False memories when viewing overlapping scenes

Filip Dechterenko [Corresp., 1] , **Jiri Lukavsky** [1]

[1] Institute of Psychology, Czech Academy of Sciences, Prague, Czech Republic

Corresponding Author: Filip Dechterenko
Email address: dechterenko@praha.psu.cas.cz

Humans can memorize and later recognize many objects and complex scenes. In this study, we prepared large photographs and presented participants with only partial views to test the fidelity of their memories. The unpresented parts of the photographs were used as a source of distractors with similar semantic and perceptual information. Additionally, we presented overlapping views to determine whether the second presentation provided a memory advantage for later recognition tests. Experiment 1 (N=28) showed that while people were good at recognizing presented content and identifying new foils, they showed a remarkable level of uncertainty about foils selected from the unseen parts of presented photographs (false alarm, 59%). The recognition accuracy was higher for the parts that were shown twice, irrespective of whether the same identical photograph was viewed twice or whether two photographs with overlapping content were observed. In Experiment 2 (N=28), the memorability of the large image was estimated by a pre-trained deep neural network. Neither the recognition accuracy for an image part nor the tendency for false alarms correlated with the memorability. Finally, in Experiment 3 (N=21), we repeated the experiment while measuring eye movements. Fixations were biased toward the center of the original large photograph in the first presentation, and this bias was repeated during the second presentation in both identical and overlapping views. Altogether, our experiments show that people recognize parts of remembered photographs, but they find it difficult to reject foils from unseen parts, suggesting that their memory representation is not sufficiently detailed to rule them out as distractors.

1 **False memories when viewing overlapping scenes**

5 Filip Děchtěrenko[1] and Jiří Lukavský[1]

7 [1] Institute of Psychology, Czech Academy of Sciences

9 Corresponding Author:

10 Filip Děchtěrenko[1]

11 Hybernská 8, 110 00, Prague, Czech Republic

12 Email address: dechterenko@praha.psu.cas.cz

Abstract

Humans can memorize and later recognize many objects and complex scenes. In this study, we prepared large photographs and presented participants with only partial views to test the fidelity of their memories. The unpresented parts of the photographs were used as a source of distractors with similar semantic and perceptual information. Additionally, we presented overlapping views to determine whether the second presentation provided a memory advantage for later recognition tests. Experiment 1 (N=28) showed that while people were good at recognizing presented content and identifying new foils, they showed a remarkable level of uncertainty about foils selected from the unseen parts of presented photographs (false alarm, 59%). The recognition accuracy was higher for the parts that were shown twice, irrespective of whether the same identical photograph was viewed twice or whether two photographs with overlapping content were observed. In Experiment 2 (N=28), the memorability of the large image was estimated by a pre-trained deep neural network. Neither the recognition accuracy for an image part nor the tendency for false alarms correlated with the memorability. Finally, in Experiment 3 (N=21), we repeated the experiment while measuring eye movements. Fixations were biased toward the center of the original large photograph in the first presentation, and this bias was repeated during the second presentation in both identical and overlapping views. Altogether, our experiments show that people recognize parts of remembered photographs, but they find it difficult to reject foils from unseen parts, suggesting that their memory representation is not sufficiently detailed to rule them out as distractors.

Introduction

Every day, we are surrounded by a rich and complex visual world. Previous research has shown that we are able to memorize a large amount of visual information (Standing, 1973). In recent years, these findings have been replicated for objects (Brady, Konkle, Alvarez, & Oliva, 2008) and scenes (Vogt & Magnussen, 2007; Konkle, Brady, Alvarez, & Oliva, 2010), indicating a massive memory capacity for complex stimuli. These experiments typically use a similar procedure, in which a large number of stimuli are presented in the study part and then identical views (or foils) are later queried in the testing part.

Although a high visual capacity was observed in many studies, one might naturally ask how detailed the memories about each item are. This level of detail is usually termed fidelity. In a memory study with thousands of objects (Brady et al., 2008), participants correctly recognized target objects when presented with similar foils, which differed only in their state. Moreover, in the free recall paradigm using drawings, people show remarkable memory of objects and their spatial location (Bainbridge, Hall, and Baker, 2019).

This high-fidelity representation is challenged by findings from Vogt and Magnussen (2007). In their study, they presented participants with 400 pictures of doors, and they observed a high recognition accuracy of approximately 85%. In a follow-up manipulation, they removed all of the surrounding details of the doors, and the recognition accuracy decreased by 20%.

The observed high recognition performance might overestimate the true capacity due to the type of experimental probing. In general, participants are able to more easily provide a correct

54 response in a two-alternative forced choice (2AFC) task than in a Yes/No task. Studies directly
55 comparing both methods (Cunningham, Yassa, & Egeth, 2015; Andermane, & Bowers, 2015)
56 reported comparable results. However, Cunningham et al. (2015) concluded that the high
57 performance in the original study by Brady et al. (2010) might also be explained by gist-based
58 representations, as the same set of subjects showed significantly lower recognition performance
59 in the Yes/No task than in the 2AFC task. Thus, the representation of the studied item might be
60 noisy, and the performance in the memory studies might result from gist-based representations
61 rather than storing detailed information about the stimuli. A similar finding of a rather low-level
62 detailed representation was observed in the incidental encoding paradigm (Draschkow,
63 Reinecke, Cunningham, & Võ, 2018), in which participants' memory was tested using the
64 Yes/No task. Again, the accuracy was lower than that in the original study by Brady et al.
65 (2008). The current literature thus provides conflicting evidence regarding what people actually
66 remember when memorizing a larger set of stimuli.
67 One way to approach the question of what we remember is to use more ecologically valid
68 stimuli. The typical stimuli used in visual memory experiments are individual objects (e.g.,
69 Brady, Konkle, Alvarez, & Oliva, 2008) or photographs of scenes (e.g., Konkle, et al., 2010).
70 However, in our daily experience, we are surrounded by a continuous stream of visual
71 information. We can remember objects or specific views, but the explicit boundaries given by a
72 photograph's frame are missing. Herein, we distinguish between *scenes* (real-world
73 combinations of objects) and *photographs* (providing views of scenes). We are interested in
74 overlapping photographs that share some common content. The usage of overlapping
75 photographs allows us to study repeated exposure and assess how individual photographs are
76 integrated into coherent memory. Although people are able to recognize previously presented
77 viewpoints (Hock & Schmelzkopf, 1980), they do not integrate the viewpoint with information
78 about viewed objects within the scenes (Varakin & Loschky, 2010). During the memorization of
79 overlapping photographs, participants usually look more at the previously viewed parts rather
80 than exploring novel, previously unseen parts (Valuch, Becker, & Ansorge, 2013). We can view
81 the distinction between scenes and photographs as a part of a hierarchy and take it one step
82 further by asking about smaller parts of the photographs (here called patches) and how well
83 they are remembered when the photographs are repeated.
84 When memorizing a photograph, people usually have trouble remembering the view boundary.
85 The problem with remembering view boundaries is prominent in boundary extension (BE)
86 studies (Intraub & Richardson, 1989), where people incorrectly claim they remember seeing
87 objects beyond the border of the photograph. According to the source-monitoring account on BE
88 (Intraub, 2012), people automatically perform additional processes when seeing a photograph –
89 they amodally complete missing parts of objects or surfaces, categorize the scene and activate
90 contextual associations. The observed error of commission is a result of confusing actual
91 memories with these additional inferences. Due to the design differences, the actual BE effect is
92 not likely relevant for our observations. However, we share a common idea that people try to
93 position the observed content (patches in this case) into a larger context (representation of the
94 photograph). Similarly, people try to position the observed photographs into larger scenes they
95 imagine (as in BE).Making assumptions about the scene beyond the view is not limited to BE
96 paradigm. In particular, people can make judgments about more distant views, e.g., whether two
97 views are from the same panoramic view (Robertson, Hermann, Mynick, Kravitz, & Kanwisher,

98   2016), but their performance is limited unless extra cues are provided and both views overlap.
99   Thus, participants are sensitive to scene layout to a limited extent (as shown by Sanocki &
100   Epstein, 1997, and others).
101   In the present study, we studied visual memory of scenes from different perspectives. As
102   previous studies showed conflicting evidence regarding what people actually remember of each
103   studied photograph, we worked with larger photographs and presented participants with only
104   partial views of these photographs. Thus, we subsequently presented an alternative view
105   partially overlapping with the original view and tested whether this second presentation provided
106   a memory advantage for later recognition tests. The unseen parts of the photograph were also a
107   suitable source of distractors, as they showed the same semantic category and usually similar
108   texture (in the case of natural scenes) or scene-related objects (of bar stools in restaurants).
109   This method provides a stricter measure of human recognition memory and allowed us to select
110   foils similar to the presented stimuli, which might be difficult in the case of whole scenes.
111   Because this approach may increase the number of false alarm errors, we decided to use the
112   Yes/No paradigm in all experiments.
113   In Experiment 1, we studied whether participants were able to recognize parts of the presented
114   photographs. We were interested in whether the repeated presentation of the whole photograph
115   would lead to increased recognition of smaller parts, similar to the findings in the study by
116   Konkle et al. (2010). Additionally, querying participants with lures selected from either unseen
117   parts or presented parts allowed us to assess the level of detail of recognition performance
118   ("This area looks like part of the kitchen I saw" vs "This area is the particular part of the kitchen I
119   saw").
120   Our paradigm allowed us to better determine whether the performance in recognition tasks is
121   more a result of vague resemblance of the memorized photograph or more a result of the
122   detailed memory of a particular photograph. The main purpose of Experiment 1 was to explore
123   the general performance in recognizing stimuli by their smaller parts using estimates from signal
124   detection theory framework. This general paradigm was further extended in two follow-up
125   experiments. In Experiment 2, we focused on whether the recognition accuracy is explained by
126   the properties of individual images, namely, their memorability. We estimated the average
127   recognition accuracy (memorability) for each image using a deep neural network, and we tested
128   to what extent participants recognized parts of highly memorable images or were lured by
129   similar foils. Finally, in Experiment 3, we performed eye-tracking analysis and assessed how the
130   repeated content was studied. We tested whether participants preferred to rescan the area they
131   had viewed during the first presentation or explore the new areas and how the dwell times on
132   parts of an image correlated with the actual recognition accuracy.

Experiment 1

134   Previous research has shown that people are able to recognize previously presented
135   photographs with high accuracy (e.g., Standing, 1973; Konkle, et al, 2010). However, to
136   correctly recognize a photograph or reject previously unseen photographs, the content of the
137   photograph needs to stand out from the remaining presented stimuli either semantically or
138   perceptually (Lukavský & Děchtěrenko, 2017; Standing, 1973; Konkle, et al, 2010; Bylinskii,
139   Isola, Bainbridge, Torralba, & Oliva, 2015). Can we recognize a presented photograph based on

140   only a part of the image? One would naturally suspect that memory accuracy would be lower for
141   smaller parts, as memory for whole photographs can benefit from memorizing multiple parts.
142   However, even when recognizing smaller parts of presented photographs, people may still have
143   only vague feelings of familiarity. Therefore, are we able to correctly reject parts of photographs
144   that have not been previously presented? If so, this may suggest that the memorized
145   photographs are stored in great detail, while in the case of high false alarms, the stored
146   representation may actually have minimal details.
147   The main goal of Exp. 1 was to assess the human memory of parts of photographs, which we
148   approached in two ways. First, we tested whether people could learn from a larger photograph
149   and respond about its parts. We presented participants with photographs, but in the recognition
150   phase, we used small patches from the presented photographs instead of asking about the
151   whole presented photographs. Participants were tested after every five trials (immediate
152   recognition) and after presentation of the complete set of photographs (delayed recognition).
153   Second, we tested whether people were able to learn more about the smaller parts when they
154   encountered the larger photograph twice, either in the identical form or in the overlapping views.
155   In the overlapping views, only half of the scene content was presented in the second viewing,
156   and new content was exposed.
157   This first experiment was designed to assess general performance on this paradigm and later
158   elaborated in Experiments 2 and 3. All experiments were approved by the Ethics Committee of
159   the Institute of Psychology, Czech Academy of Sciences (approval number PSU-
160   203/Brno/2020).

161 Materials & Methods

162 Participants

163   Twenty-eight subjects (mean age = 21.54 years, SD = 2.35 years; 7 males) participated in the
164   experiment. All participants had normal or corrected-to-normal vision, and none had previously
165   participated in this type of experiment. The subjects provided written consent before the start of
166   the experiment.

167 Stimuli and Apparatus

168   For the experiment, we obtained pairs of photographs that overlapped in 50% of their content.
169   Briefly, we selected a set of photographs and split each of them into two overlapping views (the
170   scheme is visualized in Fig. 1). The photographs were obtained from the SUN database (Xiao,
171   Hays, Ehinger, Oliva, & Torralba, 2010). We selected 11 categories to contain both natural and
172   man-made scenes (indoor: candy store, bar, shopping mall, and atrium; outdoor: basilica, cabin,
173   and slum; natural: creek, forest, canyon, and mountain). As the SUN database contains many
174   photographs with various formats, we selected only landscape-oriented photographs larger than
175   900×600 px. This resulted in a dataset with 1183 photographs.
176
177   These photographs were resized to 900×600 px while maintaining the original aspect ratio. If
178   one dimension was larger than the specified size, we cropped the remaining area and visually

179 inspected the photograph for possible cropping artifacts. The resulting photograph was divided
180 into thirds of 300×600 px each, and the training stimuli consisted of two adjacent thirds
181 (600×600 px). For the creation of test stimuli, we extracted smaller patches from each
182 photograph. The patches were 256×256 px (43% × 43% of the stimulus or 18% of the area)
183 located either in the upper or lower part of the photograph (randomly selected). From each
184 photograph of 900×600 px, we created all six smaller patches (upper/lower × left/central/right
185 patch). Each participant was presented with a random set of both training stimuli (on average 6
186 photographs per category, SD = 3.78) and corresponding test patches.
187
188
189 --- Insert Figure 1 here. ---
190
191 The experiment was programmed in PsychoPy (Peirce, 2007) and presented on a 22" LCD
192 screen (1920 x 1080 px). The participant sat approximately 50 cm from the screen, and the
193 head position was not controlled.

194 Design

195 The experiment consisted of two parts: the study part and the testing part. In the study part,
196 participants were presented with the photographs and tested for immediate recognition. In the
197 testing part, participants were tested for the recognition of photographs from the whole study
198 part. We manipulated two factors in this study: level of repetition of the photograph during the
199 study part and the selection of the smaller patch with respect to the corresponding learned
200 photograph.
201 Regarding the level of repetition, the presented photographs were classified into one of three
202 types (see Fig. 2). First, *identical trials* (2 x 55 images) featured the identical photograph
203 presented twice per participant. Second, *overlapping trials* (2 x 55 images) presented two
204 different views from the same original photograph that shared the middle portion (left view
205 followed by right view or vice versa). Third, *non-repeated trials* (110 images) contained unique
206 photographs with no repetition.
207 Smaller patches were selected either from the presented parts of photographs, unseen parts of
208 presented photographs or from completely novel photographs. Different patch conditions were
209 present in immediate and delayed recognition.
210 For immediate recognition, the testing trials featured 66 patches of three conditions (22 trials per
211 condition). The patch was either from the presented part of the non-repeated photograph (*old*),
212 from an unpresented part of the non-repeated photograph (*lure*), or from a novel, unseen
213 photograph (*new*). Each of the three patch types was presented 22 times.
214 For delayed recognition, the testing trials featured 200 patches of four possible conditions (50
215 trials per condition). Similar to immediate recognition, patches from parts of images that
216 participants saw once (*old-1)*, patches from the unseen parts of the presented stimuli (*lure*) and
217 patches from novel, unseen images (*new)* were presented. We also presented patches from the
218 parts of images that participants saw twice (*old-2)*. The number of patches per condition and
219 their relationship to the presentation trial conditions are summarized in Table 1. For all patches,
220 we evenly distributed whether the patch was selected from the upper or lower part of the original
221 image.

222
223   --- Insert Figure 2 here. ---
224

Procedure

226   As shown in Figure 3, the study part consisted of 396 trials: 330 presentation trials in which a
227   photograph (600×600 px) was shown in the center of the screen for 3 s for participants to study
228   and 66 testing trials. One testing trial was presented after each of the 5 presentation trials, and
229   the task was to decide whether the presented patch was part of one of the preceding five
230   images (with unlimited time). Participants were instructed to memorize the presented
231   photographs as accurately as possible, and they were informed that they would be presented
232   with smaller patches and that their task would be to respond, regardless of whether they saw
233   the exact patch within the photograph. Participants were informed that some patches may look
234   similar to previously presented photographs but that they should really respond only if that patch
235   was exactly the one that was presented previously (we did not tell them that some patches
236   would be from the unseen parts of the photographs). Participants were also aware that in the
237   study part, patches would be related only to the 5 previously seen photographs, while they
238   would later be queried for the whole set. Additionally, they were informed that we were
239   interested in recognition accuracy, not in the speed of the response (although they were asked
240   not to spend too much time on each trial if they were unsure).
241   The order of trials was randomized for each participant with the constraint that each block of 5
242   trials contained at least one non-repeated trial.
243
244
245   --- Insert Figure 3 here. ---
246
247   In each testing trial, we showed a single patch, and participants were asked whether they had
248   seen this exact patch during the study part. The patch was shown in the center of the screen,
249   not at its original position during the presentation. Participants responded with either the left key
250   (patch was shown) or right key (patch was not shown), while verbal cues (words *old* and *new*)
251   were provided below each image to reduce the possibility of response mistakes.
252   Although we asked participants to prioritize accuracy over response time, we analyzed the
253   response time to detect the potential speed versus accuracy trade-off.
254
255   --- Insert Table 1 here. ---

Data Analysis

257   All materials and scripts are available at the Open Science Framework (https://osf.io/469er/).
258   Data were analyzed with the R software (R Core Team, 2019) using generalized linear mixed
259   models with a binomial link function (lme4 package, Bates, Maechler, Bolker, & Walke, 2015),
260   i.e., hierarchical logistic regression. We tested the accuracy of different patch types separately
261   in immediate and delayed recognition. Both situations were tested with a model containing two
262   random factors (varying intercepts for both participants and images), one fixed factor as the

263   independent variable (patch type) and the participant's response (correct yes/no) as the
264   dependent variable. For immediate recognition, the patch type had three levels (new, lure, and
265   old), while four levels were used for delayed recognition (new, lure, old-1, and old-2).
266   Models were compared with a null model containing only random factors using the likelihood
267   ratio test. To further compare the relative difference in false alarms between the lure and the
268   new patches, we reran the regressions on the subset of data containing the new and lure
269   patches only and expressed the relative difference between conditions using the odds ratio.
270   Similar approach was performed in case of delayed recognition to express the difference in
271   odds ratios between old-2 and old-1 patches. The differences from the chance level were tested
272   using one-sample t-tests. Finally, for delayed recognition, we tested the differences between
273   presentation trial types using a two-sample t-test for each of the patch types (e.g., for old-2
274   patches, we tested the differences between repeated and overlapping trials).
275   We also analyzed the data with the signal detection theory (SDT) approach to assess the
276   changes in sensitivity between immediate and delayed recognition (Green & Swets, 1968). We
277   used the additivity property of the sensitivity measure (d') to measure the perceptual distance
278   between stimuli separately for immediate and delayed recognition. We first computed z-scores
279   for responses of type "seen" using the approach described by Macmillan & Creelman (2004) to
280   compute d'. The distances between the z-scores defined the sensitivity, and the bias was set to
281   zero. We computed d' and bias on the responses for novel patches (corresponding to false
282   alarms) and old patches (corresponding to hits) to determine the differences in sensitivity and
283   bias between the immediate test and delayed test. We corrected the hit rate and false alarm
284   rate by adding 0.5 to all cells to avoid perfect performance (which we observed in three
285   participants) as suggested by Macmillan & Creelman (2004). Changes in d'/bias were measured
286   using a paired t-test. The expected decrease in sensitivity between immediate and delayed
287   recognition corresponded to decreased memory capabilities over time. On the other hand,
288   theoretical changes in bias would represent changes in response pattern; in particular, lower
289   bias values correspond to more false alarm responses.

290   # Results

291   For immediate recognition, participants performed well in identifying new patches they had not
292   seen previously (accuracy 81% or 19% false alarm rate). The accuracy for old patches (73%)
293   was lower. Participants struggled when responding to the lure patches from areas they had not
294   seen (accuracy 41% or 59% false alarm rate). The model treating patches differently had a
295   significantly better fit than the null model with random factors only ($\chi^2(2) = 247.37$; $p < .001$).
296   Furthermore, participants scored significantly greater than chance (chance level: 50%, $p < .001$)
297   under all conditions. When comparing the conditions directly, participants were 6.6 times more
298   likely (95% CI [5.1, 8.7]) to score a false alarm for lure patches than for new patches.
299   For delayed recognition, the accuracies were lower: 69% for old-2 patches, followed by new
300   (59%) and old-1 (58%). The FA rate for lure patches was 49%. The differences between patch
301   types were significant ($\chi^2(3) = 91.09$; $p < .001$). Participants scored significantly greater than
302   chance for new, old-1, and old-2 patches ($p \le .003$), but the performance was not significantly
303   different from random guesses for lure patches ($p = .639$). Pairwise comparisons revealed that
304   participants were 1.4 times more likely to record a false alarm (95% CI [1.2, 1.7]) for lure

305 patches than for new patches. Similarly, participants were 1.6 times more likely to correctly
306 recognize the target when the patch was presented twice (95% CI [1.4, 1.9]). Because each of
307 the test patches was selected from two types of trials during the presentation part in the delayed
308 recognition (see Table 1), we assessed differences in recognition accuracy and the FA rate
309 under these conditions. None of the differences were significant ($ps ≥ .293$), indicating no effect
310 on accuracy (e.g., when old-2 patches were selected from the identical vs the central part of
311 overlapping trials) or false alarms (when lures were selected from unseen parts of identical trials
312 vs non-repeated trials). The results of both parts are depicted in Fig. 4.
313
314 -- Insert Figure 4 here. –
315
316 Based on the SDT framework, the perceptual distance in immediate recognition between old
317 and new patches was 1.52, that for old and lure patches was 0.40 and that for lure and new
318 patches was 1.12. The distance between lure and old patches is also known as pseudo d'
319 (Dosher, 1984). For delayed recognition, the perceptual distances were smaller (old-2–old-1: d'
320 = 0.28, old-1–lure: d' = 0.25, lure–new: d' = 0.20).
321 The differences in performance calculated as d' were significantly decreased between
322 immediate recognition (mean d' = 1.55, SD = 0.6) and delayed recognition (mean d' = 0.44, SD
323 = 0.19, $t(27) = 10.51$, $p < .001$). Regarding the response pattern, the bias was not different
324 between immediate (mean bias = 0.14, SD = 0.33) and delayed recognition (mean bias = 0.01,
325 SD = 0.33, $t(27) = 1.58$, $p = .125$). Thus, participants did not systematically prefer any of the
326 responses (old/new).
327 The correlation between accuracy and reaction time was not significant for immediate
328 recognition and delayed recognition ($p ≥ .446$).
329

330 # Discussion

331 Two main findings emerge from the results. First, although participants exhibited high
332 recognition accuracy for presented patches and correctly rejected novel patches, they also
333 recorded a high false alarm rate when presented with a patch from the hidden part of a
334 presented photograph. This finding shows a remarkable level of uncertainty about these
335 distractors. Second, the recognition accuracy was higher for patches that were shown twice, but
336 the accuracy was not significantly different when the same photograph was seen twice or when
337 two photographs with overlapping content were viewed. Similarly, we observed a similar false
338 alarm rate for lures irrespective of whether they were selected from the unseen parts of identical
339 or non-repeated trials.
340 We obtained similar results by analyzing the percentage of correct responses and using the
341 SDT approach. As expected, participants' sensitivity decreased after the delay. Interestingly, we
342 did not observe significant differences in bias after the delay. The results show that participants'
343 implicit preference for either old or new responses was balanced and did not change over time.

Experiment 2

Experiment 1 showed that although participants were able to recognize small areas from the presented stimuli, they were also prone to false alarms when they were queried with unseen parts of photographs. This suggests that participants appeared to base their decision only on rough semantics and perceptual similarity to the presented photographs. In research concerning visual memory, it is generally observed that some photographs are more recognizable than others. This recognition score is commonly referred to as memorability (Isola, Xiao, Torralba, & Oliva, 2011). Memorable photographs have higher perceptual organization (Goetschalckx, Moors, Vanmarcke, & Wagemans, 2018) and cannot be predicted by simple visual features such as color (Isola, et al., 2011) or spatial frequency (Bainbridge, Dilks, & Oliva, 2017). An unresolved question is to what extent memorable photographs consist of memorable parts and regions (Khosla, Xiao, Torralba, & Oliva, 2012). In the context of our experimental design, are memorable photographs more recognizable by their parts and correctly rejected in the case of unseen parts? To answer these questions, in Experiment 2, we studied how the capability to recognize the whole scene corresponds to the recognition of individual patches. We utilized the same protocol (and the same photographs) for all participants, which allowed us to compute the average recognition score for each presented patch (memorability score) and average false alarm rate for unseen parts (lureability score). The idea of using the lureability score was derived from previous findings that people show consistency in false alarms as well (Bainbridge & Rissman, 2018).

Although the optimal approach would be to obtain memorability and "lureability" scores for whole photographs and compare them with patch data, this experiment would require a separate large memory study, as our selected dataset did not contain behavioral memorability scores. Instead, we estimated the memorability of the whole photographs with a convolutional neural network (Khosla, Raju, Torralba, & Oliva, 2015).

We proposed several hypotheses about the relationship between the memorability of whole photographs and their parts. According to previous research, memorable photographs (Isola, Xiao, Torralba, & Oliva, 2011) should be easier to remember because their content differs from that of images from similar categories. Therefore, we expected that the hit rate and correct rejections should be high for old and new patches of highly memorable images (Bylinskii et al., 2015). For the lure patches, two competing hypotheses exist. First, highly memorable images are memorable because of their content; therefore, the false alarm rate (incorrectly marking images as old) should be low (accuracy should be high). Alternatively, highly memorable images are memorable because of their style or gist. Consequently, the accuracy should be low because they would be prone to false alarms caused by distractors from unseen parts, which extends the presented content.

Materials & Methods

Participants

Twenty-eight subjects (mean age = 21.39 years, SD = 3.46 years; 4 males) participated in Experiment 2. All participants had normal or corrected-to-normal vision, and none had

384   participated in Experiment 1 or a similar experiment previously. Again, all subjects provided
385   written consent before the start of the experiment.

Stimuli and Apparatus

387   The apparatus was identical to that used in Experiment 1. The stimuli were similar to those used
388   in Experiment 1: the only difference was that all participants were presented with an identical set
389   of photographs.

Procedure and Data Analysis

391   The procedure was identical to that in Experiment 1. We computed Fleiss' κ to measure the
392   agreement of participants' responses. Memorability was estimated using the pretrained MemNet
393   net (Khosla, Raju, Torralba, & Oliva, 2015). We estimated memorability for both the left and
394   right two-thirds of the original photographs, and because the difference in estimated
395   memorability between both parts was small (mean difference = 0, SD = .04), we used the
396   average value of those two scores as the memorability score for the whole photograph. For
397   each patch, we computed the average hit rate (memorability, for old and new patches) or false
398   alarm rate (lureability, for lure patches). We correlated the model-estimated memorability for
399   whole photographs using MemNet with the average patch memorability/lureability.

Results

401   The results for both immediate and delayed recognition were similar to those obtained in
402   Experiment 1 (Fig. 5). The participants showed significant agreement in accuracy (immediate
403   recognition: Fleiss' κ = .19, z = 30.1, $p < .001$; delayed recognition: Fleiss' κ = .14, z = 38.1, $p <$
404   .001). We observed variability in the average recognition/false alarm rate between images (see
405   Fig. 6)[1].
406
407   -- Insert Figure 5 here. –
408
409   -- Insert Figure 6 here. –
410
411   The model-estimated memorability ranged from 0.48 to 0.83 (mean = 0.66, SD = 0.07). When
412   averaged per each category, highest memorability was observed for candy store (mean = 0.79,
413   SD = 0.05), while mountains showed lowest memorability (mean = 0.60, SD = 0.06). Averaged
414   model-estimated memorability values per category were similar to the behavioral memorability
415   values described in Bylinskii et al. (2015).
416   For immediate recognition, we observed a positive (though nonsignificant) relationship between
417   the model-estimated memorability of the photograph and correct recognition of the patch ($r =$
418   .40, $p = .064$) and a negligible correlation between memorability and correct rejections in the
419   case of new distractors ($r = .06$, $p = .285$) when aggregating responses per patch. In the case of

---

[1] We do not show similar plots for test patches as the use of four patch types made the plot
difficult to read.

420 lureability (average false alarm rate), the correlation with model-estimated memorability was
421 nonsignificant for all patch types (|rs| ≤ 0.2; ps ≥ .164). Taken together, model-estimated
422 memorability for whole scenes did not explain the recognition performance for small patches[2].

Discussion

424 Experiment 2 confirmed the results of Experiment 1. A visual inspection of the patches that
425 showed high consistency in memory performance revealed that participants probably
426 recognized the photographs using either highly memorable objects in the photograph or the
427 semantic uniqueness of the photograph ("there was only one white mountain"). We estimated
428 the memorability using a pretrained neural network and assessed whether this model-estimated
429 memorability of the whole photograph affected the memory performance for parts (patches). In
430 our experiment, the model-estimated memorability of the whole photograph was not a good
431 predictor of remembering its part (all correlations were small, with the exception of immediate
432 recognition). The lureability or tendency for false alarm errors was not explained by the model-
433 estimated memorability of the whole photograph. Consequently, we were not able to distinguish
434 between our competing hypotheses about memorable photographs. Our main finding is that the
435 model-estimated memorability of the whole photographs does not correspond to the recognition
436 of the respective parts. Our main limitation is the usage of model-estimated memorability
437 instead of using empirical estimates. The model was trained to estimate memorability as the hit
438 rate, and it did not naturally capture the average similarity in false alarm rates.

Experiment 3

440 The previous two experiments consistently showed that people have difficulty with correctly
441 recognizing photographs by their smaller parts. We found that repeated presentations of the
442 stimuli lead to increased recognition accuracy, but we found no benefit of showing overlapping
443 content instead of showing whole photographs twice. Thus, we aimed to investigate how people
444 study the repeated content and measured their eye movements. The natural question that we
445 aimed to address was whether participants utilize this experimental manipulation in their
446 strategy (and observed recognition accuracy). Do they spend time fixating on the shared
447 content (potentially merging the representations together) or do they tend to devote more time
448 exploring the novel content? This strategy may also be reflected in their recognition
449 performance, as Olejarczyk, Luke, and Henderson (2014) observed that participants showed
450 higher recognition accuracy during incidental encoding of the parts that they fixated on longer.
451 Similarly, Valuch, Becker, and Ansorge (2013) found that people fixate more often on the parts
452 they have fixated on in previous presentations.

---

[2] Similar estimates were observed for the Spearman correlation coefficient.

Materials & Methods

Participants

Twenty-one subjects (all female, mean age= 21.86 years, SD = 3.05 years) participated in this experiment. None had participated in the previous experiments, and all participants had normal or corrected-to-normal vision.

Stimuli and Apparatus

Experiment 3 was programmed using custom scripts in MATLAB with the Psychtoolbox extension (Brainard, 1997; Cornelissen, Peters, & Palmer, 2002; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Trials were presented on a 22'' LCD display with a resolution of 1680×1050 and a 60 Hz refresh rate. Eye movements were measured using Eyelink II (SR Research, Canada). Participants were seated 50 cm from the monitor, and the head was positioned on a chin rest. We used the same stimuli and protocols as in Experiment 1 (each participant was presented with random selection of stimuli). In this setting, the images subtended 19.4° × 19.4°, and the test patches subtended 8.3 × 8.3°.

Procedure and Data Analysis

The eye tracker was calibrated using 9-point calibration before each part of the experiment, and drift correction was performed before each trial. Otherwise, the procedure was identical to that used in Experiments 1 and 2. All stimuli (both whole photographs and smaller patches) were presented in the center of the screen.

Fixation and saccades were detected using the algorithm supplied with the eye tracker. Our main goal was to analyze whether participants fixated more on the overlapping parts of the photographs than the nonoverlapping parts. Additionally, we report simple descriptive statistics of eye movements (fixation duration and saccade amplitude) for overall picture.

Simple descriptive statistics. Our main dependent variables were the fixation duration and saccade amplitude. We retained all tracking data, irrespective of whether participants fixated on the presented stimuli[3]. We analyzed eye-tracking data from the presentation trials (identical, overlapping, and unique). We did not analyze the eye-tracking data in the test trials due to differences in the presentation times.

We used two linear mixed models with either saccade amplitude and fixation duration as dependent variables and the type of trial (identical, overlapping, or unique) and first/second presentation as fixed factors to determine the significance of differences in the fixation duration and saccade amplitudes between trial types and repeated presentations. Participants and stimuli were treated as random factors, which allowed us to fit varying intercepts for each participant and photograph. We compared the significance of differences in both parameters using a model comparison approach (as in Experiment 1), but the continuous nature of the dependent variable (and thus using a linear mixed model, instead of a generalized model)

---

[3]We repeated the analysis for the data with the fixation outside the presented stimuli discarded, but the results were similar.

489   allowed us to report the results using F tests with Satterthwaite's method (Giesbrecht & Burns,
490   1985; Kuznetsova, Brockhoff, & Christensen, 2017), which is more accessible to the reader
491   given the similarity with nonhierarchical statistical methods. We obtained similar results using
492   both methods.
493   **Overlapping content**. In this analysis, our main dependent variable was dwell time. We did not
494   define areas of interest based on the locations of the patches for two reasons. First, the patch
495   areas were small; thus, the results would be more prone to spatial measurement error during
496   eye tracking. Second, the patches were selected from fixed coordinates within the photograph
497   without any notion of the semantic content. Thus, binning of the fixation data into these
498   particular areas would be arbitrary, and conclusions would be difficult to infer. Consequently, we
499   computed the dwell times separately for each third of the image only.
500   We analyzed the dwell times with respect to the first and second presentations to assess
501   whether participants fixated more on portions of the images that were already presented or on
502   novel content. For clarity, the meanings of novel and overlapping content reversed the eye-
503   tracking data for the stimuli displaying the right two-thirds of original image as they were
504   measured in the left-two thirds[4]. Using this reversal, we treated the data as we always showed
505   first for the left two-thirds of the original image, while in the second presentation of overlapping
506   trials, we are showing the right two-thirds. We denote the thirds of the original image as *lateral*
507   (left third)*, central*, and *contralateral* (right third) to make our results more salient. All identical
508   and unique trials and the first presentation of overlapping trials thus displayed the lateral and
509   central thirds, while the second presentation of overlapping trials displayed the central and
510   contralateral thirds.
511   When participants viewed the photographs for the first time (irrespective of whether the stimuli
512   were identical, overlapping, or unique), we expected null differences in dwell times between
513   halves (central third vs lateral third). When viewing photographs for the second time,
514   participants might employ different viewing strategies in the case of identical and overlapping
515   trials. When viewing identical photographs for the second time, do participants use the same
516   strategy (e.g., because one half contains more informative content) or do the dwell times
517   change (e.g., because they want to explore the parts they fixated on less during the first
518   presentation)? For overlapping trials, do participants fixate the central third more, as it shares
519   content with the previous presentation, or rather do they spent more time on the novel
520   (contralateral) third? We used linear mixed models with two fixed factors denoting whether
521   fixation was occurring in the central or lateral/contralateral part and whether it was the first or
522   second presentation to test these hypotheses. We used dwell time as our outcome variable and
523   participants as a random factor. We used all data from all three types of trials (as they always
524   showed lateral and central thirds) to test the hypothesis regarding the first presentation, while
525   we separately analyzed data for identical (as lateral and central thirds were presented in both
526   presentations) and overlapping trials (as the second presentation showed central and
527   contralateral thirds) to test the hypotheses regarding the dwell time pattern in the second
528   presentation.

---

[4] Without this approach, we would need to report the analysis separately for trials showing the left
two-thirds and trials showing right two-thirds during the first presentation. We obtained similar results from
this separate analysis.

529 Finally, we also correlated the averaged patch recognition accuracy with the dwell time in image
530 third from which the patch was selected. Because Experiments 1 and 2 showed low memory
531 accuracy in delayed recognition, we used the recognition accuracy from the immediate
532 recognition condition only.
533

Results

Accuracy

536 The memory accuracy in immediate recognition showed a similar trend as that observed in
537 Experiments 1 and 2 (immediate recognition: lure: 41% or 59% FA, old: 78%, new: 83%).

Fixation Duration and Saccade Amplitude

539 The basic descriptive statistics of eye movements are presented in Table 2. Fixation durations
540 were not different between the types of trials ($F(2, 66047) = 2.88$, $p = .056$), while for identical
541 and overlapping trials, fixation durations were significantly longer in second presentations than
542 in the first presentations ($F(1, 68218) = 13.15$, $p < .001$, Δ=6 ms, 95% CI [3, 10]). Fixation
543 durations were similar in all types of trials for the first presentation (identical, overlapping, and
544 non-repeated). The interaction between type of trial and first/second presentation was not
545 significant ($F(1, 68209) = 1.83$, $p = .177$).
546 For the saccade amplitudes, we did not observe differences between types of trials ($F(1, 60162)$
547 $= .057$, $p = .945$), presentation ($F(1, 62376) = 1.43$, $p = .232$), or the interaction ($F(1, 62366) =$
548 0.335, $p = .563$).
549
550 --- Insert Table 2 here. ---

Overlapping Content

552 An examination of the eye-tracking data revealed significant differences in fixation locations.
553 When viewing the photographs for the first time, we observed a strong preference for the central
554 third (mean = 1551 ms, SD = 603 ms) over the lateral third (mean = 1204 ms, SD = 594 ms;
555 $F(1, 8967) = 762.66$, $p < .001$). This pattern was repeated during the second presentation for
556 identical trials (central: mean = 1568 ms, SD = 653 ms; lateral: mean = 1196 ms, SD = 612 ms;
557 $F(1, 2180.1) = 191.61$, $p < .001$). In the case of overlapping trials, we again observed a
558 preference for the central third (mean = 1543 ms, SD = 601 ms) over the novel contralateral part
559 (mean = 1195 ms, SD = 606 ms; $F(1, 2215.1) = 188.85$, $p < .001$). This preference for the
560 central part was not observed for recognition accuracy, as we did not detect a correlation
561 between immediate recognition accuracy and dwell time ($r = -.01$, $p = .801$). Taken together,
562 these results indicate a strong bias for the central parts (without a behavioral effect on
563 recognition). This finding was surprising, as the participants were fixating on the central part of
564 the original image, even during the first presentation of the image.
565 We speculated that the preference for the middle part of the whole image might result from the
566 artificial crop introduced by creating the stimuli. Photographers usually do not take photographs

567 randomly but rather position interesting content in the central part of a photograph. The artificial
568 crop could disrupt the natural boundaries of the photographs, and this artificial border might be
569 distracting for the participants. We conducted a small control experiment in which participants
570 (n=8) were presented with the stimuli from Experiment 1 and their task was to respond whether
571 the photograph originally continued to the left or to the right to assess this hypothesis. Each
572 participant was presented with 593 stimuli and correctly detected the crop in the image in 72%
573 (SD = 5%) of cases (50% guessing level). Therefore, people are able to estimate the
574 composition of a particular view relative to a larger unseen photograph.
575
576

577 # Discussion

578 The eye-tracking analysis of the overlapping paradigm showed that fixations tend to be slightly
579 longer when observing identical photograph a second time. The main finding from the eye-
580 tracking experiment was preference for the central part of the original photograph. Participants
581 were able to detect the central part, even in the first presentation, without knowing in which
582 direction the scene should continue. The stimuli were presented in the center, and this
583 preference for the central part meant that participants were looking slightly to the left or to the
584 right from the center of the screen (depending on whether the original stimuli continued to the
585 left or right). By conducting a small control experiment, we also found that the preference for the
586 central part might result from the artificial crop introduced when creating the stimuli.
587 Photographers do not take photos at random but position the interesting content in the center.
588 This finding emphasizes the importance of precise stimuli selection. Small artificial
589 modifications, such as cropping an image, can influence eye movements.

590 # General Discussion

591 In a series of three experiments, we explored to what extent people make correct recognition
592 statements about partial scene views. Our experiments differed from other visual memory
593 experiments in three aspects. First, we did not show identical views in the recognition phase
594 (both immediate and delayed). We showed partial views (referred to as patches) covering 18%
595 of the original photograph. Second, we used the unpresented parts of the photographs as a
596 source of foils that generally showed similar semantic content. Finally, we repeatedly presented
597 the same photographs (either identical or overlapping views) and tested whether they would be
598 recognized better.
599 Our experiments show that people recognize photographs from their partial views well if they
600 are queried immediately (within a minute), and the recognition decreases to nearly chance level
601 when the participants are queried after the end of the study phase. Although people recognize
602 photographs from their partial views, the performance is worse than in previous studies, where
603 participants were queried with stimuli identical to those presented in the training phase (Konkle
604 et al., 2010; Vogt & Magnussen, 2007; Andermane & Bowers, 2015; Lukavský & Děchtěrenko,
605 2017). Smaller patches contain less information than the whole photographs, which may be one

606 source of errors. Patches were not taken to highlight a specific object, which might make
607 recognition more difficult, especially after longer delays. Numerous studies have shown a high
608 capacity of visual memory for objects (Brady et al., 2010) and scenes (Konkle et al., 2010),
609 including the capacity to correctly recognize state changes in objects. However, matching the
610 stored image representation with a presented partial view was difficult. Our experiments show
611 that some search is possible within short-term memory but substantially impaired when long-
612 term memory is tested.
613 Memory performance was significantly affected when people were asked about partial views
614 from unseen parts of the photograph. People were particularly prone to false alarm errors (59%
615 in Experiment 1). This finding further supports our argument that the stored representation is
616 either not sufficiently detailed or is difficult to search. People might instead rely on high-level
617 information (semantic information about the scene) or heuristics about the photograph's texture
618 and style.
619 The high proportion of false alarms in lure patches is consistent with the boundary extension
620 effect and the source-monitoring account (Intraub, 2012). We provided people with foils from
621 unseen parts of the photograph, which potentially share many properties with the original view.
622 Moreover, people might be able to detect a common layout of the underlying scene. Gottesman
623 (2011) showed that partial views of scenes prime distance judgments in areas that were not
624 visible in the original view. However, our setup substantially differed from the usual BE
625 experiments. Namely, we presented people only with the patch from the unseen part during the
626 recognition phase, while BE is usually observed when presenting the wide-angle view of the
627 original photograph (showing both the previously seen and the unseen parts). Our lure patches
628 depicted content 3.5% to 46.5% beyond the border, while the differences studied in BE
629 experiments are more subtle. The BE account would likely predict more false alarms in repeated
630 presentations, which we did not observe. Our experiment does not exclude the potential role of
631 BE. In accordance with the BE account, our results further emphasize that our memory is not
632 sufficiently detailed to exclude more difficult lures, and we may be confused by a similar style or
633 underlying layout.
634 In addition to using smaller patches instead of whole scenes, our design also employed the
635 Yes/No paradigm instead of 2AFC, as in previous studies. Using the Yes/No paradigm makes
636 the task more difficult, as shown both theoretically (Macmillan & Creelman, 2004) and
637 empirically (Cunningham, et al., 2015; Andermane & Bowers, 2015). A reasonable assumption
638 is that the use of 2AFC in our task would lead to higher accuracy.
639 We expected that repeated presentations would provide additional viewing time and lead to
640 better encoding. This expectation was confirmed, and the additional presentation increased the
641 accuracy from 58% to 69% (in Experiment 1). Additionally, we were interested in whether the
642 additional presentation would exert a different effect if identical views or overlapping views were
643 presented. We did not propose a specific hypothesis; the overlapping views might exert multiple
644 effects. If people fuse their representations into a single form, this process may cost additional
645 time and effort, which are not spent on encoding details, resulting in poorer memory than
646 identical repeated views. Alternatively, fusion may lead to deeper processing of the input and
647 better encoding along levels of the processing framework. However, we consistently observed
648 no difference. Our experiments may not have been sufficiently sensitive to measure the
649 difference or people did not fuse the photographs into a single representation.

650   We also presented the same set of stimuli to all subjects, which allowed us to compute patch
651   memorability (for seen patches) and patch lureability (for unseen patches). In the case of both
652   immediate and delayed recognition, the correlation between patch memorability (or lureability
653   for unseen patches) and memorability estimated from a pretrained neural network was small.
654   Therefore, either the model-estimated memorability of the whole photograph was not a good
655   estimate for the memorability of its regions or memorability (estimated as the hit rate) does not
656   guarantee rich memory representation. In other words, people do not notice changes even in
657   highly memorable photographs. This claim is supported by recent findings from Broers and
658   Busch (2019), who explored recollection and familiarity with respect to image memorability.
659   They reported large differences in stimuli – some highly memorable stimuli showed high
660   recollection, while the remaining stimuli showed a similar relationship for familiarity. A visual
661   inspection of patches with high false alarm rates and high hit rates also supports this claim.
662   Patches with a high hit rate either contained a highly informative object, or the whole
663   photograph was highly distinct from the remaining stimuli. Patches with high false alarm rates
664   either contained textures similar to seen parts or contained objects that were also present in the
665   presented part. The results of visual inspection are subjective in nature and need to be verified
666   in further studies. Taken together with the finding that model-estimated memorability was similar
667   for scenes sharing only 50% of the content, the memorability score estimated by the neural
668   network (and showing correlation with large-scale memorability studies) did not appear to
669   correspond with the amount of encoded details.
670   The eye-tracking analysis showed slightly longer fixations in second presentations. Otherwise,
671   the participants' eye movement patterns were similar, regardless of whether the stimuli were
672   presented twice or whether they overlapped. The most surprising finding was that participants
673   fixated more on the part corresponding to the central part of the original photograph, even when
674   the photograph was presented for the first time. This finding dominated all other possible
675   viewing strategies, such as fixating on the overlapping parts more or fixating on the novel
676   content during the overlapping condition during the second presentation. We confirmed this
677   preference in a small control experiment, which indicated that participants were able to detect
678   the artificial crops introduced when creating the stimuli. These results confirm that participants
679   are aware of the overall scene layout and could use this information for further judgments
680   (Gottesman, 2011). Furthermore, the findings emphasize the importance of careful stimuli
681   selection and controlling for possible confounders. Photographs of visual scenes are usually
682   taken with a certain spatial layout in mind, which would easily be disrupted when a photograph
683   is cropped. We did not detect increased recognition of the parts with higher dwell times. This
684   result contradicts findings from Olejarczyk, Luke, and Henderson (2014), who reported higher
685   recognition accuracy for parts on which participants fixated more. In our case, we did not
686   measure dwell times for small patches but rather for larger semantic areas (whole thirds), which
687   might mask the possible benefit of a prolonged viewing time. The observed strong preference
688   for the central third might hide more subtle effects.

Conclusions

690   The capacity and fidelity of visual memory have been important research topics in recent years,
691   and the fidelity or nature of stored information is an unresolved question. To address the
692   importance of storing details or simple gist-based representation, we used a novel methodology

693 with photograph patches and repeated presentations. Taken together, the study has two main
694 findings.
695
696 First, people can recognize the remembered photographs from their partial views. However,
697 their memory performance is worse than usually observed, and they are highly prone to false
698 alarm errors in distractors, which were selected from the unseen parts of the presented
699 photographs. We suggest that this vulnerability to false alarms arises from the semantic or gist
700 information people use in their memory decisions. The probability that the photograph will be
701 recognized from a partial view cannot be explained by model-estimated memorability for the
702 whole photograph. The eye-tracking analysis confirmed that the probability that a photograph
703 will be recognized by its partial view does not depend on the time spent looking at the stimulus.
704 Second, participants showed higher accuracy for parts of scenes that they saw twice. A
705 surprising finding was the existence of a small fixation bias toward the center of the original,
706 larger scene from which the particular view was selected. This result suggests that participants
707 are sensitive to the general layout of the scene.
708

Acknowledgments

710 This research was supported by Czech Science Foundation grants (GA 16-07983S and GA 20-
711 06894S) and RVO68081740.
712

References

714 Andermane, N., & Bowers, J. S. (2015). Detailed and gist-like visual memories are forgotten at
715 similar rates over the course of a week. *Psychonomic Bulletin & Review*, 22(5), 1358–1363. doi:
716 10.3758/s13423-015-0800-0
717
718 Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2019). Drawings of real-world scenes during free
719 recall reveal detailed object and spatial information in memory. *Nature Communications*, 10(1),
720 1-13.
721 Bainbridge, W.A., Dilks, D.D., & Oliva, A. (2017). Memorability: A stimulus-driven
722 perceptual neural signature distinctive from memory. NeuroImage, 149, 141-152.Bainbridge,
723 W.A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and
724 subjective recognition during episodic retrieval. *Scientific Reports*, 8,
725 8679.
726 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
727 Using lme4. *Journal of Statistical Software*, 67(1). https://doi.org/10.18637/jss.v067.i01
728 Broers, N., Potter, M. C., & Nieuwenstein, M. R. (2018). Enhanced recognition of memorable
729 pictures in ultra-fast RSVP. *Psychonomic Bulletin & Review*, 25(3), 1080–1086.
730 https://doi.org/10.3758/s13423-017-1295-7
731 Broers, N., & Busch, N. (2019). The effect of intrinsic image memorability on recollection and
732 familiarity. *PsyArXiv*. https://doi.org/10.31234/osf.io/hvzyw
733 Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic
734 effects on image memorability. *Vision Researc*h, 116(March), 165–178.
735 https://doi.org/10.1016/j.visres.2015.03.005

736   Cunningham, C. A., Yassa, M. A., & Egeth, H. E. (2015). Massive memory revisited: Limitations
737   on storage capacity for object details in visual long-term memory. *Learning & Memory*, 22(11),
738   563–566. https://doi.org/10.1101/lm.039404.115
739   Draschkow, D., Reinecke, S., Cunningham, C. A., & Võ, M. L.-H. (2019). The lower bounds of
740   massive memory: Investigating memory for object details after incidental encoding. *Quarterly*
741   *Journal of Experimental Psychology*, 72(5), 1176–1182.
742   https://doi.org/10.1177/1747021818783722
743   Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic)
744   associations: A speed-accuracy study. *Cognitive Psychology*, 16(4), 519-555.
745   Giesbrecht, F. G., & Burns, J. C. (1985). Two-stage analysis based on a mixed model: large-
746   sample asymptotic theory and small-sample simulation results. *Biometrics*, 477-486.
747   Gottesman, C. V. (2011). Mental layout extrapolations prime spatial processing of scenes.
748   Journal of Experimental Psychology. *Human Perception and Performance*, 37(2), 382–395.
749   https://doi.org/10.1037/a0021434
750   Goetschalckx, L., Moors, P., Vanmarcke, S., & Wagemans, J. (2019). Get the Picture?
751   Goodness of Image Organization Contributes to Image Memorability. *Journal of Cognition*, 2(1),
752   22. https://doi.org/10.5334/joc.80Green, D. M., & Swets, J. A. (1966). Signal Detection Theory
753   and Psychophysics. New York: John Wiley and Sons.
754   Hock, H. S., & Schmelzkopf, K. F. (1980). The abstraction of schematic representations from
755   photographs of real-world scenes. *Memory & Cognition*, 8(6), 543-554.
756   Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of*
757   *Experimental Psychology: Learning, Memory, and Cognition*, 15(2), 179–187.
758   https://doi.org/10.1037/0278-7393.15.2.179
759   Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews:*
760   *Cognitive Science*, 3(1), 117–127. https://doi.org/10.1002/wcs.149
761   Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? In 2011
762   *Computer Vision and Pattern Recognition* (pp. 145–152). IEEE.
763   https://doi.org/10.1109/CVPR.2011.5995721
764   Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image
765   memorability at a large scale. In *Proceedings of the IEEE International Conference on*
766   *Computer Vision* (pp. 2390-2398).
767   Khosla, A., Bainbridge, W. A., Torralba, A., & Oliva, A. (2013). Modifying the memorability of
768   face photographs. In *Proceedings of the IEEE International Conference on Computer Vision*
769   (pp. 3200–3207). IEEE. https://doi.org/10.1109/ICCV.2013.397
770   Khosla, A., Xiao, J., Torralba, A., & Oliva, A. (2012). Memorability of image regions. In
771   *Advances in Neural Information Processing Systems* (pp. 296-304).
772   Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010. Scene Memory Is More Detailed Than
773   You Think. *Psychological Science*, 21(11), 1551–1556.
774   https://doi.org/10.1177/0956797610385359
775   Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear
776   mixed effects models. *Journal of Statistical software*, 82(13), 1-26.
777   Lukavský, J., & Děchtěrenko, F. (2017). Visual properties and memorising scenes: Effects of
778   image-space sparseness and uniformity. *Attention, Perception, & Psychophysics*, 79(7), 2044–
779   2054. https://doi.org/10.3758/s13414-017-1375-9

780   Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
781   Olejarczyk, J. H., Luke, S. G., & Henderson, J. M. (2014). Incidental memory for parts of scenes
782   from eye movements. *Visual Cognition*, 22(7), 975-995.
783   Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science*, 374-
784   378.
785   Robertson, C. E., Hermann, K. L., Mynick, A., Kravitz, D. J., & Kanwisher, N. (2016). Neural
786   representations integrate the current field of view with the remembered 360 panorama in scene-
787   selective cortex. *Current Biology*, 26(18), 2463-2468.
788   Valuch, C., Becker, S. I., & Ansorge, U. (2013). Priming of fixations during recognition of natural
789   scenes. *Journal of Vision*, 13(3), 3-3.
790   Varakin, D. A., & Loschky, L. (2010). Object appearance and picture-specific viewpoint are not
791   integrated in long-term memory. *The Quarterly Journal of Experimental Psychology*, 63(6),
792   1181–1200. https://doi.org/10.1080/17470210903254639
793   Vogt, S., & Magnussen, S. (2007). Long-Term Memory for 400 Pictures on a Common Theme.
794   *Experimental Psychology*, 54(4), 298–303. https://doi.org/10.1027/1618-3169.54.4.298
795   Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the
796   recognition of faces. *Memory & Cognition*, 20(3), 291–302. https://doi.org/10.3758/BF03199666
797

Figure 1

Example of visual images.

(A) Source image used to create the visual stimuli, (B) The view presented to the participants. In the recognition phase, we asked either about patches of the original view (C) or about patches beyond the original view (D).

A B C

D

Figure 2

Three types of experimental stimuli.

Repeated and overlapping stimuli were presented twice, while non-repeated trials were presented only once. Color denotes how many times a part of an image was presented.

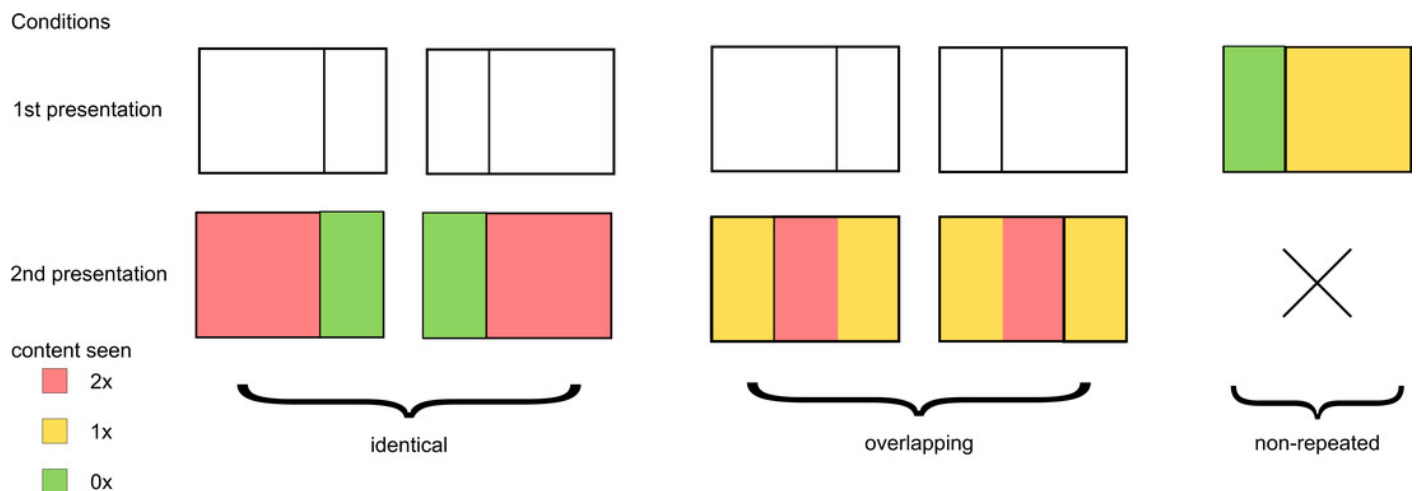



Figure 3

Experimental scheme depicting examples of stimuli in each part.

In the study part, patches were selected from the previous 5 images, while in the testing part, the queried patches were selected from the whole set.

Figure 4

Accuracy in immediate (left) and delayed (right) recognition.

Vertical lines denote bootstrapped standard errors of the mean. Individual points denote the average score per participant.

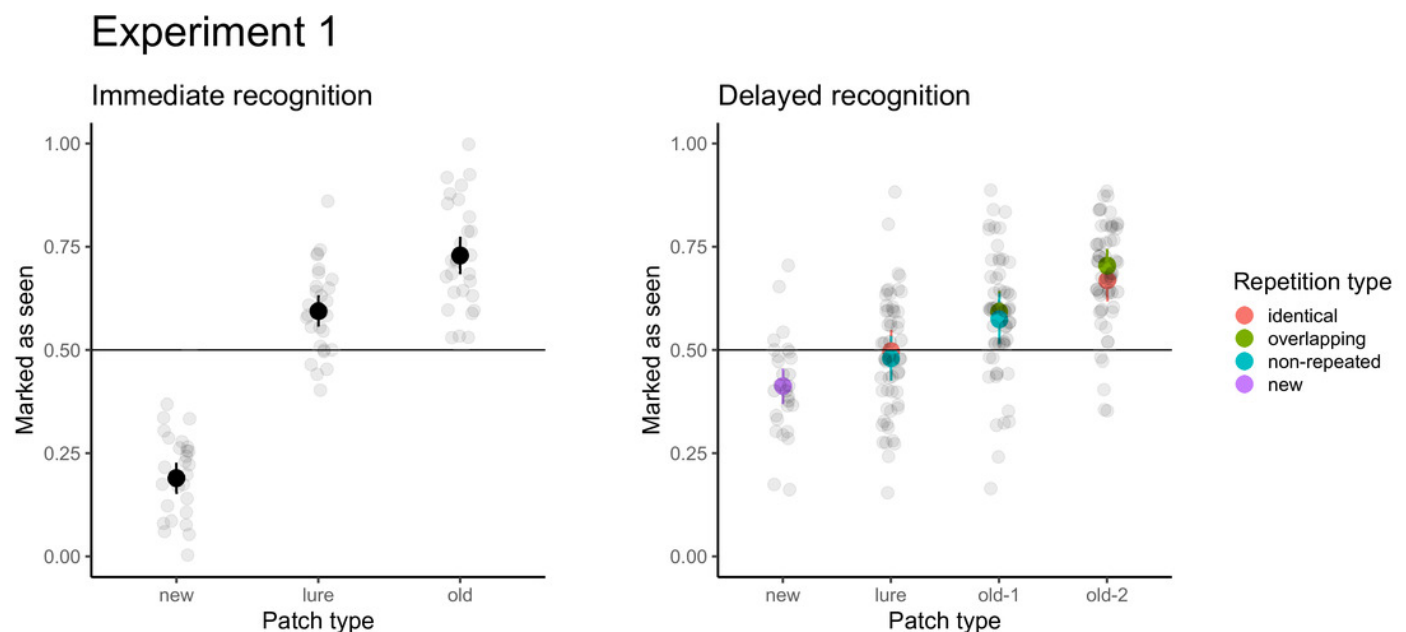



Figure 5

Accuracy for immediate recognition and delayed recognition in Experiment 2.

Vertical lines denote bootstrapped standard errors of the mean. Individual points denote the average score per participant.

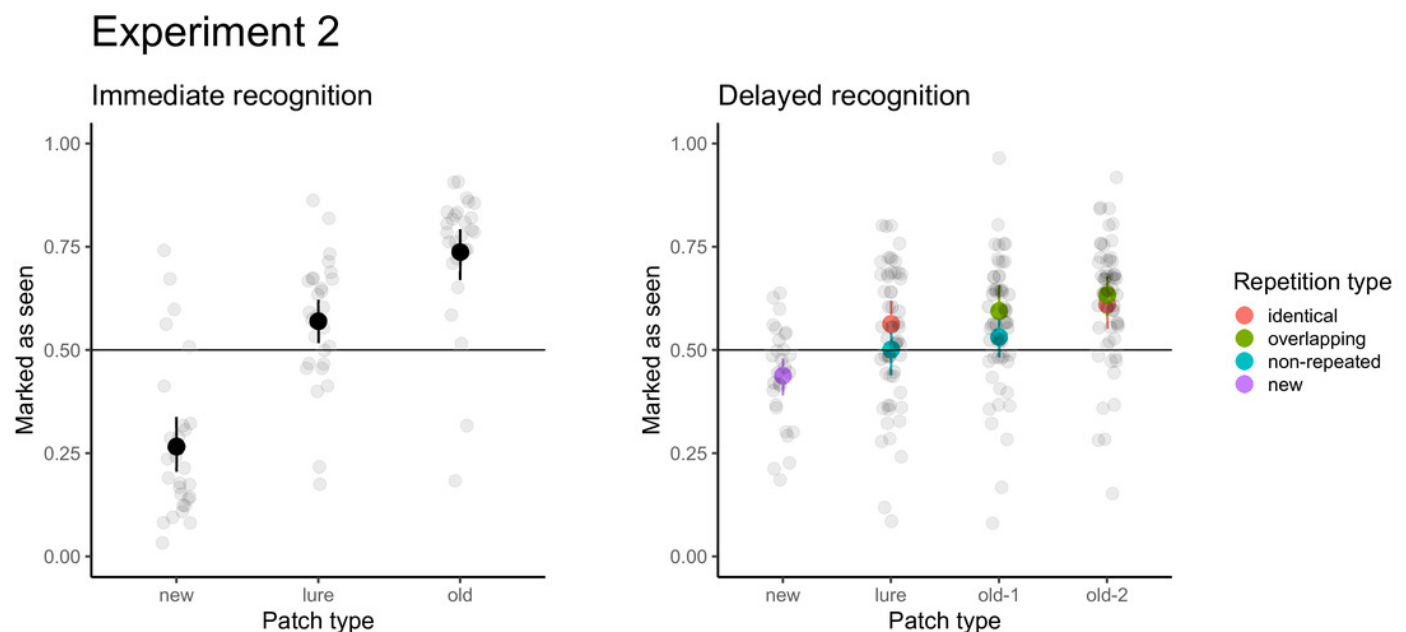
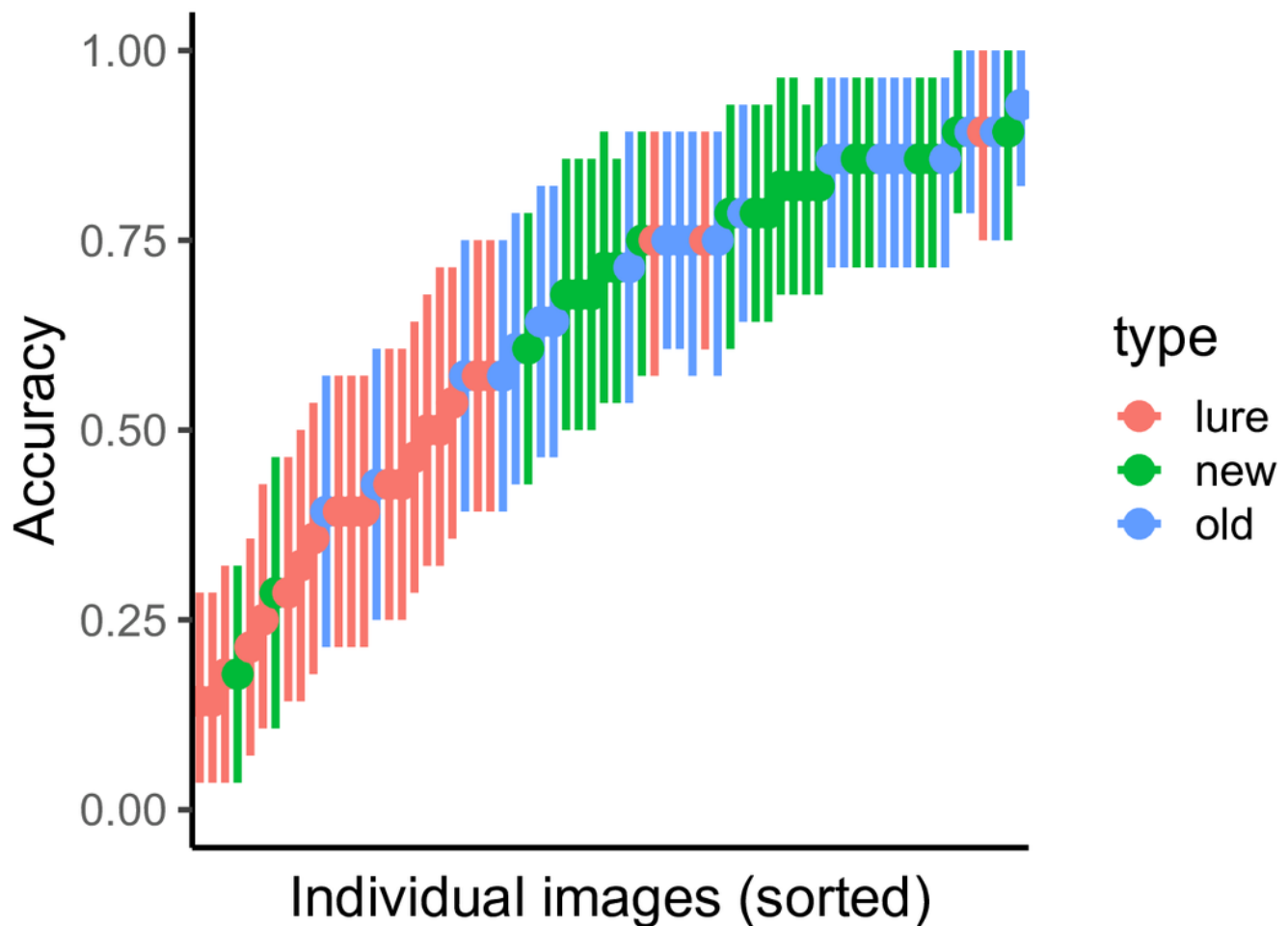



Figure 6

Variance in accuracy for immediate recognition in Experiment 2.

Vertical lines denote bootstrapped standard errors of the mean.

Breakdown of presentation trials for patch selection.

1 Table 1:

2 Breakdown of presentation trials for patch selection.

| Patch type | Presented images | | | | | | New images |
| | Identical trial | | Overlapping trial | | Non-repeated trial | | |
	Presented part	Unseen part	Central part	Noncentral part	Presented part	Unseen part	Unseen part
Old-2	25		25				
Old-1				25	25		
Lure		25				25	
New							50

Fixation durations and saccade amplitudes for each trial type and presentation.

Means with SD shown in parentheses.

1   Table 2: Fixation durations and saccade amplitudes for each trial type and presentation (means
2   with SD shown in parentheses).

variable	trial type	1st presentation	2nd presentation
Fixation duration	non-repeated	276.51 (42.10)	-
	identical	270.92 (36.14)	280.69 (48.60)
	overlapping	274.50 (40.41)	279.40 (48.94)
Saccade amplitude	non-repeated	4.10 (0.71)	-
	identical	4.11 (0.70)	4.07 (0.80)
	overlapping	4.14 (0.70)	4.06 (0.80)