lao, J., de Lissa, P., Caldara, R., & Mielle, S. (2019). Developing
535 attentional control in naturalistic dynamic road crossing situations. *Scientific Reports*, *9*(1),
536 4176. <https://doi.org/10.1038/s41598-019-39737-7>
- 537 Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing
538 patterns. *Vision Research*, *11*(9), 929–938. [https://doi.org/10.1016/0042-6989\(71\)90213-6](https://doi.org/10.1016/0042-6989(71)90213-6)
- 539 O’Connell, T. P., & Walther, D. B. (2015). Dissociation of salience-driven and content-driven spatial
540 attention to scene category with predictive decoding of gaze patterns. *Journal of Vision*, *15*(5),
541 20. <https://doi.org/10.1167/15.5.20>
- 542 Parag, O., & Vakil, E. (2018). Distinct eye movements for different cognitive processes as expressed in
543 the face recognition task. *Memory*, *26*(4), 524–534.
544 <https://doi.org/10.1080/09658211.2017.1381265>
- 545 Park, J., Jin, Y., Ahn, S., & Lee, S. (2019). The Impact of Design Representation on Visual Perception:
546 Comparing Eye-Tracking Data of Architectural Scenes Between Photography and Line Drawing.
547 *Archives of Design Research*, *32*(1), 5–29. <https://doi.org/10.15187/adr.2019.02.32.1.5>
- 548 Ramey, M. M., Henderson, J. M., & Yonelinas, A. P. (2020). The spatial distribution of attention
549 predicts familiarity strength during encoding and retrieval. *Journal of Experimental Psychology:*
550 *General*, *149*(11), 2046–2062. <https://doi.org/10.1037/xge0000758>
- 551 Ramey, M. M., Yonelinas, A. P., & Henderson, J. M. (2020). Why do we retrace our visual steps?
552 Semantic and episodic memory in gaze reinstatement. *Learning & Memory*, *27*(7), 275–283.
553 <https://doi.org/10.1101/lm.051227.119>
- 554 Riche, N., Duvinage, M., Mancas, M., Gosselin, B., & Dutoit, T. (2013). Saliency and Human Fixations:
555 State-of-the-Art and Study of Comparison Metrics. *2013 IEEE International Conference on*
556 *Computer Vision*, 1153–1160. <https://doi.org/10.1109/ICCV.2013.147>
- 557 Rolls, E. T. (2016). Pattern separation, completion, and categorisation in the hippocampus and
558 neocortex. *Neurobiology of Learning and Memory*, *129*, 4–28.
559 <https://doi.org/10.1016/j.nlm.2015.07.008>
- 560 Schomaker, J., & Wittmann, B. C. (2017). Memory Performance for Everyday Motivational and
561 Neutral Objects Is Dissociable from Attention. *Frontiers in Behavioral Neuroscience*, *11*(June), 1–
562 13. <https://doi.org/10.3389/fnbeh.2017.00121>
- 563 Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working
564 memory and long-term memory. *Attention, Perception, & Psychophysics*, *80*(5), 1035–1056.
565 <https://doi.org/10.3758/s13414-018-1522-y>
- 566 Shakespeare, T. J., Pertzov, Y., Yong, K. X. X., Nicholas, J., & Crutch, S. J. (2015). Reduced modulation
567 of scanpaths in response to task demands in posterior cortical atrophy. *Neuropsychologia*, *68*,
568 190–200. <https://doi.org/10.1016/j.neuropsychologia.2015.01.020>
- 569 Sharot, T., Davidson, M. L., Carson, M. M., & Phelps, E. A. (2008). Eye Movements Predict Recollective
570 Experience. *PLoS ONE*, *3*(8), e2884. <https://doi.org/10.1371/journal.pone.0002884>
- 571 Toh, W. L., Rossell, S. L., & Castle, D. J. (2011). Current visual scanpath research: a review of
572 investigations into the psychotic, anxiety, and mood disorders. *Comprehensive Psychiatry*,
573 *52*(6), 567–579. <https://doi.org/10.1016/j.comppsy.2010.12.005>
- 574 Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruity influence eye
575 movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, *59*(11),

576 1931–1949. <https://doi.org/10.1080/17470210500416342>

577 Vidal, M., Turner, J., Bulling, A., & Gellersen, H. (2012). Wearable eye tracking for mental health
578 monitoring. *Computer Communications*, 35(11), 1306–1311.
579 <https://doi.org/10.1016/j.comcom.2011.11.002>

580 von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis?
581 *Journal of Memory and Language*, 65(2), 109–127. <https://doi.org/10.1016/j.jml.2011.02.004>

582 Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., & Zhao, Q. (2015).
583 Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye
584 Tracking. *Neuron*, 88(3), 604–616. <https://doi.org/10.1016/j.neuron.2015.09.042>

585 Wei, Z., Adeli, H., Hoai, M., Zelinsky, G., & Samaras, D. (2017). Predicting Scanpath Agreement during
586 Scene Viewing using Deep Neural Networks. *Journal of Vision*, 17(10), 749.
587 <https://doi.org/10.1167/17.10.749>

588 Williams, C. C., & Castelhana, M. S. (2019). The Changing Landscape: High-Level Influences on Eye
589 Movement Guidance in Scenes. *Vision*, 3(3), 33. <https://doi.org/10.3390/vision3030033>

590 Wynn, J. S., Bone, M. B., Dragan, M. C., Hoffman, K. L., Buchsbaum, B. R., & Ryan, J. D. (2016).
591 Selective scanpath repetition during memory-guided visual search. *Visual Cognition*, 24(1), 15–
592 37. <https://doi.org/10.1080/13506285.2016.1175531>

593 Wynn, J. S., Ryan, J. D., & Buchsbaum, B. R. (2020). Eye movements support behavioral pattern
594 completion. *Proceedings of the National Academy of Sciences*, 117(11), 6246–6254.
595 <https://doi.org/10.1073/pnas.1917586117>

596

597

598

599
600

Tables

	N° Picture										Total
	1	2	3	4	5	6	7	8	9	10	
Total nb of fixations retained	652	708	670	644	626	656	680	658	653	664	6611
Nb of 1 st fixation retained	4	3	3	4	4	2	4	5	2	6	37

601 **Table 1** Total number of fixations and number of first fixations retained for analysis for each target
602 picture.

603
604

	N° picture										Total
	1	2	3	4	5	6	7	8	9	10	
Number of ROI	3	2	2	1	2	1	2	2	3	2	20

605 **Table 2** Number of data-driven Regions Of Interest identified in each target picture during study phase.

606

N° picture	% of target pictures correctly recognized	% of distractor pictures correctly rejected	L2-norm mean
1	78.57	52.38	0.3619
2	69.05	64.29	0.40743
3	73.81	85.71	0.35782
4	80.95	78.57	0.35135
5	88.10	40.48	0.40763
6	85.71	83.33	0.34745
7	80.95	69.05	0.37707
8	88.10	83.33	0.38137
9	61.90	59.52	0.39328
10	88.10	85.71	0.36134
<i>Total</i>	<i>79.52</i>	<i>70.24</i>	

607 **Table 3** For each target picture, % of participants who recognized target correctly, rejected distractor
608 correctly, and L2-norm mean.

609

610

611 **Figures**

612
613

614 **Figure 1.** Scenes presented to participants during study and recognition phases (*only scenes used for*
615 *the eye-tracking data analysis are presented*).

616
617

618 **Figure 2.** Experimental design. In the study phase (left), participants were presented with pictures
619 and were asked to watch carefully. In the recognition phase (right), participants were presented with
620 pictures again and were asked to decide whether they had seen them in the study phase or not.
621 Pictures in the recognition phase were presented until the participant responded.

622
623

624 **Figure 3.** Schematic representation of the procedures for eye-tracking data analysis.

625 *Note: in order to represent the MDS space in the figure, the size of the n-dimensional space was*
626 *chosen equal to 3 (but in the analysis, 8 was used).*

627
628

629 **Figure 4.** Recognition accuracy: percentage of pictures correctly recognized (targets) or correctly
630 rejected (distractors and new) in recognition phase (** $p < 0.01$, *** $p < 0.001$).

631 *Note: The red line indicates the median, the bottom and top edges of the box indicate the 25th and*
632 *75th percentiles, respectively. The whiskers extend to the most extreme data points not considered*
633 *outliers, and the outliers are plotted individually using the '+' symbol*

634
635

636 **Figure 5.** Data-driven ROIs ($p < 0.01$ Bonferroni corrected, significant area marked by dark line) and
637 visual saliency (obtained using GBVS) for scene viewing during study phase.

638
639

640 **Figure 6.** Relative fixation duration in study phase corresponding to subsequent target hits or misses
641 (A), correct rejections and false alarms (B) in recognition phase (*** $p < 0.001$).

642 *Note: The red line indicates the median, the bottom and top edges of the box indicate the 25th and*
643 *75th percentiles, respectively. The whiskers extend to the most extreme data points not considered*
644 *outliers, and the outliers are plotted individually using the '+' symbol*

645

646

647 **Figure 7.** Spearman's rank correlation between visual saliency and fixation duration, during study
648 phase, in whole picture for each target picture. The numbers above each bar correspond to Spearman's
649 rank coefficient.

650

651

652 **Figure 8.** Stress value and dimensional-space for the 10 scenes; 0.075 corresponds to the maximum
653 accepted value for the stress value.

654

655

656 **Figure 9.** Relationship between average MDS-distance and percentage of: targets correctly
657 recognized (A), distractors correctly rejected (B) (see table 3).