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

Kirsten Ziman[1,2], Madeline R. Lee[1], Alejandro R. Martinez[1], Ethan D. Adner[1], and Jeremy R. Manning[1,†]

[1]Dartmouth College

[2]Princeton University

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

March 25, 2023

Abstract

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

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

Introduction

Our brains' cognitive systems detect and exploit patterns in our prior and ongoing experiences, enabling us to function and adapt in an ever-changing world. However we do not attend to or treat all types of remembered or incoming information equally, and our ability to flexibly adapt our thinking and behaviors can vary markedly with the specific set of concepts or tasks relevant to a given setting or situation (**????**). There is also substantial variability across people with respect to which aspects of experience (sensory, social, emotional, etc.) are noticed, discriminated between, and acted upon (**?**). This implies that the same physical (objective) experience may give rise to very different perceived (subjective) experiences across people (**??**).

The aspects of our experience we attend may be under our volitional control or may be unconscious or automatic (**?**). Both volitional and unconscious attention may be expressed overtly, for example through intentional eye movements (**?**) or covertly, without any volitional physical change (**?**). Prior work has explored the similarities and differences in the neural basis of overt versus covert attention (**??**) as well as the behavioral and neural underpinnings of volitional versus unconscious attention (**?**) and their differential effects on memory. There is a general consensus that sustained volitional attention enhances memory relative to unconscious attentional processes (**??**). However, volitional attention takes many forms, such as attention to particular spatial locations or attention to particular visual features or other stimulus properties. How different *types* of volitional attention combine (or compete) to enhance memory remains an open question. Volitional covert attention is of particular interest in that it allows us to dynamically and intentionally manipulate our experience, even when our sensory input remains largely static (i.e., constant physical stimuli, retinal image, etc. **??**).

Here we examine the ways different types of volitional covert attention interact to affect memory. We designed an experimental paradigm (following **?**) that asked participants to attend to a series of presented composite image pairs while keeping their gaze fixed on a central point. The image pairs comprised a left and right image, each constructed by blending an image of a face and place. The stimuli and presentation durations were constant across the two experiments, but the experiments differed in how often we asked participants to change the focus of their attention with respect to image category (face versus place) and image location (left versus right). After the participants attended to a series of images, we used a recognition memory test to assess which aspects of the presented images had been encoded into memory. In both experiments we found that the images participants covertly attended to were better recognized than other images, supporting the notion that attention enhances memory encoding (i.e., they rated attended images as more familiar than unattended images **?**). After maintaining the focus of attention to a single image category and location (Sustained Attention Experiment), participants also recognized the attended-category image at the unattended location, and (to a lesser extent) the unattended-category image at the attended location. After more rapidly varying their focus of attention (Variable Attention Experiment), participants showed a similar boost in recognition for the unattended-category image at the attended location, but they did not recognize images at the unattended location. This suggests that participants were able to shift the location of their covert attentional focus more rapidly than they were able to shift their focus of covert attention to stimulus features. We also found differences in the timecourses of these memory effects, suggesting that the impact of location-based attention on memory persists on the order of several seconds longer than the impact of feature-based attention.

Materials and methods

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

Participants

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

We used a voluntary pre-experimental survey to collect self-reported demographic information about each partipant. All 53 participants elected to fill out the survey. Participants ranged in age from 18–21 years old (mean: 18.7 years; standard deviation: 0.8 years). Participants reported their genders as female (34 particiapnts), male (18 participants), and gender non-binary (1 participant). Participants reported their ethnicities as Not Hispanic or Latino (44 particiapnts), Hispanic or Latino (7 participants), or declined to respond (2 participant). Participants reported their race as White (37 participants), Asian (13 participants), American Indian or Alaska Native (4 participants), Black or African American (2 participants), and Other (1 participant). Note that each participant could report one or more racial categories, as they deemed

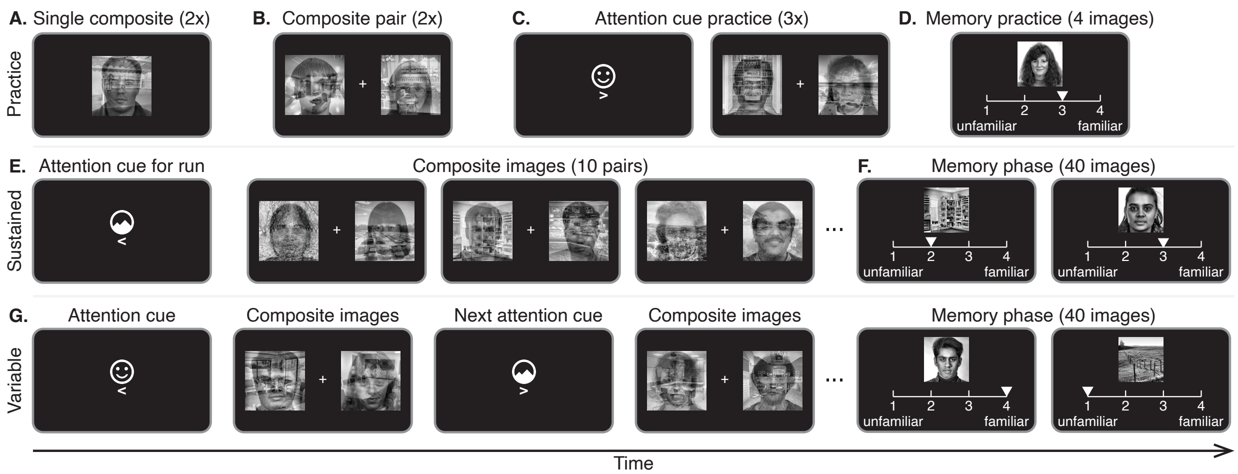

Figure 1. **Experimental paradigm. A.–D. Practice phase. A.** Composite face/place image. **B.** A single pair of composite images and a central fixation cross. **C.** One attention cue practice trial. **D.** Familiarity judgement practice trial. **E. Sustained attention condition.** Participants receive 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).

appropriate.

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

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

Stimulus selection and presentation

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

In addition to the face and place images, we presented (in white) attention cues to direct the participant's focus of attention. The attention cues comprised a stylized icon of a face or mountain peaks, directing attention to the face or place component of the images, respectively; and a left- or right-facing angled bracket, directing attention to the left or right image, respectively (e.g., Figs. **??**C, E, and G).

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

Eyetracking

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

Experimental paradigm

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

Practice phase

Several participants in pilot versions of our experiments reported that they found it difficult to modulate the focus of their attention quickly on command. We therefore designed a practice sequence to orient the participant to the process of quickly modulating the focus of their attention without moving their eyes. The experimenter remained in the testing room throughout the practice phase and answered any questions about the experiment. The practice sequence builds up incrementally to provide a gradual on-ramping for the participant prior to beginning the main experimental tasks that we focused on in our analyses.

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

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

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

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

Practice reaction time probe. After practicing modulating their focus of attention to a series of composite image pairs, we introduced a reaction time probe after each image presentation, whereby we presented either an $\times$ or $\circ$ on either the left or the right of the screen (not shown). We asked the participant to press the 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 did not impose a time limit on their responses, other than asking participants to respond as quickly as they were able. Participants practiced three trials of modulating their focus of attention to a pair of composite images (3 s), and reacting as quickly as possible to the $\times$ or $\circ$ symbol presented after each composite image pair. The reaction time probe was intended to keep participants continually engaged in modulating the focus of their attention.

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

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

Task blocks

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

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

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

Memory phase (Fig. ??F). After the presentation phase of each task block, we asked the participant to rate the familiarity (on a 1–4 scale, as during the practice phase) of a succession of face and place images. Each image was preceded by a 1 s fixation cross, and participants had up to 2 s to input their rating of each image. Participants made a total of 40 familiarity judgements, about 20 face images and 20 place images. Of these, 20 of the images (10 faces and 10 places) were drawn randomly from the (attended and unattended) composite images that the participant had viewed during the presentation phase. The remaining 20 images

(10 faces and 10 places) were novel images that the participant had not encountered during any part of the experiment. At the end of each memory block, the participant was given the opportunity to take a short break. When they were ready to continue with the next task block, they indicated their readiness to the experimenter. The experimenter then entered the testing room, re-ran the eyetracker calibration sequence, and exited the testing room prior to the next task block.

JRM STOPPED HERE

Results

We ran two volitional covert attention experiments; in the Sustained Attention Experiment we asked participants to *sustain* the focus of their attention over a succession of 10 stimulus presentations per block whereas in the Variable Attention Experiment we asked participants to *vary* their focus of attention with each new stimulus (also for a total of 10 stimulus presentations per block). Each stimulus comprised a pair of composite images (one on the left and one on the right side of the display), where each composite comprised an equal blend of a unique face and a unique place image. We followed the presentation phases of each experimental block with a memory phase, where participants performed a recognition memory task by rating the familiarity of previously experienced and novel face and place images (see *Materials and methods*, Fig. **??**).

We considered each image we probed in the memory phase with respect to whether (and how) the participant had experienced and/or attended to that image during the presentation phase (Fig. **??**). For example, if a given face image had been presented as part of the composite image on the right of the screen following an `attend to the face on the right` instruction, then that face image would have been the focus of the participant's attention during one stimulus presentation. The face image on the left (unattended side) during that trial would have been captured by feature-based attention, but not by location-based attention; and the place image (unattended category) on the right during that trial would have been captured by location-based attention, but not by feature-based attention. The place image on the left during that trial remained outside of the focus of both feature-based and location-based attention. In this way, we categorized each of the images that participants experienced during the presentation phases of each experiment in terms of whether they fell under the scope of feature-based and/or location-based attention. We also categorized novel face and place images that we asked participants to judge during the memory phase (i.e., images that the participant hadn't seen before) as belonging to the attention-cued or uncued category. These novel images were intended to serve as a baseline for comparison. For example, we sought to measure how people rated the familiarity of new images in general. We also used differences

Figure 2. **Familiarity ratings for attended, unattended, and novel stimuli.** Each split violin plot displays the distribution of within-participant average familiarity ratings given to faces (left, darker colors) and scenes (right, lighter colors) during the memory phases of each experiment. As shown in the legend (bottom), the colors indicate whether each image had been viewed during the presentation phase (`old`) or not; whether the given images matched the `attended category`; and/or whether the given images matched the `attended location`. The colored lines above each set of violin plots denote statistical differences (**pos**itive or **neg**ative differences in mean, collapsing over image category, assessed via two-tailed *t*-tests) between the distributions centered on the endpoints of each line. The line thicknesses denote *p*-values as indicated in the legend. Asterisks denote differences between the face versus scene distributions (assessed via two-tailed *t*-tests). Panel **a.** displays results from the Sustained Attention Experiment and panel **b.** displays results from the Variable Attention Experiment. Figure S1 displays these results broken down by participant cohort.

in familiarity ratings between novel images of the attention-cued versus uncued categories to evaluate potential *prospective* affects of modulating the focus of feature-based attention.

Participants in both experiments rated stimuli they attended as more familiar than unattended or novel stimuli (Fig. **??**; Sustained Attention Experiment: $t(59) = 13.42, p = 1.39 \times 10^{-19}, d = 1.73$; Variable Attention Experiment: $t(52) = 12.40, p = 3.70 \times 10^{-17}, d = 1.70$). Participants in both experiments also rated attended scene stimuli as more familiar than attended face stimuli (Sustained Attention Experiment: $t(59) = 7.69, p = 1.85 \times 10^{-10}, d = 0.99$; Variable Attention Experiment: $t(52) = 7.08, p = 3.63 \times 10^{-9}, d = 0.97$). These findings indicate that aspects of experience that are captured by the focus of attention are easier to recognize later, and that attention may preferentially benefit memory for some aspects of experience to a greater extent than other aspects of experience.

Participants in the Sustained Attention Experiment experiment also rated unattended images that matched the attended category as more familiar than unattended images that did not match the attended category (*incidental feature-based attention*; $t(59) = 7.17, p = 1.40 \times 10^{-9}, d = 0.93$). This pattern did not hold for participants in the Variable Attention Experiment ($t(52) = 0.44, p = 0.66, d = 0.06$). Participants in both experiments displayed an incidental effect of location-based attention, whereby they rated unattended-category images at the attended location as more familiar than unattended-category images at the unattended location (Sustained Attention Experiment: $t(59) = 4.32, p = 6.05 \times 10^{-5}, d = 0.56$; Variable Attention Experiment: $t(52) = 2.57, p = 0.01, d = 0.35$). This indicates that aspects of our experience that incidentally overlap with the focus of our location-based attention are easier to recognize later, even if those aspects do not fall within the intentional focus of our feature-based attention.

We defined the *memory benefit of feature-based attention* as the difference in mean familiarity ratings of attended stimuli and unattended stimuli from the attended category versus the mean familiarity ratings of unattended stimuli from the unattended category. Similarly, we defined the *memory benefit of location-based attention* as the difference in mean familiarity ratings of attended stimuli and unattended stimuli from the attended location versus the mean familiarity ratings of unattended stimuli from the unattended location. Participants in the Sustained Attention Experiment displayed a larger memory benefit of feature-based attention than location-based attention ($t(59) = 3.25, p = 1.91 \times 10^{-3}, d = 0.42$), whereas participants in the Variable Attention Experiment displayed a larger memory benefit of location-based attention ($t(52) = 1.70, p = 9.99 \times 10^{-2}, d = 0.23$). This appeared to be driven in part by a response bias towards attended-category images. For example, relative to the most recent attention cue, participants in the Sustained Attention Experiment rated attended-category novel images as more familiar than unattended-category novel images ($t(59) = 7.04, p = 2.35 \times 10^{-9}, d = 0.91$), whereas participants in the Variable Attention Experiment responded similarly to attended-category and unattended-category novel images ($t(52) = 0.23, p = 0.82, d = 0.03$). The size of this attended-category response bias for novel images was reliably smaller than the memory benefits of feature-based attention in both experiments (Sustained Attention Experiment: $t(59) = 7.59, p = 2.68 \times 10^{-10}, d = 0.98$; Variable Attention Experiment: $t(52) = 6.74, p = 1.31 \times 10^{-8}, d = 0.93$). This indicates that tuning the focus of feature-based attention with respect to how ongoing experience is encoded in memory takes longer than the duration of a single stimulus presentation (including fixation, roughly 5.5 s). Further, once feature-based attention comes "online," it affects how *new* stimuli are processed.

We also compared the relative sizes of the memory benefits of feature-based attention and location-based attention across the two experiments. Participants in the Sustained Attention Experiment displayed a larger difference between these two effects than did participants in the Variable Attention Experiment ($t(111) = 3.48, p = 7.24 \times 10^{-4}, d = 0.66$). Further, the difference in mean familiarity ratings of attended-location images versus unattended images was greater for participants in the Variable Attention Experiment ($t(111) = 2.32, p = 0.02, d = 0.44$). The memory benefit of location-based attention reliably distinguished between attended-category versus unattended category images for participants in both experiments (Sustained Attention Experiment: $t(59) = 9.50, p = 1.68 \times 10^{-13}, d = 1.23$; Variable Attention Experiment: $t(52) = 9.72, p = 2.76 \times 10^{-13}, d = 1.33$). This indicates that tuning the focus of location-based attention with respect to how ongoing experience is encoded into memory occurs faster than the duration of a single stimulus presentation.

We also compared reaction times to probes from the attended versus unattended side. Participants in both experiments responded slightly faster on average to attended-side probes, although the differences did not reach significance (Sustained Attention Experiment: $t(59) = 1.03, p = 0.31, d = 0.13$; Variable Attention Experiment: $t(52) = 0.65, p = 0.52, d = 0.09$). Participants in the Sustained Attention Experiment displayed a slightly larger (and not statistically significant) difference in attended-side versus unattended-side reaction times than participants in the Variable Attention Experiment ($t(111) = .87, p = 0.38, d = 0.20$). Because none of these reaction time effects were statistically reliable, we are unable to draw meaningful conclusions about reaction times to the attention probes on the attended versus unattended side, nor are we able to draw meaningful conclusions about differences in those reaction times that might be due to sustained versus variable attention.

Given that participants in both experiments rated attended-location images from both the attended and unattended category as more familiar than unattended-location images, we wondered whether this was truly a consequence of covert attention, or whether it might be driven solely by where participants were looking. In other words, we sought to distinguish the memory effects of the focus of covert attention from the focus of visual gaze. We used eyetracking to measure the horizontal positions of the participants' visual fixations while each pair of composite stimuli appeared onscreen (Fig. **??**). For 94% of presentation trials (Sustained Attention Task: $\frac{4531}{4800}$ trials; Variable Attention Task: $\frac{3972}{4240}$ trials), we collected viable gaze data for analysis. In the remaining trials, participant movement or other technical issues corrupted the gaze data. We found that, though they generally followed our instruction to keep their gaze fixed on the center of the screen, participants in both experiments exhibited a slight tendency to look towards the side of the screen they were attending to (Sustained Attention Task: $t(59) = 3.61, p = 6.31 \times 10^{-4}$; Variable Attention Task: $t(52) = 2.86, p = 6.06 \times 10^{-3}$). When we limited our analyses to trials with viable gaze data where participants successfully maintained fixation on the center of the screen within 4.28° visual angle (i.e., between the rightmost boundary of the left image and the leftmost boundary of the right image) we observed (numerical) memory benefits of feature-based attention in both experiments (Sustained Attention Experiment: $t(59) = 4.73, p = 1.96 \times 10^{-5}, d = 0.67$; Variable Attention Experiment: $t(52) = 0.64, p = 0.52, d = 0.09$) and location-based attention in the Sustained Attention Experiment, but not the Variable Attention Experiment (Sustained Attention Experiment: $t(59) = 0.67, p = 0.51, d = 0.09$; Variable Attention Experiment: $t(52) = -0.79, p = 0.43, d = 0.11$). These analyses indicate that although the focus of covert attention can affect where people are looking, covert attention affects memory encoding beyond what may be accounted for by gaze position alone.

The preceding results indicate that volitional covert attention affects memory, and that feature-based and location-based covert attention may come online at different timescales. We next sought to examine how long the memory effects of feature-based and location-based covert attention persist. To gain additional insights into the timecourse of the impact of feature-based and location-based attention on memory, we used figures/exps1_2_xgaze.pdf

Figure 3. **Horizontal coordinates of visual fixations.** Each violin plot displays a distribution of participant-wise average horizontal gaze positions relative to a presented image. The coordinates have been normalized such that a value of 1.0 denotes the furthest onscreen coordinate from the central fixation point *towards* the direction of the given image, and -1.0 denotes the furthest onscreen coordinate from the central fixation point *away from* the direction of the given image. The gray bars in each panel mark the boundaries of the presented images. The violin plots are broken down by participants' familiarity ratings during the memory phases of each experiment, and by each image's relation to the attention cue while that image appeared onscreen. Panel **a.** displays fixation data for the Sustained Attention Experiment and Panel **b.** displays fixation data for the Variable Attention Experiment. Figure S2 displays these results broken down by participant cohort.

a sliding window analysis to measure how participants' familiarity ratings of attended, partially attended, unattended, and novel stimuli varied over the duration of the memory phases of both experiments. During each block within the memory phases of each experiment, participants made familiarity judgements about a total of 40 images (20 old images and 20 novel images). We computed average familiarity ratings for images appearing in each position of the memory stimulus sequence (positions 1–40), as a function of how (or whether) they were attended during the preceding presentation phase (old images) or whether they matched the category of the most recent attention cue (novel images), for overlapping 20-image sliding windows (Figs. **??**, **??**, and **??**). We first calculated these values for each participant, and then we averaged across participants.

Throughout the memory phase, participants from both experiments rated attended images as more familiar than category-matched unattended and novel images (Fig. **??**, top row). This indicates that the memory benefits of attending to specific stimulus features (versus category-general features) persist for at least as long as the duration of the memory phase (2 min). Early in the memory phase, participants in the Variable Attention Experiment also rated category-matched unattended images as more familiar than novel images. This pattern did not hold for participants in the Sustained Attention Experiment. The short-lived boost in familiarity for attended-category images at the unattended location following variable (but not sustained) attention suggests that the timecourse of feature-based attention's impact on memory depends on how long the focus of feature-based attention has been held.

Participants in both experiments also rated unattended-category images at the attended location as more familiar than unattended-category images at the unattended location and novel unattended-category images (Fig. **??**, bottom row). This pattern persisted throughout the memory phases of both experiments. This supports the notion that location-based attention enhances memory in a (partially) feature-independent way. Participants in the Variable Attention Experiment also rated unattended-category images at the unattended location as (numerically) more familiar than novel images (though this pattern was only statistically reliable towards the middle of the memory phase). This pattern did not hold for participants in the Sustained Attention Experiment. Participants in the Sustained Attention Experiment rated unattended-category images (regardless of location or novelty status) as less familiar than participants in the Variable Attention Experiment (e.g., compare heights of the lines in the bottom left versus bottom right panels of Fig. **??**). This suggests that, after sustained feature-based attention to a particular image category, the encodings of stimuli from the unattended category are suppressed. This suppression effect appears to persist at least as long as the memory phase of the experiment (2 min). Further, the magnitude of this suppression effect appears to become stronger following longer-term sustained (versus shorter-term variable) feature-based attention.

We next examined how participants' familiarity ratings varied over the duration of the memory phases of figures/Cat_Cued_TimeSeries.pdf

Figure 4. **Familiarity ratings over time for images that matched the cued or uncued image category.** Each curve reflects the average familiarity ratings for attended, unattended, and novel images (denoted in the legends on the right) within a succession of overlapping 20-image sliding windows. Error ribbons denote 95% confidence intervals, computed across participants. Panels **a.** and **c.** display results from the Sustained Attention Experiment and Panels **b.** and **d.** display results from the Variable Attention Experiment. The paired horizontal lines at the bottom of each panel denote timepoints when the given pair of curves was statistically distinguishable (i.e., the topmost line color was statistically greater than the bottommost line color at $\alpha = 0.05$, via a paired two-tailed t-test.) The gray lines in Panels b and d reflect familiarity ratings of novel stimuli from both the most recently attended and unattended image categories. Figures S3 and S4 display these results broken down by participant cohort.

figures/Side_Cued_TimeSeries_2.pdf

Figure 5. **Familiarity ratings over time for images that matched the cued or uncued image location.** This figure follows the same format and conventions as Figure **??**. Figures S5 and S6 display these results broken down by participant cohort.

each experiment as a function of the attended location. Participants in the Sustained Attention Experiment rated attended-category images as more familiar, at both the attended and unattended locations (Fig. **??**, left panels). In contrast, participants in the Variable Attention Experiment rated attended-category images as more familiar than unattended-category images only at the attended locations (Fig. **??**, right panels). The primary difference in familiarity ratings between Sustained Attention Experiment versus Variable Attention Experiment participants was in how they rated unattended-category images. Participants in the Variable Attention Task rated unattended-category images as more familiar than did participants in the Sustained Attention Task; this pattern persisted throughout the memory phase and held for images at both the attended and unattended locations (compare heights of blue lines in the left versus right panels of Fig. **??**). This suggests that lower familiarity ratings of unattended-category stimuli in the Sustained Attention Experiment are due to suppression of the encoding of unattended-category features (e.g., as opposed to enhancement of the encoding of attended-category features). This suppression effect appears to build over an interval that is longer than the duration of a single stimulus presentation (5.5 s).

Finally, we examined how participants in both experiments rated the familiarity of novel images. These ratings provide insights into how modulating the focus of attention to *prior* stimuli can affect the perception of *future* stimuli. We considered how participants rated novel stimuli that matched versus did not match the category (face or scene) of the most recent attention cue. Participants in the Sustained Attention Experiment rated attended-category novel stimuli as more familiar than unattended-category novel stimuli throughout the entire memory phase (Fig. **??**a). By contrast, participants in the Variable Attention Experiment rated attended-category and unattended-category novel images similarly throughout the memory phase (Fig. **??**b). The difference in familiarity ratings between the two experiments primarily reflects that participants in the Variable Attention Experiment rated unattended-category novel images as more familiar than did participants in the Sustained Attention Experiment (e.g., compare blue lines in Fig. **??**a versus **??**b). This may indicate that the processing of unattended-category novel stimuli is suppressed following sustained attention (e.g., as opposed to enhancing the processing of attended-category novel stimuli). The differences in familiarity between attended-category versus unattended-category novel images increased slightly but reliably throughout the memory phases of both experiments (Fig. **??**, bottom panels; slope of regression line fit to differences in Sustained Attention Experiment: $\beta = 2.05 \times 10^{-3}, p = 1.07 \times 10^{-14}$; slope of regression for Variable Attention Experiment: $\beta = 1.54 \times 10^{-3}, p = 5.63 \times 10^{-13}$). This might reflect a process whereby the suppression of unattended features continues to build for a short duration even after the participant stops volitionally focusing on the attended features.

figures/Novel_Timecourse_Panels.pdf

Figure 6. **Familiarity ratings over time for attended-category and unattended-category novel images.** The top panels of this figure follow the same formats and conventions as Figures **??** and **??**. The bottom panels display the differences between the timecourses of familiarity ratings of attended-category and unattended-category novel stimuli. The error ribbons denote 95% confidence intervals, computed across participants. The dotted horizontal line in each panel denotes a difference of 0– i.e., the value above which attended-category novel images were rated as more familiar than unattended-category novel images. The black lines in the bottom panels display linear regression fits to the data (error ribbons denote 95% confidence intervals). Figures S7 and S8 display these results broken down by participant cohort.

Figure 7. **Timecourse of encoding and retrieval effects of feature-based and location-based attention.** The focus of location-based attention (green) may be modulated quickly and rapidly enhances memory encoding for nearby stimuli. The focus of feature-based attention (purple) is modulated more slowly, and serves to suppress unattended stimulus features. The effects of location-based attention on memory persist for longer than the effects of feature-based attention. Sustained attention (darker shading) yields more robust enhancement and suppression than shorter-term (variable) attention (light shading). The vertical gray line on the left denotes the duration of a single stimulus presentation (the upper bound by which location-based attention begins to affect encoding, and the lower-bound by which feature-based attention begins to affect encoding). The vertical gray line on the right separates encoding from immediate subsequent retrieval of the encoded information.

Discussion

We ran two covert attention experiments that asked participants to sustain or vary the focus of their covert attention, respectively. When participants held the focus of their feature-based (face versus scene) and location-based (left versus right) attention for a sustained interval, they judged stimuli they had seen as familiar when they overlapped with respect to the features and locations they had attended. The increase in familiarity was larger for attended-feature images than attended-location images. The increase also extended to novel stimuli from the attended image category. By contrast, when participants varied the focus of their feature-based and location-based attention more rapidly, the boost in familiarity for feature-matched stimuli was smaller than that for location-matched stimuli, and did not extend to novel stimuli. Our findings suggest that participants were able to more rapidly modulate their focus of location-based attention than their focus of feature-based attention. The tuning of location-based attention appears to be mediated by enhanced encoding and faster processing at the attended location. The tuning of feature-based attention appears to be mediated by a suppression in the encoding and processing of unattended stimulus features. This suppression effect also affects how new stimuli are processed, and it persists for a longer duration following an interval when the focus of feature-based attention was held constant over a longer duration. Taken together, our findings suggest that feature-based and location-based attention are mediated by different mechanisms and affect memory at different timescales (Fig. **??**).

The notion that location-based attention operates at a faster timescale than feature-based attention is supported by prior work on the deployment of visual attention (**??**). Our findings that location-based attention enhances the processing of attended stimuli whereas feature-based attention suppresses the processing of unattended stimuli is also consistent with prior work on location-based attention (e.g., **?**) and feature-based attention (e.g., **?**). Our finding that people better remember attended stimuli also follows prior work on interactions between attention and memory (**???????**). Whereas much of this prior work focused on elucidating the neural basis of these interactions, our work extends these prior studies by elucidating the specific and separable behavioral impacts of feature-based attention (inhibition with a slow onset) and location-based attention (enhancement with a fast onset) on subsequent memory. Both of these effects persisted throughout the 2 min memory phases of both experiments. Therefore future work is needed to elucidate the longevity of these effects beyond 2 minutes.

Another important area for future study concerns how the flow of information between different brain structures is modulated according to the focus of volitional attention– particularly with respect to pathways from primary sensory regions (e.g., V1, A1) to regions implicated in encoding ongoing experiences into memory (e.g, medial temporal lobe structures such as the hippocampus and entorhinal cortex, prefrontal cortex, etc.). For example, several studies suggest that attention serves to modulate the *gain* of specific neural circuits (**????**), effectively facilitating or inhibiting the flow of specific neural representations (**??**). Prior work suggests that feature-based attention may be supported by changes in connectivity with the thalamus (**?**), whereas location-based attention may be supported by changes in connectivity with primary visual cortex (**?**). That feature-based and location-based attention are mediated by different brain structures may explain why these different aspects of attention operate on different timescales and affect memory differently. A strong test of this hypothesis would entail directly measuring neural activity patterns as people modulate their focus of attention (e.g., using functional magnetic resonance imaging or electroencephalography), and then using neural decoding approaches (e.g., **???**) to follow how neural representations of attended (or unattended) stimuli are transfered from primary sensory regions, to higher order sensory regions, to memory areas. If the effects of attention on memory are mediated by changes in network dynamics, the transmission rates of the representations of attended stimuli from primary sensory regions to memory areas should be facilitated relative to the transmission rates of unattended stimuli. Further, variability in these neural changes (e.g., as a participant focuses their attention with more or less success) should track with behavioral measures of memorability.

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

Author Contributions

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

Data and code availability

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

Acknowledgements

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