

1 Category-based and location-based volitional covert
2 attention affect memory at different timescales

3 Kirsten Ziman^{1,2}, Madeline R. Lee¹, Alejandro R. Martinez¹,

Ethan D. Adner¹, and Jeremy R. Manning^{1,†}

¹Dartmouth College

²Princeton University

[†]Address correspondence to jeremy.r.manning@dartmouth.edu

4 March 26, 2023

5 **Abstract**

6 Our ongoing subjective experiences, and our memories of those experiences, are shaped by our prior
7 experiences, goals, and situational understanding. These factors shape how we allocate our attentional
8 resources over different aspects of our ongoing experiences. These attentional shifts may happen overtly
9 (e.g., when we change where we are looking) or covertly (e.g., without any explicit physical manifestation).
10 Additionally, we may attend to what is happening at a specific spatial location (e.g., because we think
11 something important is happening there) or we may attend to particular features irrespective of their
12 locations (e.g., when we search for a friend's face in a crowd versus a location on a map). We ran a covert
13 attention experiment with two conditions that differed in how long they asked participants to maintain the
14 focus of the categories and locations they were attending. Later, the participants performed a recognition
15 memory task for attended, unattended, and novel stimuli. Participants were able to shift the location of
16 their covert attentional focus more rapidly than they were able to shift their focus of covert attention to
17 stimulus categories, and the effects of location-based attention on memory were longer-lasting than the
18 effects of category-based attention.

19 **Keywords:** covert attention, spatial attention, category-based attention, recognition memory, perception

20 Introduction

21 Our brains' cognitive systems detect and exploit patterns in our prior and ongoing experiences, enabling
22 us to function and adapt in an ever-changing world. However we do not attend to or treat all types of
23 remembered or incoming information equally, and our ability to flexibly adapt our thinking and behaviors
24 can vary markedly with the specific set of concepts or tasks relevant to a given setting or situation (Chun
25 & Turk-Browne, 2007; Aly & Turk-Browne, 2017; Hardt & Nadel, 2009; Ranganath & Ritchey, 2012). There
26 is also substantial variability across people with respect to which aspects of experience (sensory, social,
27 emotional, etc.) are noticed, discriminated between, and acted upon (E. Hunt et al., 1989). This implies that
28 the same physical (objective) experience may give rise to very different perceived (subjective) experiences
29 across people (Freeman & Simoncelli, 2011; Chang et al., 2021).

30 The aspects of our experience we attend may be under our volitional control or may be unconscious
31 or automatic (Jacoby et al., 1992). Both volitional and unconscious attention may be expressed overtly, for
32 example through intentional eye movements (Hoffman & Subramaniam, 1995) or covertly, without any
33 volitional physical change (Engbert & Kliegl, 2003). Prior work has explored the similarities and differences
34 in the neural basis of overt versus covert attention (Posner et al., 1987; A. R. Hunt & Kingstone, 2003) as well
35 as the behavioral and neural underpinnings of volitional versus unconscious attention (Dijksterhuis & Aarts,
36 2010) and their differential effects on memory. There is a general consensus that sustained volitional attention
37 enhances memory relative to unconscious attentional processes (Uncapher et al., 2011; Turk-Browne et al.,
38 2013). However, volitional attention takes many forms, such as attention to particular spatial locations
39 or attention to particular visual features or other stimulus properties. How different *types* of volitional
40 attention combine (or compete) to enhance memory remains an open question. Volitional covert attention is
41 of particular interest in that it allows us to dynamically and intentionally manipulate our experience, even
42 when our sensory input remains largely static (i.e., constant physical stimuli, retinal image, etc. Yi et al.,
43 2006; O'Craven et al., 1999).

44 Here we examine the ways different types of volitional covert attention interact to affect memory. We
45 designed an experimental paradigm (following Posner, 1980) that asked participants to attend to a series
46 of presented composite image pairs while keeping their gaze fixed on a central point. The image pairs
47 comprised a left and right image, each constructed by blending an image of a face and place. The stimuli
48 and presentation durations were constant across the two experiments, but the experiments differed in how
49 often we asked participants to change the focus of their attention with respect to image category (face versus
50 place) and image location (left versus right). After the participants attended to a series of images, we used
51 a recognition memory test to assess which aspects of the presented images had been encoded into memory.

52 In both experiments we found that the images participants covertly attended to were better recognized than
53 other images, supporting the notion that attention enhances memory encoding (i.e., they rated attended
54 images as more familiar than unattended images; Yonelinas, 2002). After maintaining the focus of attention
55 to a single image category and location (Sustained Attention Experiment), participants also recognized
56 the attended-category image at the unattended location, and (to a lesser extent) the unattended-category
57 image at the attended location. After more rapidly varying their focus of attention (Variable Attention
58 Experiment), participants showed a similar boost in recognition for the unattended-category image at
59 the attended location, but they did not recognize images at the unattended location. This suggests that
60 participants were able to shift the location of their covert attentional focus more rapidly than they were able
61 to shift their focus of covert attention to stimulus features. We also found differences in the timecourses
62 of these memory effects, suggesting that the impact of location-based attention on memory persists on the
63 order of several seconds longer than the impact of feature-based attention.

64 Materials and methods

65 We ran a total of 53 participants in across two experimental conditions (Fig. 1). The two conditions differed in
66 how often we cued participants to change the focus of their attention. All code and documentation pertaining
67 to our experiments and analyses, along with the experimental stimuli and data, may be downloaded from
68 <http://www.github.com/ContextLab/attention-memory-task>.

69 Participants

70 A total of 53 Dartmouth College undergraduate students enrolled in our study (30 in the sustained attention
71 condition and 23 in the variable attention condition). Following a pilot study using a similar experimental
72 design, we aimed to enroll 30 participants in each condition. However, we fell short of our enrollment
73 target in the variable attention condition when in-person testing was discontinued at our institution due to
74 the COVID-19 pandemic. All participants had self-reported normal or corrected-to-normal vision, memory,
75 and attention. Participants gave written consent to enroll in the study under a protocol approved by the
76 Committee for the Protection of Human Subjects at Dartmouth College.

77 We used a voluntary pre-experimental survey to collect self-reported demographic information about
78 each participant. All 53 participants elected to fill out the survey. Participants ranged in age from 18–21
79 years old (mean: 18.7 years; standard deviation: 0.8 years). Participants reported their genders as female
80 (34 participants), male (18 participants), and gender non-binary (1 participant). Participants reported their

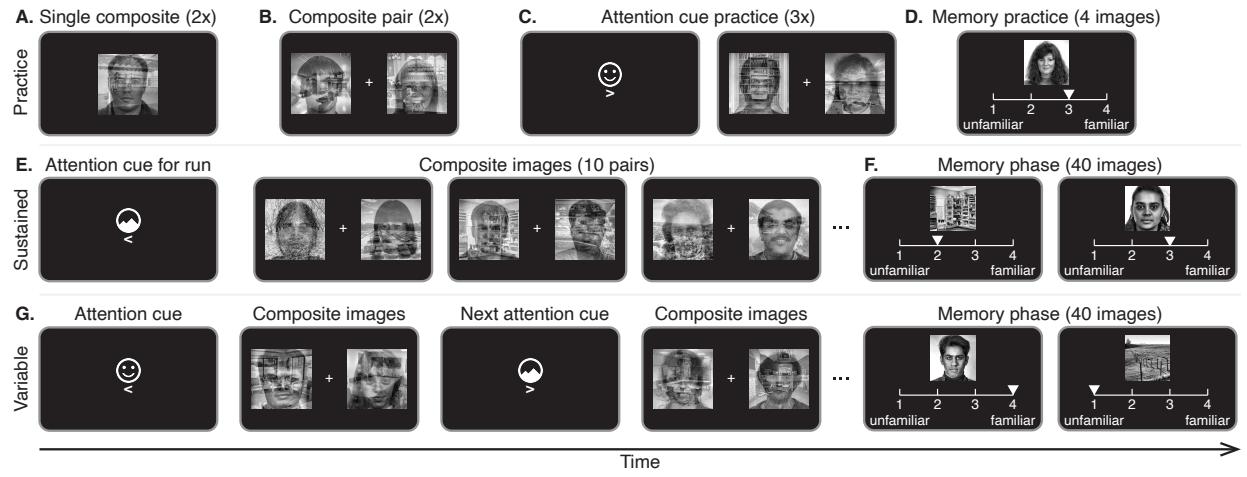


Figure 1. Experimental paradigm. A.-D. Practice phase. A. Composite face/place image. B. A single pair of composite images and a central fixation cross. C. One attention cue practice trial. D. Familiarity judgement practice trial. **E. Sustained attention condition.** Participants receive an attention cue, followed by a sequence of 10 composite image pairs (Panel E), and then they make of familiarity judgements on each of 40 face and place images, presented in sequence (Panel F). **G. Variable attention condition.** Participants study a succession of 10 composite image pairs, each preceded by an attention cue. After studying the images, they make a series of 40 familiarity judgements about face and place images, as in the sustained attention condition (Panel F).

81 ethnicities as Not Hispanic or Latino (44 participants), Hispanic or Latino (7 participants), or declined to
82 respond (2 participant). Participants reported their race as White (37 participants), Asian (13 participants),
83 American Indian or Alaska Native (4 participants), Black or African American (2 participants), and Other
84 (1 participant). Note that each participant could report one or more racial categories, as they deemed
85 appropriate.

86 Fourty-nine participants reported having no reading impairments, and 4 participants reported having
87 reading impairments such as dyslexia. Fifty participants reported having normal color vision and 3 reported
88 having abnormal color vision such as colorblindness. Fifty participants reported taking no medications and
89 having no recent injuries. One participant reported that they had recently “hit [their] head very hard.”
90 Another reported having taken concerta in the past, but mentioned they had not taken it recently. One
91 participant reported using amphetamines regularly, but also clarified that they were not currently on
92 amphetamines at the time of testing.

93 We also asked participants to self-report on their sleep, alertness, and coffee consumption. Participants
94 reported having gotten between 4 and 9 hours of sleep on the night prior to testing (mean: 6.9 hours;
95 standard deviation: 1.3 hours). Participants reported their alertness at the time of testing, and we converted
96 their responses to point values as follows: “Very alert” (5 points), “A little alert” (4 points), “Neutral” (3
97 points), “A little sluggish” (2 points), and “Very sluggish” (1 point). Across all participants, the full range of
98 alertness values were used (maximum: 5; minimum: 1; mean: 3.44; standard deviation: 1.0). Participants
99 reported having consumed between 0 and 2 cups of coffee so far on their testing day (mean: 0.3 cups;
100 standard deviation: 0.5 cups).

101 **Stimulus selection and presentation**

102 Participants viewed photographs of faces, places, and composite images each comprising an equal blend
103 of one face image and one place image. The pool of 360 face images included photographs of adult human
104 male and female faces selected from the FERET database (Phillips et al., 1998). The pool of 360 place
105 images included photographs of indoor and outdoor places selected from the SUN database (Xiao et al.,
106 2010). The images we used from both databases came from a stimulus subset that was manually curated by
107 Megan deBettencourt (personal communication). All images were resized to 256×256 pixels, converted to
108 greyscale, and processed so that every image was matched for mean contrast and intensity. We selected 20
109 face images and 20 place images from the stimulus pool to use in the instructional and practice phases of
110 the experiments (Fig. 1A–D).

111 In addition to the face and place images, we presented (in white) attention cues to direct the participant’s

112 focus of attention. The attention cues comprised a stylized icon of a face or mountain peaks, directing
113 attention to the face or place component of the images, respectively; and a left- or right-facing angled
114 bracket, directing attention to the left or right image, respectively (e.g., Figs. 1C, E, and G).

115 Our experiment was conducted in a sound- and light-attenuated testing room containing a chair, desk,
116 and 27-inch iMac desktop computer (display resolution: 2048×1152). The participant sat in the chair and
117 rested their chin on a chin rest located 60 cm from the display. The active portion of the display screen
118 occupied 52.96° (width) and 31.28° (height) of the participant's field of view from the chin rest. Stimuli were
119 sized to occupy 6.7° (width and height) of the participant's field of view from the chin rest. We maintained
120 a black background (with any text displayed in white) throughout the experiment.

121 **Eyetracking**

122 We recorded participants' eye gaze positions using a desk-mounted video-based eyetracker with a spatial
123 resolution of 0.1° visual angle root mean squared error and a sampling rate of 30 Hz (Eye Tribe, The Eye
124 Tribe, Copenhagen, Denmark). We calibrated the eyetracker using a 9-point gaze pattern. As described
125 below, we re-calibrated the eyetracker at regular intervals throughout the experiment to protect against
126 camera drift.

127 We used the eyetracking data to home in specifically on behavioral effects related to *covert* attention, as
128 opposed to overt looking effects. Specifically, we excluded from further analysis any images from trials
129 where participants shifted their gaze (for any non-zero amount of time) to any part of the attended composite
130 image during a presentation trial (see Figs. S1, and S2).

131 **Experimental paradigm**

132 Our experiment comprised two testing conditions: a *sustained* attention condition and a *variable* attention
133 condition. Both experimental conditions comprised a practice phase followed by a series of eight task
134 blocks. Each task block was in turn comprised of a presentation phase and a memory phase. The practice
135 and presentation phases differed across the two experiments, and the memory phases were identical across
136 the two experiments. Both experiments were implemented using PsychoPy (Peirce et al., 2019).

137 **Practice phase**

138 Several participants in pilot versions of our experiments reported that they found it difficult to modulate
139 the focus of their attention quickly on command. We therefore designed a practice sequence to orient the
140 participant to the process of quickly modulating the focus of their attention without moving their eyes.

¹⁴¹ The experimenter remained in the testing room throughout the practice phase and answered any questions
¹⁴² about the experiment. The practice sequence builds up incrementally to provide a gradual on-ramping for
¹⁴³ the participant prior to beginning the main experimental tasks that we focused on in our analyses.

¹⁴⁴ **Practice shifting the focus of category-based attention to elements of a single composite image.** At the
¹⁴⁵ start of the practice phase, we instructed the participant to look at a single composite (face-place blend)
¹⁴⁶ image at the center of the screen, and to try to bring the face component of the image into greater focus
¹⁴⁷ by attending to it (Fig. 1A). After pressing a button on the keyboard to indicate that they had done so, we
¹⁴⁸ displayed a second composite image and instructed the participant to bring the place component of the new
¹⁴⁹ composite image into focus. Again, they pressed a button to indicate that they had done so.

¹⁵⁰ **Practice shifting the focus of category-based and location-based attention while viewing two composite
151 images.** Next, we asked the participant to stare at a fixation cross presented in the center of the screen
¹⁵² while two composite images were displayed on the left and right side of the screen, respectively (Fig. 1B). We
¹⁵³ first instructed the participant to attend to the place component of the left image without moving their eyes.
¹⁵⁴ Participants practiced shifting their attention, and they pressed a button on the keyboard to indicate that
¹⁵⁵ they had done so. We then displayed a second pair of composite images and instructed the participant to
¹⁵⁶ attend to the face component of the right image. Again, the participant shifted their attention in a self-paced
¹⁵⁷ manner, and pressed a button to indicate when they had successfully done so.

¹⁵⁸ **Practice sustaining category-based and location-based attention over a series of composite image pairs.**
¹⁵⁹ We asked participants in the sustained attention condition to practice holding their focus of category-based
¹⁶⁰ and location-based attention constant (to the face component of the right image) while viewing a series of
¹⁶¹ three composite image pairs presented in succession (Fig. 1C).

¹⁶² **Practice varying category-based and location-based attention over a series of composite image pairs.**
¹⁶³ We asked participants in the variable attention condition to practice varying their focus of category-based
¹⁶⁴ and location-based attention while viewing a series of three composite image pairs, each presented after a
¹⁶⁵ different attention cue (Fig. 1C).

¹⁶⁶ **Practice reaction time probe.** After practicing modulating their focus of attention to a series of composite
¹⁶⁷ image pairs, we introduced a reaction time probe after each image presentation, whereby we presented
¹⁶⁸ either an \times or \circ on either the left or the right of the screen (not shown). We asked the participant to press the
¹⁶⁹ 1 key as quickly as possible when they saw an \times , or the 3 key as quickly as possible when they saw an \circ . We

170 did not impose a time limit on their responses, other than asking participants to respond as quickly as they
171 were able. Participants practiced three trials of modulating their focus of attention to a pair of composite
172 images (3 s), and reacting as quickly as possible to the × or ◦ symbol presented after each composite image
173 pair. The reaction time probe was intended to keep participants continually engaged in modulating the
174 focus of their attention.

175 **Practice recognition memory task.** Finally, we asked the participant to practice reporting familiarity on
176 a recognition memory task (Fig. 1D). We presented a single face or place image at the center of the screen,
177 and asked them to press a button to indicate how “familiar” the image seemed: 1 (very confident they had
178 not seen the image), 2 (somewhat confident they had not seen the image), 3 (somewhat confident they had
179 seen the image), or 4 (very confident that they had seen the image). We instructed the participant to go with
180 their “gut reaction” in the event that they were unsure of how to respond. We allowed the participant up to
181 2 s to provide their response. We gave participants a total of four practice images to rate.

182 After completing the practice phase of the experiment, the participant read the instructions for the task
183 blocks (described next). The experimenter gave participants a chance to ask any remaining questions about
184 the experiment. After answering the participant’s questions, the experimenter calibrated the eyetracker and
185 exited the testing room.

186 **Task blocks**

187 During each task block we asked the participant to modulate their attention while viewing a series of
188 10 composite image pairs (each followed by a reaction time probe), and then we tested the participant’s
189 memory using 40 familiarity judgements. Each participant completed a total of eight task blocks.

190 **Sustained attention condition: presentation phase (Fig. 1E).** Participants viewed an attention cue (1.5 s)
191 instructing them to attend to either the face or place component of either the left or right images in each
192 to-be-viewed composite pair. Next we displayed 10 composite images in succession (each preceded by a
193 fixation cross and proceeded by a reaction time probe). All possible attention cue pairs appeared exactly
194 twice across the eight task blocks.

195 **Variable attention condition: presentation phase (Fig. 1G).** Participants viewed a succession of 10 atten-
196 tion cues (1.5 s), each followed by a fixation cross (1 s), composite image (3 s), and a reaction time probe.
197 The attention cues were selected pseudorandomly across trials within each block, with the constraints that
198 no single attention cue pair could appear more than three times across the 10 composite image pairs within

199 a single task block.

200 **Memory phase (Fig. 1F).** After the presentation phase of each task block, we asked the participant to rate
201 the familiarity (on a 1–4 scale, as during the practice phase) of a succession of face and place images. Each
202 image was preceded by a 1 s fixation cross, and participants had up to 2 s to input their rating of each
203 image. Participants made a total of 40 familiarity judgements, about 20 face images and 20 place images. Of
204 these, 20 of the images (10 faces and 10 places) were drawn randomly from the (attended and unattended)
205 composite images that the participant had viewed during the presentation phase. The remaining 20 images
206 (10 faces and 10 places) were novel images that the participant had not encountered during any part of the
207 experiment. At the end of each memory block, the participant was given the opportunity to take a short
208 break. When they were ready to continue with the next task block, they indicated their readiness to the
209 experimenter. The experimenter then entered the testing room, re-ran the eyetracker calibration sequence,
210 and exited the testing room prior to the next task block.

211 Results

212 We ran a volitional covert attention experiment with two conditions; in the sustained attention condition
213 we asked participants to *sustain* the focus of their attention over a succession of 10 stimulus presentations
214 per block whereas in the variable attention condition we asked participants to *vary* their focus of attention
215 with each new stimulus (also for a total of 10 stimulus presentations per block). Each stimulus comprised
216 a pair of composite images (one on the left and one on the right side of the display), where each composite
217 comprised an equal blend of a unique face and a unique place image. We followed the presentation phases
218 of each experimental block with a memory phase, where participants performed a recognition memory
219 task by rating the familiarity of previously experienced and novel face and place images (see *Experimental*
220 *paradigm*, Fig. 1).

221 We first wondered whether (and how) shifts in covert attention might affect participants' ratings during
222 the recognition memory task (Fig. 2). To ensure that our findings were not conflated with where people
223 were physically looking, we excluded from further analysis any images presented during trials where the
224 participant's gaze touched on any part of the attended composite image (see *Eyetracking*, Figs. S1 and S2).
225 For the remaining trials, the participants kept their gaze focused on a fixation cross at the center of the
226 screen while *covertly* shifting the focus of their attention to the cued category component of the composite
227 image on the cued location. In other words, during these remaining trials, participants' physical (external)
228 experiences of the face and place components of every presented composite image remained relatively

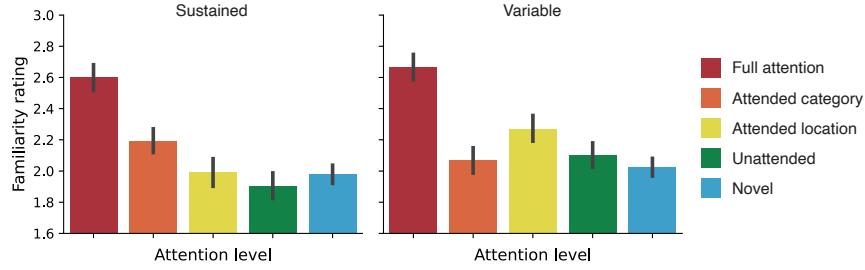


Figure 2. Familiarity by attention level. The bars display the average familiarity ratings participants gave to images from the same category and location as the attention cue (fully attended), the same category (but opposite location) as the attention cue (attended category), the same location (but opposite category) as the attention cue (attended location), the opposite category and location as the attention cue (unattended), or novel images. The left panel displays familiarity ratings from the sustained attention condition and the right panel displays familiarity ratings from the variable attention condition. All error bars denote across-participant bootstrap-estimated 95% confidence intervals. For results sub-divided by stimulus category, see Figure S3.

229 constant across trials (up to our ability to accurately measure where participants were looking using the
230 eyetracker).

231 Simply by encoding their prior experiences into memory, we reasoned that participants should rate
232 *any* presented images as more familiar than novel images, regardless of whether they were following the
233 attention cues. We confirmed that this prediction held in both the sustained ($t(29) = 8.856, p < 0.001$) and
234 variable ($t(22) = 5.144, p < 0.001$) conditions. In addition, to the extent that participants were following the
235 attention cues, there *internal* experiences of each image should depend on their internal focus of attention
236 during each image presentation. For example, we expected that the attended-category component of
237 the composite image at the attended location might be better recognized than the other composite image
238 components. Indeed, participants in both experimental conditions rated these “fully attended” images as
239 more familiar than category-matched image components from unattended locations (sustained: $t(29) =$
240 $6.893, p < 0.001$; variable: $t(22) = 6.938, p < 0.001$), location-matched images from the unattended category
241 (sustained: $t(29) = 6.710, p < 0.001$; variable: $t(22) = 7.633, p < 0.001$), unattended images that were
242 neither from the attended category nor the attended location (sustained: $t(29) = 8.470, p < 0.001$; variable:
243 $t(22) = 7.256, p < 0.001$), or novel images they had never seen before (sustained: $t(29) = 10.259, p < 0.001$;
244 variable: $t(22) = 7.874, p < 0.001$).

245 We also wondered whether the ways participants attended to or remembered the images might depend
246 on image-specific properties like the images’ categories. We repeated the analysis displayed in Figure 2

separately for face and place images (Fig. S3). The same general patterns held for each category, as when we combined across the two stimulus categories as reported above. For example, fully attended face and place images were both rated as more familiar than the category-matched images from the unattended location (sustained: $ts(29) \geq 3.36, ps \leq 0.002$; variable: $ts(22) \geq 6.205, ps \leq 0.001$), attended-location images from the unattended category (sustained: $ts(29) \geq 5.886, ps \leq 0.001$; variable: $ts(29) \geq 4.277, ps \leq 0.001$), images from the unattended category and location (sustained: $ts(29) \geq 6.628, ps \leq 0.001$; variable: $ts(29) \geq 5.624, ps \leq 0.001$), and novel images (sustained: $ts(29) \geq 5.987, ps \leq 0.001$; variable: $ts(29) \geq 5.132, ps \leq 0.001$). Taken together, the above results suggest that what we remember is guided in part by what we attended to, even after accounting for where we look or what specifically we are looking at.

Splitting participants' responses to face versus place images also revealed that participants often rated attended (and partially attended) place images as more familiar than attention-matched face images (compare dark versus light bars in Fig. S3). We hypothesized that this might be explainable by some property of the relevant cognitive processes or by properties of the stimuli themselves. To help elucidate this distinction, we examined individual exemplars of the face and place images used in our paradigm (Fig. 3A). By design, the face images had consistent head sizes, viewing angles, expressions, and so on. In contrast, the place images varied more substantially across images. For example, some place images depicted human-made structures; others depicted natural scenes; some depicted indoor views; others depicted outdoor views; etc. This can also be seen by averaging the pixel intensity values across images, separately for the face and place stimuli (Fig. 3B). Whereas the average face image retains many of the landmarks characteristic of most faces (e.g., clearly defined hair, eyes, nose, mouth, head shape, etc.), the average place image does not show place-specific features as clearly, aside from a general tendency for the tops of place images to be lighter than the bottoms of place images. We also computed the pairwise similarities across images from each stimulus category (Fig. 3C) and found that face images tended to be much more similar to each other than place images (Fig. 3D; $t(115258) = 254.764, p < 0.001$). This analysis indicated to us that our experimental paradigm was not well-suited to identifying cognitively meaningful stimulus category differences, since participants' category-specific judgements may be confounded with within-category image similarity differences.

Next, we turned to identifying potential differences in participants' behaviors across the two experimental conditions. The main difference between the conditions' procedures was in how long participants were asked to maintain the same focus of category-based and location-based attention, across successive image presentations. Therefore, differences in participants' behaviors across the two conditions might reflect differences in the timescales of the relevant cognitive processes. We compared participants' familiarity ratings for images at each attention level across the two conditions. We saw no evidence that people rated fully attended

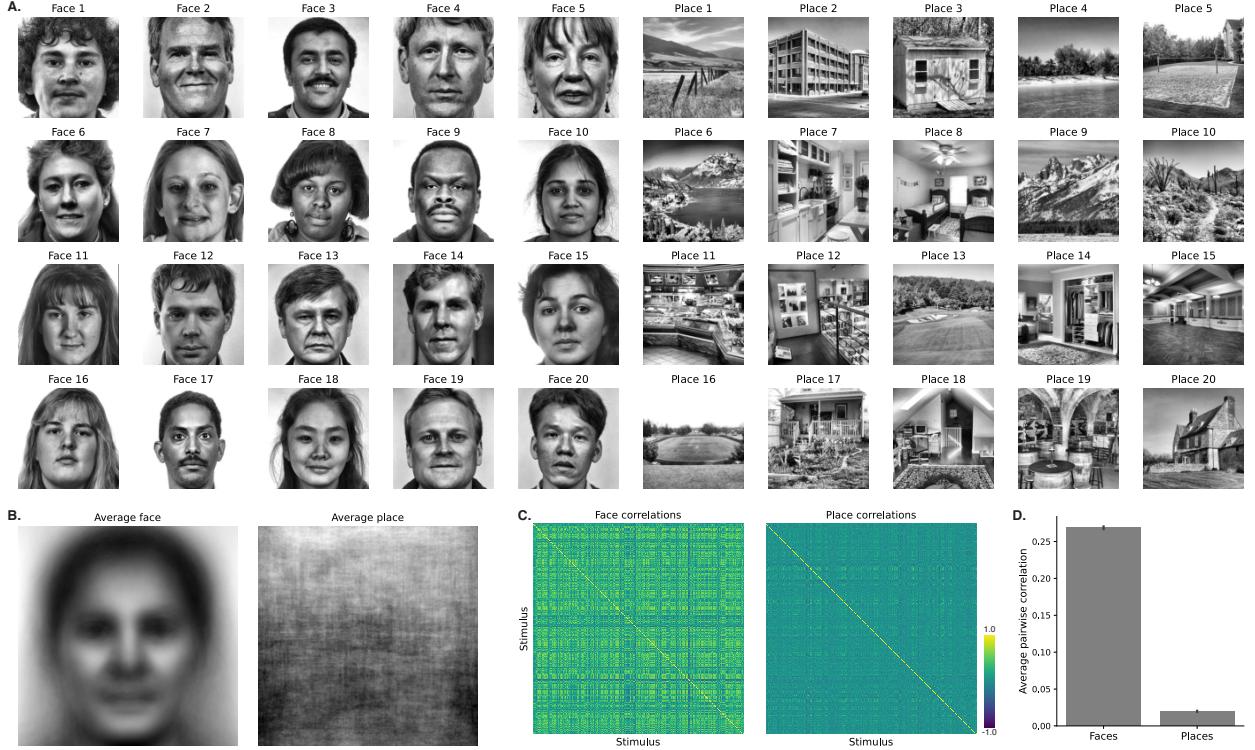


Figure 3. Stimulus examples and properties. **A. Example images from each stimulus category.** Randomly chosen subsets of 20 face images (left) and 20 place images (right) are displayed. **B. Across-image averages.** Each panel displays the average image, taken across all 360 face images (left) and 360 place images (right). **C. Pairwise correlations.** Each row and column of the matrices displays the correlation (across pixels) in intensity values for one pair of face images (left) or place images (right). **D. Average pairwise correlations.** The bar heights denote the average pairwise correlations between face and place images. Error bars denote across-pair bootstrap-estimated 95% confidence intervals.

280 images ($t(51) = -0.649, p = 0.519$), attended-category (but not location) images ($t(51) = 1.163, p = 0.250$),
281 or novel images ($t(51) = -0.435, p = 0.665$) differently across the two conditions. However, participants in
282 the variable attention condition rated attended-location (but unattended category) images as more famili-
283 iar than participants in the sustained attention condition ($t(51) = 2.174, p = 0.034$). We found a trending
284 effect for unattended category and location images, whereby participants in the variable attention condi-
285 tion tended to rate these images as more familiar than participants in the sustained attention condition
286 ($t(51) = 1.600, p = 0.116$). We also observed some within-condition differences in how participants rated
287 partially attended versus novel images. In the sustained attention condition, participants rated attended-
288 category images at the unattended location as more familiar than novel images ($t(29) = 6.205, p < 0.001$),
289 but participants in the variable attention condition did not show this pattern ($t(22) = 1.042, p = 0.309$).
290 On the other hand, whereas participants in the sustained attention condition showed no reliable differ-
291 ences in familiarity between attended-location images from the unattended category and novel images
292 ($t(29) = 0.165, p = 0.870$), participants in the variable attention condition rated attended-location images as
293 more familiar than novel images ($t(22) = 3.026, p = 0.006$). Taken together, our analyses highlight several
294 familiarity differences in partially attended images from the attended category or at the attended location
295 across the two experimental conditions. These differences suggest that category-based versus location-based
296 attention operate over different timescales.

297 Given the above results suggesting potential differences in the timescales of category-based and location-
298 based attention, we carried out two additional exploratory analyses aimed at identifying other timing effects.
299 First, we wondered whether participants' familiarity ratings might show serial position effects analogous to
300 those reported in classic recognition memory studies (e.g., Neath, 1993; McElree & Dosher, 1989; Wickelgren
301 & Norman, 1966). For each participant, for each composite image pair (presented in each trial during the
302 study phase of the experiment), we labeled each composite's face and place image component according
303 to whether it matched the cued category and/or location. We discarded any face or place images that
304 did not appear in the participants' memory phase. We tagged the remaining (probed) images with the
305 familiarity ratings that participants would later give the images during the memory phase and plotted these
306 ratings against the images' presentation positions (Figs. 4A and B). Across both experimental conditions and
307 across all serial positions, we generally found that the average ordering of familiarity ratings by attention
308 level (Fig. 2) were preserved. This suggests that the encoding-related affects of attention on subsequent
309 recognition memory are relatively stable over time (e.g., we did not observe clear primacy or recency
310 effects during the study phase of the experiment). Second, we carried out an analogous analysis to identify
311 potential serial position effects of recall order. For each probed item a participant rated during the memory
312 phase, we assigned the image a label according to whether the participant's attention cue (at the time

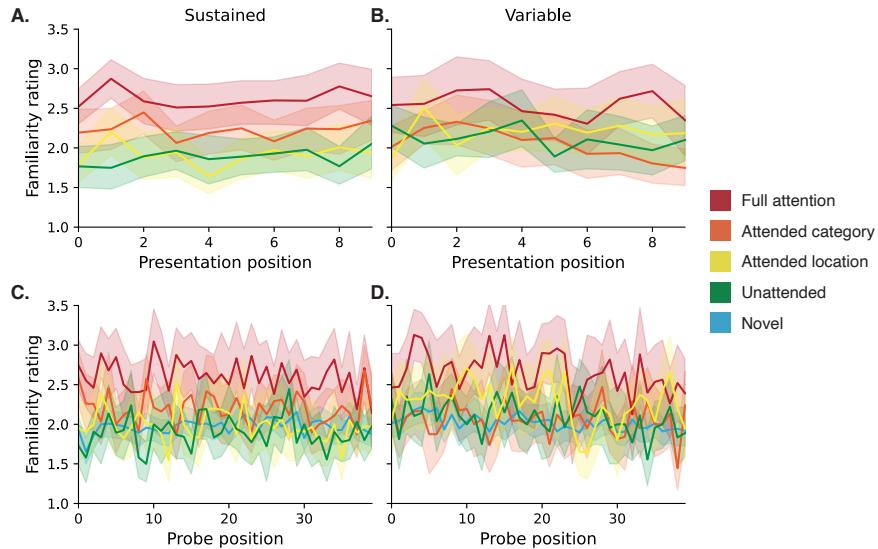


Figure 4. Familiarity ratings by serial positions and attention level. **A. Subsequent familiarity ratings by presentation position (sustained attention condition).** The curves' colors denote the attention levels of each presented image. The *x*-axis denotes the presentation positions of each image within the sequence of 10 composite image pairs during the run when it was presented. The *y*-axis denotes the average familiarity ratings later given to the corresponding items. **B. Subsequent familiarity ratings by presentation position (variable attention condition).** This panel is in the same format as Panel A, but displays ratings for the variable attention condition. **C. Familiarity ratings by memory probe position (sustained attention condition).** The curve's colors denote the attention levels (or novelty) of each probe image. The *x*-axis denotes the position of each probed image within the sequence of 40 images that participants judged during the memory phase of the experiment. The *y*-axis denotes the average familiarity ratings given to the corresponding probes. **D. Familiarity ratings by memory probe position (variable attention condition).** This panel is in the same format as Panel C, but displays ratings for the variable attention condition. All panels: error ribbons denote across-participant bootstrap-estimated 95% confidence intervals.

313 the image was presented as part of its composite pair) matched the image's category and/or location, or
314 whether the image was novel (i.e., not presented during the study phase). Again, we found that (across
315 both experimental conditions and all probe positions) in general the average ordering of familiarity ratings
316 by attention level (Fig. 2) were preserved. This suggests that the retrieval-related affects of attention on
317 subsequent recognition memory are relatively stable over time (e.g., we did not observe clear primacy or
318 recency effects during the memory phase of the experiment).

319 Our finding that location-attended items appear to receive a familiarity boost in the variable attention
320 condition but not the sustained attention condition is consistent with two possible interpretations. One
321 possibility is that focusing attention requires just a brief trigger (in this case, an attention cue), but different
322 forms of attention (in this case, category-based attention versus location-based attention) require different
323 amounts of time to "ramp up" to full efficacy such that they begin to affect memory encoding. For example,
324 if category-based attention ramps up more slowly than location-based attention, this might explain why the
325 relative ordering of category-matched versus location-matched unattended images changes between the
326 sustained attention and variable attention conditions (orange and yellow bars in Fig. 2). A second possibility
327 is that each attention cue provides a "boost" to memory encoding for the relevant aspects of one's experience
328 (e.g., image categories, spatial locations), but the size of the boost varies across different forms of attention.
329 If so, the *number* of successive attention cues one receives should predict how effectively the attended and
330 partially attended images are encoded. We developed a sequence length analysis to distinguish between
331 these possibilities. For each probed (target) image that had been presented as a composite image pair, we
332 computed the number of matching attention cues the participant had received by the time the given images
333 were presented (up to and including the image's composite pair). We computed these sequence lengths by
334 defining "matching" cues in three ways: (a) a match means the cues are for the same category *and* location,
335 (b) a match means that the cues are for the same category, and (c) a match means that the cues are for the
336 same location. This yielded, for each target image, an associated count of how many matching attention
337 cues the participant had received up to and including that image's presentation. As shown in Figure S4, we
338 used linear regressions of familiarity rating on sequence lengths to identify potential sequence length effects.
339 For both experimental conditions, all attention levels, and for all three approaches to defining matching cue
340 sequence lengths, we found virtually no reliable associations between cue sequence length and familiarity.
341 This finding is most parsimonious with the first possibility mentioned above— i.e., that attention may be
342 guided by a brief trigger, but that different forms of attention may take differently long to affect memory
343 encoding.

344 Finally, we wondered whether participants' familiarity ratings might be influenced in part by response
345 bias effects. For example, a participant who had recently received an "attend to face images" cue might

346 rate a probed face images as more “familiar” even if they had no specific memories for the given image.
347 Indeed, participants in the sustained attention condition do show some response biases. For example,
348 participants tended to rate novel images as more familiar if they came from the just-cued category (Fig. S5A;
349 $t(29) = 4.371, p < 0.001$). Responses biases are more difficult to evaluate in the variable attention condition.
350 For example, should cue recency be defined as the number studied composite image pairs that came between
351 the image whose familiarity the participant is judging and the most recent save-category cue (i.e., distance
352 to the nearest same-category probe)? Or might response biases instead arise when a given category is cued
353 more often near the end of the just-studied list? We assigned each probed image a label based on these two
354 measures of category-matched cue recency. We then used linear regressions of familiarity on these recency
355 measures to identify potential response biases (Fig. S5B). Altogether we found no evidence that participants
356 in the variable attention condition tended to rate images as more familiar solely due to how recently they
357 had received a same-category attention cue.

358 Discussion

359 We ran a covert attention experiment with two conditions that asked participants to sustain or vary the
360 focus of their covert attention, respectively. We then administered recognition memory tests that asked
361 participants to rate the “familiarity” of attended, unattended, and novel images. In our analyses, we used
362 eyetracking data to focus in on trials where participants were specifically varying their focus of *covert*
363 attention (i.e., with no change in external behavior), as opposed to simply moving their eyes to look at the
364 to-be-attended images. In both conditions, we found that participants recognized images from the attended
365 category and location more readily than unattended or novel images. This effect was substantial and
366 robust (in both conditions) and also held for individual image categories. Whereas prior work has focused
367 primarily on *overt* changes in attention (e.g., changes in eye movements associated with attention cues),
368 we show that what people *covertly* attend to also affects how they remember their ongoing experiences.
369 Specifically, covert attention appears to boost memory encoding such that the focus of covert attention is
370 recognized more readily later on.

371 We also found that partially attended images (e.g., from the attended category but unattended location,
372 or at the attended location but unattended category) were rated as more familiar than novel images. How-
373 ever, these encoding benefits differed across the two experimental conditions. In the sustained attention
374 condition, attended-category images from the unattended location were rated as more familiar than novel
375 images, but attended-location images from the unattended category were not. Participants in the variable
376 attention condition showed the opposite pattern. The variable attention participants rated attended-*location*

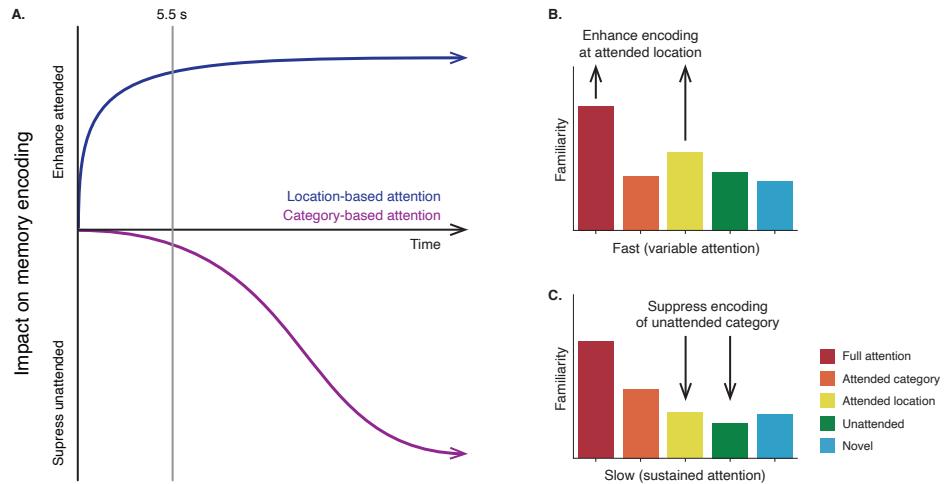


Figure 5. How do covert location-based and category-based attention affect memory encoding?

A. Hypothesized timecourses of the impact of location-based and category-based attention on memory encoding. Shifting the focus of location-based attention increases memory encoding at the attended locations (blue curve). This increase can be observed after 5.5 s (the duration of one presentation from the variable attention condition). Shifting the focus of category-based attention suppresses memory encoding for the unattended category. However, this suppression effect occurs relatively slowly (longer than the duration of a single image presentation in the experiment).

B. Location-based attention. Focusing covert attention on one *location* enhances encoding of stimuli at the attended location (red and yellow bars), regardless of stimulus category. (This panel is based on the variable attention results presented in Fig. 2.)

C. Category-based attention. Focusing covert attention on one *category* suppresses encoding of stimuli from the unattended category (yellow and green bars), regardless of spatial location. (This panel is based on the sustained attention results presented in Fig. 2.)

377 images from the unattended category as more familiar than novel images, but they rated attended-category
378 images from the unattended location similarly to novel images. Because the primary difference between
379 the sustained and variable attention conditions was the duration of the attention cues, our analyses of par-
380 tially attended images suggest that the effects of different aspects of attention on memory may unfold over
381 different timescales (Fig. 5). Specifically, location-based attention appears to affect memory encoding rela-
382 tively quickly, which would explain why attended-location images from both the attended *and* unattended
383 category receive a memory encoding benefit in the variable attention condition. In contrast, category-based
384 attention appears to affect memory encoding more slowly, and it appears to operate in a “suppressive”
385 manner (i.e., suppressing encoding of the unattended category, as opposed to enhancing encoding of the
386 attended category). This would explain why unattended-category images at the attended location are rated
387 as *less* familiar in the sustained attention condition.

388 The notion that location-based attention operates at a faster timescale than category-based attention
389 is supported by prior work on the deployment of visual attention (Soto & Blanco, 2004; Stoppel et al.,
390 2007). Our findings that location-based attention enhances the processing of attended stimuli whereas
391 category-based attention suppresses the processing of unattended stimuli is also consistent with prior work
392 on location-based attention (e.g., Itti & Koch, 2001) and category-based attention (e.g., Moher et al., 2014).
393 Our finding that people better remember attended stimuli also follows prior work on interactions between
394 attention and memory (Paller & Wagner, 2002; Chun & Turk-Browne, 2007; Aly & Turk-Browne, 2016, 2017;
395 Wittig et al., 2018; Morrison et al., 2014; Balestrieri et al., 2021). Whereas much of this prior work focused
396 on elucidating the neural basis of these interactions, our work extends these prior studies by elucidating
397 the specific and separable behavioral impacts of location-based attention (enhancement with a fast onset)
398 and category-based attention (inhibition with a slow onset) and on subsequent recognition memory. Both
399 of these effects persisted throughout the 2 min memory phases of both experiments. Therefore future work
400 is needed to elucidate the longevity of these effects beyond 2 minutes.

401 Another important area for future study concerns how the flow of information between different brain
402 structures is modulated according to the focus of volitional attention—particularly with respect to pathways
403 from primary sensory regions (e.g., V1, A1) to regions implicated in encoding ongoing experiences into
404 memory (e.g., medial temporal lobe structures such as the hippocampus and entorhinal cortex, prefrontal
405 cortex, etc.). For example, several studies suggest that attention serves to modulate the *gain* of specific
406 neural circuits (Treue & Trujillo, 1999; Chance et al., 2002; Eldar et al., 2013; Salinas & Thier, 2000), effectively
407 facilitating or inhibiting the flow of specific neural representations (Vartanian et al., 2007; LaRocque et al.,
408 2014). Prior work suggests that category-based attention may be supported by changes in connectivity
409 with the thalamus (Schneider, 2011), whereas location-based attention may be supported by changes in

410 connectivity with primary visual cortex (Noudoost et al., 2010). That category-based and location-based
411 attention are mediated by different brain structures may explain why these different aspects of attention
412 operate on different timescales and affect memory differently. A strong test of this hypothesis would entail
413 directly measuring neural activity patterns as people modulate their focus of attention (e.g., using functional
414 magnetic resonance imaging or electroencephalography), and then using neural decoding approaches (e.g.,
415 Haxby et al., 2001; Norman et al., 2006; Manning et al., 2018; Owen et al., 2021) to follow how neural
416 representations of attended (or unattended) stimuli are transferred from primary sensory regions, to higher
417 order sensory regions, to memory areas. If the effects of attention on memory are mediated by changes in
418 network dynamics, the transmission rates of the representations of attended stimuli from primary sensory
419 regions to memory areas should be facilitated relative to the transmission rates of unattended stimuli.
420 Further, variability in these neural changes (e.g., as a participant focuses their attention with more or less
421 success) should track with behavioral measures of memorability.

422 Which aspects of our ongoing experiences we choose to attend affects how we process and remem-
423 ber those experiences later. Different forms of attention—e.g., to specific stimulus categories or spatial
424 locations—operate and affect memory at different timescales, and are likely mediated by different brain
425 networks. Elucidating the behavioral and neural consequences of volitional changes in attention is central
426 to discovering how our thoughts, feelings, goals, and situational understanding fluctuate from moment to
427 moment.

428 **Author Contributions**

429 JRM and KZ developed the concept for this study. Experiment code was written by KZ and ARM, and
430 testing and data collection were conducted by MRL and KZ. KZ, MRL, ARM, and JRM analyzed the data.
431 JRM supervised the project. All authors contributed to writing and editing the manuscript.

432 **Data and code availability**

433 All of the data analyzed in this manuscript, along with all of the code for running our experiment and
434 carrying out the analyses may be found at <http://www.github.com/ContextLab/attention-memory-task>.

435 **Acknowledgements**

436 This work was supported in part by NSF EPSCoR Award Number 1632738 to JRM and by NSF CAREER
437 Award Number 2145172 to JRM. The content of this manuscript is solely the responsibility of the authors
438 and does not necessarily represent the official views of our supporting organizations. We are grateful
439 for useful discussions with the EPSCoR Attention Consortium, particularly with Marian Berryhill, Gideon
440 Caplovitz, Patrick Cavanagh, Theresa Desrochers, and Peter Tse. We thank Megan deBettencourt for useful
441 discussions, pointers to our experimental stimuli, and for providing code to facilitate stimulus preparation.
442 We also appreciate useful discussions with Paxton Fitzpatrick, Kevin Hartstein, Talia Manning, Lucy Owen,
443 and Michael Ziman. Finally, we thank Christina Liu and Eowyn Pak for their help running pilot versions
444 of the experiments reported here.

445 **References**

- 446 Aly, M., & Turk-Browne, N. (2016). Attention promotes episodic encoding by stabilizing hippocampal
447 representations. *Proceedings of the National Academy of Sciences, USA*, 113(4), 420–429.
- 448 Aly, M., & Turk-Browne, N. B. (2017). How hippocampal memory shapes, and is shaped by, attention. In
449 *The hippocampus from cells to systems* (pp. 369–403). Springer.
- 450 Balestrieri, E., Ronconi, L., & Melcher, D. (2021). Shared resources between visual attention and visual
451 working memory are allocated through rhythmic sampling. *European Journal of Neuroscience*, 55(11),
452 3040–3053.
- 453 Chance, F. S., Abbott, L. F., & Reyes, A. D. (2002). Gain modulation from background synaptic input.
454 *Neuron*, 35(4), 773–782.
- 455 Chang, L. J., Jolly, E., Cheong, J. H., Rapuano, K., Greenstein, N., Chen, P.-H. A., & Manning, J. R. (2021). En-
456 dogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects
457 affective experience. *Science Advances*, 7(17), doi.org/10.1126/sciadv.abf7129.
- 458 Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in*
459 *Neurobiology*, 17(2), 177–184.
- 460 Dijksterhuis, A., & Aarts, H. (2010). Goals, attention, and (un)consciousness. *Annual Review of Psychology*,
461 61, 467–490.

- 462 Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, 16(8), 1146.
- 463
- 464 Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, 43(9), 1035–1045.
- 465
- 466 Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14, 1195–1201.
- 467 Hardt, O., & Nadel, L. (2009). Cognitive maps and attention. *Progress in Brain Research*, 176, 181–194.
- 468 Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and
469 overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- 470 Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception*
471 and *Psychophysics*, 57(6), 787–795.
- 472 Hunt, A. R., & Kingstone, A. (2003). Covert and overt voluntary attention: linked or independent? *Cognitive*
473 *Brain Research*, 18(1), 102–105.
- 474 Hunt, E., Pellegrino, J. W., & Yee, P. L. (1989). Individual differences in attention. *Psychology of Learning and*
475 *Motivation*, 24, 285–310.
- 476 Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2,
477 194–203.
- 478 Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: attention, awareness, and
479 control. *American Psychologist*, 47(6), 802–809.
- 480 LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in
481 short-term memory? it's a matter of attention. *Frontiers in Human Neuroscience*, 8, 5.
- 482 Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., ... Norman, K. A. (2018).
483 A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*,
484 180, 243–252.
- 485 McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: the time course of
486 recognition. *Journal of Experimental Psychology: General*, 118, 346–373.
- 487 Moher, J., Lakshmanan, B. M., Eggeth, H. E., & Ewen, J. B. (2014). Inhibition drives early feature-based
488 attention. *Psychological Science*, 25(2), 314–324.

- 489 Morrison, A. B., Conway, A. R. A., & Chein, J. M. (2014). Primacy and recency effects as indices of the focus
490 of attention. *Frontiers in Human Neuroscience*, 8(6), 1–14.
- 491 Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory and Cognition*, 21, 689–698.
- 492 Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern
493 analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430.
- 494 Noudoost, B., Chang, M. H., Steinmetz, N. A., & Moore, T. (2010). Top-down control of visual attention.
495 *Current Opinion in Neurobiology*, 20(2), 183–190.
- 496 O’Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional
497 selection. *Nature*.
- 498 Owen, L. L. W., Chang, T. H., & Manning, J. R. (2021). High-level cognition during story listening is
499 reflected in high-order dynamic correlations in neural activity patterns. *Nature Communications*, 12(5728),
500 doi.org/10.1038/s41467-021-25876-x.
- 501 Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in
502 Cognitive Sciences*, 6(2), 93–102.
- 503 Peirce, J. W., Gray, J., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., … Lindeløv, J. (2019).
504 PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203.
- 505 Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The feret database and evaluation procedure for
506 face-recognition algorithms. *Image and Vision Computing*, 16(5), 295–306.
- 507 Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- 508 Posner, M. I., Walker, J. A., & adn R D Rafal, F. A. F. (1987). How do the parietal lobes direct covert attention.
509 *Neuropsychologia*, 25(1), 135–145.
- 510 Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behavior. *Nature Reviews
511 Neuroscience*, 13, 713–726.
- 512 Salinas, E., & Thier, P. (2000). Gain modulation: A major computational principle of the central nervous
513 system. *Neuron*, 27(1), 15–21.
- 514 Schneider, K. A. (2011). Subcortical mechanisms of feature-based attention. *The Journal of Neuroscience*,
515 31(23), 8643–8653.

- 516 Soto, D., & Blanco, M. J. (2004). Spatial attention and object-based attention: a comparison within a single
517 task. *Vision Research*, 44(1), 69–81.
- 518 Stoppel, C. M., Boehler, C. N., Sabelhaus, C., Heinze, H.-J., Hopf, J. M., & Shoenfeld, M. A. (2007). Neural
519 mechanisms of spatial- and feature-based attention: a quantitative analysis. *Brain Research*, 1181(21),
520 51–60.
- 521 Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque
522 visual cortex. *Nature*, 399(6736), 575–579.
- 523 Turk-Browne, N. B., Golomb, J. D., & Chine, M. M. (2013). Complementary attentional components of
524 successful memory encoding. *NeuroImage*.
- 525 Uncapher, M. R., Hutchinson, J. B., & Wagner, A. D. (2011). Dissociable effects of top-down and bottom-up
526 attention during episodic encoding. *The Journal of Neuroscience*.
- 527 Vartanian, O., Martindale, C., & Kwiatkowski, J. (2007). Creative potential, attention, and speed of infor-
528 mation processing. *Personality and Individual Differences*, 43(6), 1470–1480.
- 529 Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition
530 memory. *Journal of Mathematical Psychobiology*, 3, 316–347.
- 531 Wittig, J. H., Jang, A. I., Cocjin, J. B., Inati, S. K., & Zaghloul, K. A. (2018). Attention improves memory
532 by suppressing spiking-neuron activity in the human anterior temporal lobe. *Nature Neuroscience*, 21,
533 808–810.
- 534 Xiao, J., Hays, J., Ehinger, K., Oliva, A., & Torralba, A. (2010). SUN database: large-scale scene recognition
535 from abbey to zoo. In *IEEE conference on computer vision and pattern recognition*.
- 536 Yi, D. J., Kelley, T. A., Marois, R., & Chun, M. M. (2006). Attentional modulation of repetition attenuation
537 is anatomically dissociable for scenes and faces. *Brain Research*.
- 538 Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal
539 of Memory and Language*, 46, 441–517.