Research

Why do we retrace our visual steps? Semantic and episodic memory in gaze reinstatement

Michelle M. Ramey,[1,2,3] Andrew P. Yonelinas,[1,2] and John M. Henderson[1,3]

[1]Department of Psychology, University of California, Davis, California 95616, USA; [2]Center for Neuroscience, University of California, Davis, California 95618, USA; [3]Center for Mind and Brain, University of California, Davis, California 95618, USA

When we look at repeated scenes, we tend to visit similar regions each time—a phenomenon known as *resampling*. Resampling has long been attributed to episodic memory, but the relationship between resampling and episodic memory has recently been found to be less consistent than assumed. A possibility that has yet to be fully considered is that factors unrelated to episodic memory may generate resampling: for example, other factors such as semantic memory and visual salience that are consistently present each time an image is viewed and are independent of specific prior viewing instances. We addressed this possibility by tracking participants' eyes during scene viewing to examine how semantic memory, indexed by the semantic informativeness of scene regions (i.e., *meaning*), is involved in resampling. We found that viewing more meaningful regions predicted resampling, as did episodic familiarity strength. Furthermore, we found that meaning interacted with familiarity strength to predict resampling. Specifically, the effect of meaning on resampling was attenuated in the presence of strong episodic memory, and vice versa. These results suggest that episodic and semantic memory are each involved in resampling behavior and are in competition rather than synergistically increasing resampling. More generally, this suggests that episodic and semantic memory may compete to guide attention.

[Supplemental material is available for this article.]

Most people have heard of retracing one's steps to find something lost, but this idea goes beyond a useful adage: Retracing one's steps may be an important part of episodic memory. For example, returning to the location in which one learned information enhances the ability to retrieve that information by reinstating the encoding context (Godden and Baddeley 1975; Smith 1979). There are also theories proposing that this effect extends to where we look—such that reinstating gaze by viewing similar regions across study and test of an image (i.e., resampling) improves episodic memory for that image—based on observations of consistent eye movement patterns between successive viewings of an image (Noton and Stark 1971). Though controversial (Henderson 2003), such theories have had a longstanding influence on the literature in visual attention and memory (Wynn et al. 2019). However, research has since indicated that the relationship of resampling with episodic memory is more tenuous than had been assumed, and the causality of the relationship has remained elusive (e.g., Locher and Nodine 1974; Holm and Mantyla 2007; Foulsham and Kingstone 2013; Valuch et al. 2013; Damiano and Walther 2019). Despite this tenuous relationship and the rapidly growing interest in resampling behavior (e.g., Wynn et al. 2019), attempts to explain what gives rise to resampling have continued to focus primarily on episodic memory as the driving factor (Holm and Mantyla 2007; Foulsham and Kingstone 2013; Valuch et al. 2013; Wynn et al. 2016, 2018; Damiano and Walther 2019), even in studies finding no relation between recognition performance and resampling (e.g., Locher and Nodine 1974; Humphrey and Underwood 2010). Because of this longstanding focus on episodic memory in the resampling literature, it is not well understood how other cognitive or visual factors may contribute to resampling behavior, and how such factors may modulate the relationship between resampling and episodic memory.

The notion that resampling image regions between successive viewings is a uniquely episodic-memory-related phenomenon began with Noton and Stark's (1971) scanpath theory. They proposed that the path traveled by the eyes during encoding was stored alongside memory representations for the visual information, and that repeating the scanpath upon subsequent viewings of the image facilitated memory retrieval. This theory was based upon observations that participants tend to produce similar scanpaths between repeated viewings of a given image, and it was assumed that this resampling behavior was a uniquely memory-related phenomenon—without any direct test of whether the repeated scanpaths related to episodic memory. In fact, the first direct tests found no relationship between resampling and memory accuracy (e.g., Locher and Nodine 1974; Humphrey and Underwood 2010). Only recently has evidence surfaced for a relationship, albeit weak, between resampling and recognition memory performance (Mantyla and Holm 2006; Holm and Mantyla 2007; Foulsham and Kingstone 2013; Valuch et al. 2013; Damiano and Walther 2015, 2019; Wynn et al. 2016; Ramey et al. 2020). Furthermore, most of the evidence thus far has been correlational. The few causal studies that have been done have found consistent evidence for an influence of episodic memory on resampling behavior, but often find no significant influence of resampling on episodic memory (Holm and Mantyla 2007; Foulsham and Kingstone 2013; Valuch et al. 2013; Damiano and Walther 2015)—despite the assumptions of scanpath theory and

Corresponding author: mmramey@ucdavis.edu
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.051227.119.

other similar theories that resampling improves episodic memory through an iterative, bidirectional process (Noton and Stark 1971).

Thus far, the evidence suggests that episodic memory and resampling are related, and that this is primarily driven by stronger memory causally increasing the extent to which similar image regions are viewed. However, a substantial amount of resampling behavior occurs that is not explained by variations in recognition memory strength (Ramey et al. 2020), which suggests that there may be additional driving forces behind resampling. One possibility is that fixating similar regions between repeated viewings of an image could simply reflect consistent guidance by factors that are present irrespective of whether the image is remembered, such as other known sources of influence on eye movement behavior (e.g., visual salience and general world knowledge; see Henderson 2003). For example, viewers tend to look at semantically informative scene regions (e.g., regions containing objects) more than uninformative regions (e.g., empty regions; Henderson and Hayes 2017). If they tend to do this for both initial and repeated viewings, then what might appear to be memory-based resampling could instead simply be due to selection of the same regions independently across viewings. Despite the intuitive appeal of and indirect support for such possibilities, they have yet to be directly investigated. The present study is thus aimed at determining whether additional factors, primarily the semantic informativeness of scene regions, are involved in resampling behavior, and whether such factors may influence the extent to which episodic memory predicts resampling.

A direct measure of the spatial distribution of semantic information in scenes was recently developed by constructing *meaning maps* based on participants' ratings of different parts of each scene (Henderson and Hayes 2017). Meaning maps are able to capture the spatial distribution of potentially useful information in a scene, such as objects and people, and thus might be expected to provide a good estimate of where people tend to look. Indeed, quantification of the amount of meaning contained in image patches has been found to quite accurately predict the spatial distribution of attention (Henderson and Hayes 2017, 2018). As described above, meaning may naturally lead to resampling behavior because the distribution of meaningful regions in scenes is unchanged between viewings, and one would thus expect attention to be driven to meaningful information similarly between viewings.

Given that meaning indexes the semantic informativeness of scene regions, it may serve as a measure of semantic memory. That is, objects are only informative insofar as we have learned that they are informative, through a lifetime of accruing semantic knowledge about the world (Tulving 1986; Saffran and Schwartz 1994). Therefore, the fact that people tend to direct their attention toward meaningful regions suggests that semantic memory consistently influences attention during naturalistic viewing. This possibility may prove particularly relevant for the resampling literature—given its longstanding focus on episodic memory—as there is emerging evidence from the visual search literature that semantic and episodic memory can interact to influence attention. For example, in addition to studies showing that episodic and semantic memory are each able to guide viewing behavior generally (Henderson 2003; Neider and Zelinsky 2006; Ryan and Shen 2020; Wynn et al. 2020), one study suggests that decreasing the extent to which semantic memory is available to guide search leads to an increase in reliance on episodic memory (Võ and Wolfe 2013). This indicates that semantic and episodic memory may trade off or compete in their guidance of attention during search. It is therefore possible that the semantic meaning of viewed scene regions may influence the relationship between episodic memory and resampling behavior as well.

In sum, the evidence thus far indicates that episodic memory is indeed related to resampling behavior, such that stronger

episodic memory increases the extent to which regions are resampled (Holm and Mantyla 2007; Foulsham and Kingstone 2013; Valuch et al. 2013; Damiano and Walther 2015). Despite this, much of resampling behavior remains unexplained by episodic memory; the resampling phenomenon is being hotly investigated in the memory literature, but it is not yet well understood why people tend to revisit regions between study and test. As mentioned above, however, there is also indirect support for the possibility that semantic memory may be able to produce resampling behavior (Henderson and Hayes 2017), and that it may interact with episodic memory to do so (Võ and Wolfe 2013; Wynn et al. 2020).

Current research

To address these possibilities, we examined how resampling behavior is related to the semantic informativeness of viewed scene regions, and whether directing attention toward semantically informative regions may modulate how episodic memory is involved in resampling. Furthermore, we examined resampling on a more fine-grained level than in prior work by developing a new fixation-by-fixation measure for assessing the extent to which each retrieval fixation was near regions that had been visited during encoding. This new measure (i.e., *refixation distance*) also allows fixation-by-fixation trends in how spatial resampling is related to memory to be examined for the first time; that is, it allows for assessment of how resampling varies over the course of a trial. We assessed recognition memory using a confidence-based memory scale, to allow for a sensitive assessment of memory strength (Ramey et al. 2019), rather than previously used dichotomous old/new judgments.

In the present experiment, participants viewed a series of real-world scenes while their eye movements were tracked. During an initial study phase, participants viewed scenes in two encoding tasks. In one encoding task, participants were asked to memorize each scene, whereas in the other, they were asked to judge each scene for its aesthetic appeal. Two encoding tasks were included to test the generalizability of any effects obtained as well as to verify that the effects were not limited to conditions in which participants intentionally encoded the scenes. During a subsequent test phase, participants viewed the same scenes that they had viewed during the study phase (i.e., old scenes) along with randomly intermixed new scenes, and were asked to provide a recognition judgment for each scene. Recognition memory awareness was measured by asking participants to rate their memory confidence for each scene on a six-point scale during the recognition judgment (Yonelinas 2002). Participants were told that if they could consciously recollect some qualitative aspect of the initial learning event, such as what they thought about when the scene was encountered earlier, they should respond "Recollect old (6);" otherwise, they rated their memory confidence by responding "I'm sure it's old (5)," "Maybe it's old (4)," "I don't know (3)," "Maybe it's new (2)," or "I'm sure it's new (1)." In a prior study of scene memory, we found that trial-by-trial resampling was consistently related to familiarity strength, but not recollection (Ramey et al. 2020); therefore, we focused the present analyses on the continuous gradient of familiarity-based memory responses from "sure new" to "sure old." However, for the sake of completeness, we also verified that the effects held when using all responses (see Supplemental Material), and we present the recollect responses in our data figures.

Semantic informativeness was quantified using meaning maps (Henderson and Hayes 2017), which capture the spatial distribution of semantic information across a scene. Attention to meaning was used as an index of attentional guidance by semantic memory and was computed on a fixation-wise basis by determining the average amount of meaning contained in the region

immediately surrounding each fixation. Furthermore, to ensure that attention to meaningful regions was not potentially confounded by bottom-up visual saliency, we also ran analyses that controlled for saliency and examined the effects of saliency on resampling (see Supplemental Material).

Because the meaning of scene regions has been shown to be a potent driver of attention, and there is no evidence to our knowledge suggesting this would not be the case across multiple viewings, we hypothesized that fixating highly meaningful regions would be a strong predictor of resampling. The potential for interaction between semantic meaning and episodic memory is less clear, but a variety of outcomes would be of theoretical relevance. First, it is possible that semantic meaning and episodic memory may have a synergistic relationship in predicting resampling. This could be the case if resampling improves memory, as some theories have proposed (Wynn et al. 2019), and if attention to meaning increases resampling. A second possibility is that meaning and episodic memory may compete with each other to guide attention, such that a stronger influence of one on any given fixation might lessen the influence of the other. This would fit with the emerging evidence of potential competitive interactions between semantic and episodic memory (Võ and Wolfe 2013).

Results

Preliminary analyses

Recognition memory accuracy

The percentage of scenes that received a recognition confidence response of "recollect," "sure old," "maybe old," "don't know," "maybe new," and "sure new," respectively, were 46%, 25%, 11%, 8%, 6%, and 4% for old scenes, and 2%, 3%, 8%, 13%, 27%, and 47% for new scenes (Fig. 2A). These results suggest that participants were able to discriminate between old and new scenes and used the full range of response options.

Study task

We included two study tasks (i.e., memorization and aesthetic judgment) to ensure that any effects obtained were robust to encoding conditions. To determine if study task affected resampling behavior, we ran three models: (1) regressing refixation distance on study task, (2) regressing refixation distance on the interaction between study task and memory strength, and (3) regressing refixation distance on the interaction between study task and meaning score. There were no significant effects of study task ($Ps > 0.08$). Subsequent analyses are thus collapsed across tasks.

Primary analyses

For a schematic of how episodic memory and semantic meaning may be able to drive resampling behavior via individual fixations,
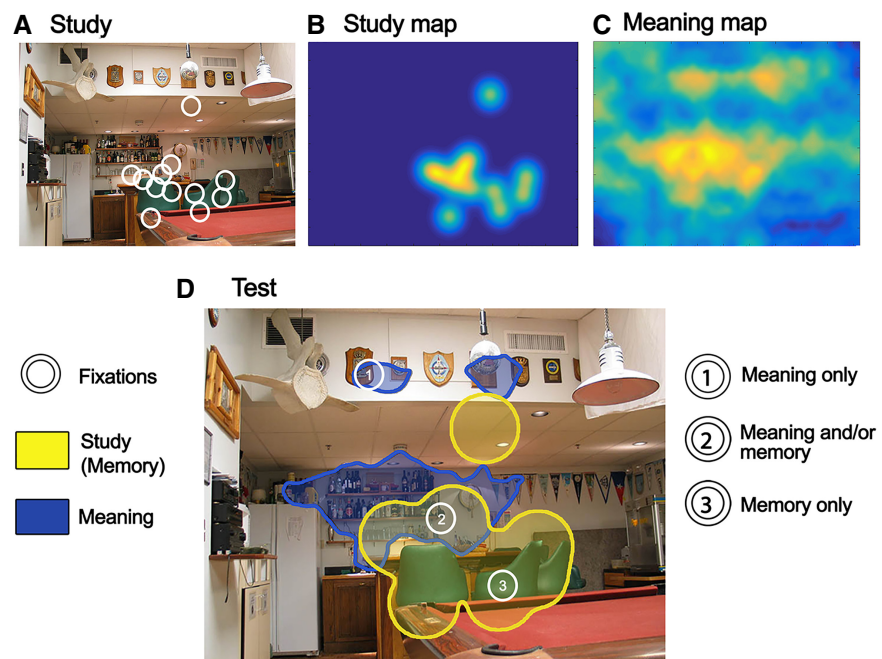

Figure 1. Conceptual framework using example fixations made during study and test of a scene. (*A*) Fixations (rings) made during study of a scene. (*B*) A smoothed heatmap of the study fixation locations from the same scene. (*C*) The meaning map of the same scene. (*D*) A combination of the scene and the brightest portions of the heatmaps in (*B,C*), to illustrate how semantic meaning and episodic memory may drive resampling. Fixation 1 represents a non-resampled fixation that was likely driven by meaning (blue). Fixation 2 is an example of resampling (i.e., is within a yellow region) that could be driven by memory or meaning given that they are overlapping. Fixation 3 is an example of resampling that was likely not driven by meaning. We hypothesized that an increase in memory strength would lead to an increase in fixations like fixations 2 and 3, whereas attention to meaning would result in fixations like 1 and 2.

see Figure 1. Note that all of the analyses below use data from old scenes in the test phase, because it was not possible to directly assess resampling in new scenes.

Episodic memory strength

To examine how resampling behavior related to recognition memory strength, we determined whether refixation distance varied across the linear gradient of "sure new" to "sure old" responses (i.e., familiarity-based memory strength). Memory strength predicted significantly decreased refixation distance, $\beta = -0.14$, $P < 0.0001$, indicating that stronger memory was related to increased resampling behavior (Fig. 2B). Note that memory strength was a trial-level measure, whereas refixation distance was a fixation-by-fixation measure; we used a nested linear mixed effects model to account for this difference, but found that the effect remained when refixation distance was aggregated by trial as well, $\beta = -0.29$, $P < 0.0001$ (and by subject; see Supplemental Material).

Semantic meaning

To determine whether the semantic informativeness of viewed scene regions predicted resampling behavior, we regressed the refixation distance of each fixation on the meaning score of each fixation in the test phase. Meaning scores were negatively related to refixation distance, $\beta = -0.28$, $P < 0.0001$, such that the tendency to resample a region was associated with the meaning of that region (Fig. 2C). This indicates that semantic memory may give rise to resampling behavior by guiding attention toward meaningful regions consistently across viewings.
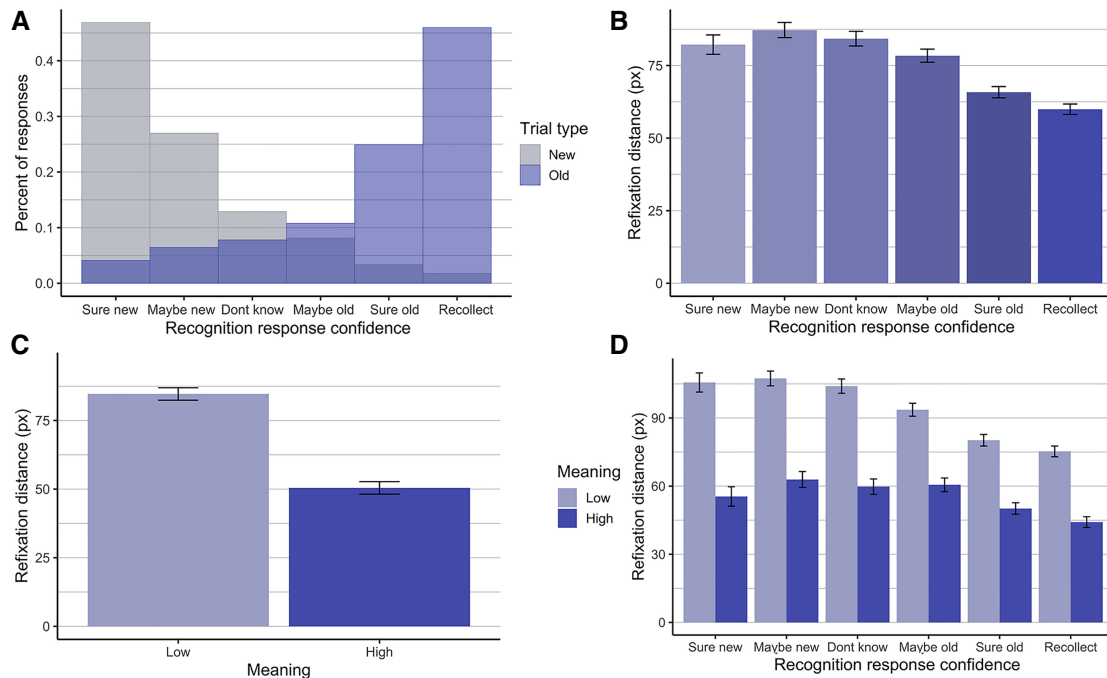
Figure 2. Results of primary resampling analyses. Note that lower refixation distance reflects more resampling behavior. (*A*) Histogram of the proportions of recognition memory responses made for old and new scenes. (*B*) Refixation distance by memory response. (*C*) Refixation distance by attention to meaning (i.e., meaning score). (*D*) Interaction between meaning score and memory response in predicting refixation distance. For (*B–D*), least-squares means derived from the linear mixed effects models used in the analyses are plotted, and the error bars represent the standard error of these estimated means from the model. For (*C,D*), meaning was dichotomized to facilitate visualization, but all analyses were done with continuous data.

Interaction between semantic meaning and episodic memory

The evidence thus far indicates that episodic memory and semantic meaning are each uniquely involved in resampling behavior: Memory strength was not related to the meaning of viewed regions (Eq. A6), and each predicted unique variance in resampling behavior (see Supplemental Material). To confirm that each variable had a main effect on resampling when the other was controlled for, we ran a model predicting refixation distance from both meaning and memory strength. As expected, memory strength, $\beta = -0.11$, $P < 0.0001$, and meaning, $\beta = -0.27$, $P < 0.0001$, each predicted resampling when the other was held constant statistically.

Given that both factors appear to be simultaneously involved in resampling, we sought to determine whether semantic meaning may modulate the relationship between episodic memory and resampling based on preliminary evidence for interactions between episodic and semantic memory (Võ and Wolfe 2013). To do this, we ran a model regressing refixation distance on the interaction between meaning score and memory strength (Eq. A7). As predicted, meaning score and memory strength interacted to predict refixation distance, $\beta = 0.06$, $P < 0.0001$, such that an increase in meaning score reduced the strength of the relation between memory strength and resampling, and vice versa (i.e., an increase in memory strength reduced the association between meaning and resampling; Fig. 2D). These results suggest that stronger semantic guidance attenuates the extent to which episodic memory strength predicts resampling, and/or that strong episodic memory may weaken the relationship between semantic memory and resampling.

Temporal analyses

Because refixation distance is a fixation-by-fixation measure of resampling, it allows for examination of how the effects of episodic memory and semantic meaning on resampling may change over the course of a trial. To determine how the effects observed thus far changed over fixations, we regressed refixation distance on the interaction between the variable of interest and the ordinal fixation number in a trial.

Episodic memory strength

The relationship between recognition memory response and refixation distance did not change over fixations in a linear fashion, $\beta = -0.003$, $P = 0.58$. However, examination of the plot (Fig. 3A; Supplemental Fig. 4) revealed what appeared to be a quadratic relationship, so we conducted an exploratory analysis of this possibility. Including a quadratic interaction term in the model, we found that the relationship between memory strength and refixation distance was significantly more pronounced toward the middle of the trial, $\beta = 0.03$, $P < 0.001$. That is, the effect of memory strength on resampling was weakest at the beginning and end of each trial, and was strongest midway through each trial.

Semantic meaning

Regressing refixation distance on the interaction between meaning score and ordinal fixation number revealed a significant linear interaction, $\beta = -0.04$, $P < 0.0001$, such that the relationship between meaning and refixation distance grew stronger over the course of the trial (Fig. 3B). Unlike memory strength, there was no significant quadratic effect, $\beta = 0.005$, $P = 0.53$.

Interaction between semantic meaning and episodic memory

We next examined how the interaction between meaning and memory strength observed above might change over time, and
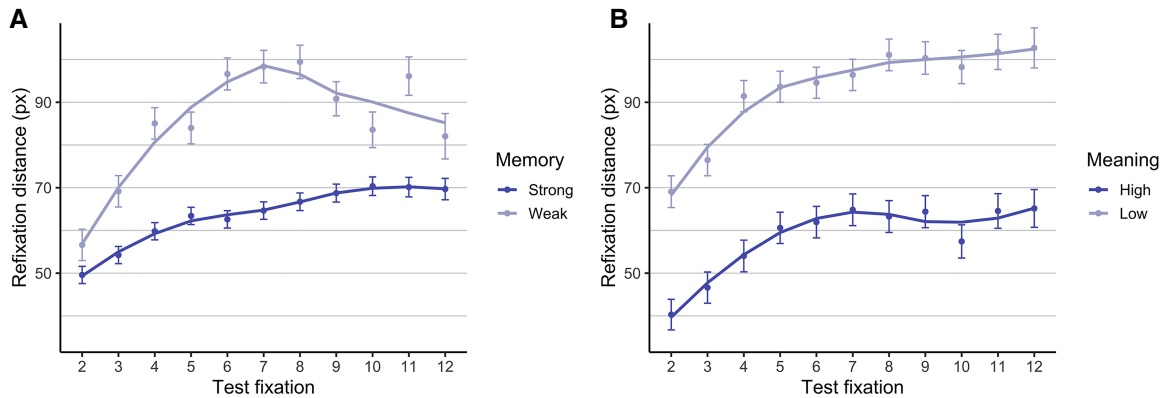
Figure 3. Results of the temporal analyses. (*A*) Refixation distance over fixations by memory strength; "strong" memory included responses of "recollect" and "sure old," and "weak" memory included responses of "sure new" and "maybe new." (All responses are shown in Supplemental Fig. 4). (*B*) Refixation distance over fixations by the meaning score of each fixation. In each plot, the x-axis is the ordinal fixation number in a test phase trial. Ninety percent of the data were included at a cutoff of 12 fixations; the plot was thus truncated at 12 fixations to reduce noise from the small number of trials containing more than 12 fixations. However, analyses included all data. Least-squares means derived from the linear mixed effects models used in the analyses are plotted, and the error bars represent the standard error of these estimated means from the model. The lines were generated using a locally weighted smoothing function, which plots local regressions to aid the eye in seeing trends. The data in the plots were dichotomized to facilitate visualization, but all analyses were done with continuous data.

found that it does not vary systematically over the course of the trial, *P*s > 0.39 (Supplemental Fig. 3; see Supplemental Material).

Taken together, the temporal analyses indicate that memory strength exhibited little effect on resampling early on in viewing, had a strong influence midtrial, and was attenuated toward the end of the trial. In contrast, meaning demonstrated an effect early in the trial that consistently increased over the course of viewing. This suggests that semantic meaning and episodic memory might have unique time courses in how they relate to resampling over the course of viewing.

Additional analyses

We ran a series of additional analyses that are presented in the Supplemental Material. First, to probe the robustness of the present effects to potentially confounding variables, we reran analyses controlling for variables such as the number of fixations per trial, potential center bias of meaning maps, and the inclusion or exclusion of recollect responses—none of which altered the pattern of results. We also created cross-subject and cross-image refixation distance baselines by randomly pairing subject and image data, and found that resampling behavior is indeed driven by both subject and image-level idiosyncrasies such as episodic memory and semantic meaning of regions, respectively. That is, the observed refixation distance values were significantly lower than the refixation distance values obtained by randomly pairing images, or by randomly pairing subjects. Furthermore, the analysis suggests that image properties—such as semantic meaning—may be responsible for the majority of resampling behavior (Supplemental Fig. 1b).

We then examined whether the bottom-up visual salience of scene regions related to resampling, and found that salience largely followed the same pattern of results as meaning. Furthermore, both meaning and salience accounted for unique variance in resampling behavior; thus, the relationship between meaning and resampling held when salience was controlled, indicating that the meaning effects observed above were driven by semantic information and not by bottom-up visual information. We also ran a combined model predicting refixation distance from memory strength, meaning, salience, center bias, and the interaction between meaning and memory strength, and found that every vari-

able significantly predicted unique variance and improved the model fit.

Discussion

In the present study, we examined how episodic and semantic memory predicted resampling behavior, operationalized as the extent to which people revisited scene regions between study and test. We tracked participants' eye movements during encoding and retrieval of scenes, and participants provided confidence-based recognition judgments for each scene during retrieval. Resampling—as well as guidance by semantic memory, indexed by attention to meaningful regions (i.e., *meaning*)—were assessed on a fixation-by-fixation basis, allowing for fine-grained analysis of temporal trends. We found that episodic memory strength and semantic meaning each predicted resampling behavior, such that stronger recognition memory and increased viewing of meaningful regions during retrieval were both related to increased resampling. Importantly, memory strength was not related to meaning, and these factors each predicted unique variance in resampling behavior. Furthermore, episodic memory strength interacted with semantic meaning to predict resampling, such that stronger semantic guidance weakened the relationship between episodic memory and resampling, and vice versa. Moreover, these effects were not due to the numbers of fixations made, center bias of viewing, nor overall similarities in viewing patterns between subjects or scenes. We also found that these effects were robust to encoding conditions (i.e., memorization and aesthetic judgment), and that the effects of meaning were not driven by visual salience; in fact, meaning and salience each predicted unique variance in resampling. Last, episodic memory and meaning had different patterns of results in how they were related to refixation distance over the course of fixations. Specifically, whereas meaning influenced viewing from the first fixation onward—and its influence continued to strengthen over the course of a trial—memory strength had little effect early on, and its effect peaked midway through the viewing period. These results suggest that episodic and semantic memory might have fundamentally different time courses in their influence on attention.

In contrast to theories assuming that resampling behavior is uniquely related to episodic memory (Noton and Stark 1971),

the present results indicate that there is a robust role for information present in the image itself—such as semantic informativeness and visual salience—in driving resampling as well. In fact, when considered in the same model, meaning was a stronger predictor of resampling than was memory strength. Additionally, computing resampling with randomly shuffled pairings of trials revealed that image content such as the meaning of regions may be a stronger contributor to resampling behavior than subject-level factors such as episodic memory. Whereas there is debate surrounding the mechanisms underlying the relationship between episodic memory and resampling (Noton and Stark 1971; Henderson 2003), the mechanisms driving the involvement of semantic memory in resampling may be more straightforward. For example, because semantic memory is known to guide attention (Henderson and Hayes 2017), it is likely that the meaning of regions predicts resampling behavior simply by guiding attention consistently with each viewing. This potential mechanism was supported by follow-up analyses indicating that increased attention to meaning at study strengthened the relationship between resampling and meaning at test (see Supplemental Material). Furthermore, the association between semantic memory and resampling may reflect consistent attention to the relationships between semantically relevant scene elements, paralleling theories of resampling in relational episodic memory (Wynn et al. 2019). In addition to the need to incorporate a role for semantic and image factors in theories of resampling, the present results point to the need to consider potential modulatory roles of these factors in the relationship between episodic memory and resampling. That is, the finding that more attention to meaning weakens the extent to which episodic memory strength predicts resampling, and vice versa, suggests that strong guidance by semantic memory may reduce the extent to which episodic memory is able to guide attention.

The apparent competition between semantic meaning and episodic memory observed in the present study also has potentially important implications for theories of attention. Specifically, many theories of attention have focused on competition between bottom-up or perceptual sources of guidance, such as image salience, and top-down cognitive factors in guiding eye movements (Van der Stigchel et al. 2009; Tatler et al. 2011). The present results, however, also point to the possibility that different top-down factors (i.e., semantic and episodic memory) may compete with each other to guide naturalistic viewing—a possibility that has been less well explored, particularly with respect to episodic memory. Harnessing meaning maps as a new method of indexing the distribution of semantic information, combined with the use of resampling as an index of attentional deployment—rather than solely as an index of episodic memory as has been done in prior work—provides a unique new window through which we can observe competition between semantic and episodic memory during naturalistic viewing for what is, to our knowledge, the first time. Using these measures, the current study suggests that fixations may be the result of a conflict between semantic and episodic memory, among other factors, to determine where attention is deployed. For example, when strong semantic guidance is present, episodic memory is less likely to "win" on any given fixation, and semantic memory would thus emerge as the stronger driver of resampling. However, further investigation of the potential interplay between episodic and semantic memory in driving attention is warranted before causal conclusions can be drawn.

In addition to implications for theories of both memory and attention, the present findings provide a potential new lens through which to view prior investigations of resampling behavior. The presently identified importance of both visual and semantic image content in resampling behavior indicates that prior findings of a relationship between episodic memory and resampling could have, in part, been driven by image content.

Specifically, in the present study, we incorporated a random effect of image in all analyses—including the null relation between recognition memory strength and attention to meaning—to ensure that we were not including potentially confounding image effects such as overall differences in meaning, salience, or memorability. The majority of prior studies, however, have not controlled for such image effects. This is a particularly important consideration because of findings that images that are more memorable also tend to elicit more similar viewing patterns between subjects (Mancas and Le Meur 2013). When image effects are not controlled for, this effect could emerge as an apparent within-subjects relation between resampling and episodic memory, when in fact it is a result of certain images leading to more stereotyped scan patterns—even between subjects—and better memory. Therefore, accounting for potential confounds of image properties may be particularly important for future investigations of resampling and memory.

Taken together, the present findings indicate that resampling behavior reflects cognitive sources of guidance besides episodic memory, and that these factors may influence the relationship between episodic memory and resampling. Future investigations aimed at uncovering other such factors that guide resampling, and how they might modulate its relationship with episodic memory, may prove fruitful. In particular, these results highlight the complex interplay of cognitive and visual factors that orchestrate how we guide our attention: it is rarely, if ever, just one factor at play. Our knowledge of the world, the task at hand, our memories, the current visual input, and likely myriad other influences are all resolved within a few hundred milliseconds to produce each movement of the eyes.

Materials and Methods

Participants

Forty-five undergraduates from the University of California, Davis completed the experiment for course credit. The sample size was selected to provide more than 98% power to detect the weakest effect of subjectively reported memory on eye movements obtained in a prior study (Ramey et al., 2019). The quality of each participant's eyetracking data was assessed by computing the mean percent signal across all trials to determine whether there was excessive track loss due to blinks or calibration loss. All participants had greater than the preselected criterion of 75% signal ($M = 94.7\%$), (Henderson and Hayes 2017), such that they lost less than 25% signal; all participants were thus retained for analysis.[4]

Stimuli

Stimuli were 200 photographs of real-world scenes. All scenes were presented in color at 1024 × 768 pixels subtending a visual angle of approximately 25° × 19° at presentation. Of these 200 scenes, 150 were presented at study and test, and 50 were presented only at test. Eighty out of the 200 scenes had been run through the meaning mapping procedure (from Henderson and Hayes 2017) and were used in analyses. Stimulus presentation was counterbalanced, such that each scene appeared in different conditions (i.e., in one of the two study tasks, or as a new lure during test; see procedure) for different participants, to mitigate stimulus effects.

Apparatus

Participants' eye movements were recorded using an SR Research EyeLink 1000+ tower mount eyetracker, sampling at 1000 Hz. A

---

[4]This data set is the same as that used in Experiment 2 of Ramey et al. (2020), which focused on a separate set of questions pertaining to episodic memory processes, but contained analyses of trial-level resampling and memory processes. Ramey et al. (2020) did not include any meaning or salience data, or fixation-level measures.
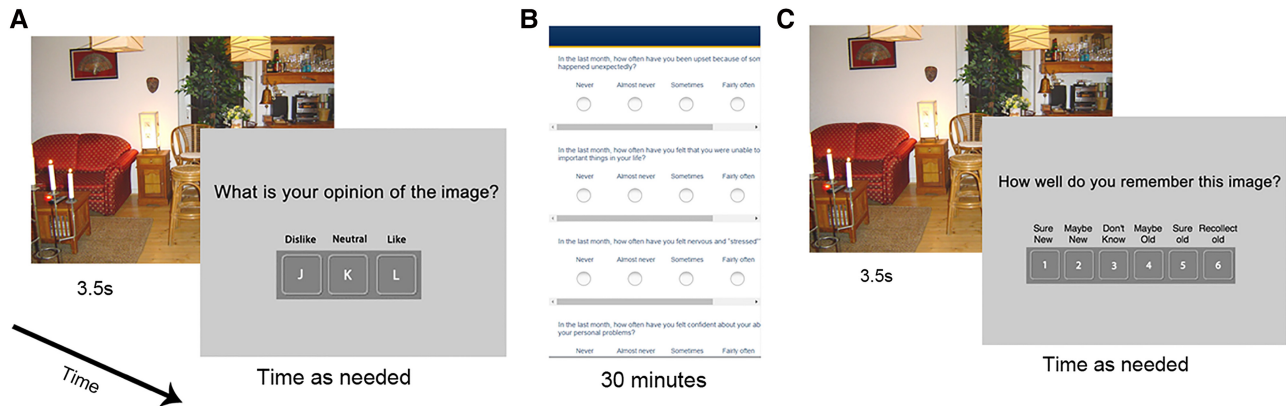
Figure 4. Illustration of the procedure. (*A*) Study phase. Half of the scenes were presented in an aesthetic judgment task (i.e., participants were instructed to judge the image aesthetically and rate it as "dislike," "neutral," or "like"), whereas the other half were presented in a memorization task (i.e., participants were instructed to memorize the image and rate it as "not memorable," "neutral," or "memorable"). (*B*) Delay between study and test, during which participants completed unrelated questionnaires. (*C*) Test phase in which participants rated their recognition confidence.

forehead and chin rest were used to reduce head movements, and eye movements were recorded from one eye though viewing was binocular. Stimuli were displayed on a monitor 85 cm from the eye-tracker, and the experiment was controlled with SR Research Experiment Builder software (SR Research 2010a).

Procedure

The experiment lasted 1.5 h and consisted of a study phase followed by a filled 30 min delay, as well as a subsequent test phase (see Fig. 4). Eye movements were recorded throughout the study and test phases. In both phases, each trial (i.e., each scene presentation) was preceded by a central fixation cross. Participants were given breaks every 50 trials and between phases, and the eyetracker was re-calibrated after each of these breaks.

Study phase

During the study phase, participants were presented with 150 unique scenes split into two task blocks: an aesthetic judgment task and a memorization task. These tasks were selected to ensure that any effects obtained were not a product of a given task, but rather generalized across tasks (as prior work has shown that eye movements vary systematically between tasks; Castelhano et al. 2009; Mills et al. 2011; Kardan et al. 2015). The order of the tasks was counterbalanced such that half of the participants completed the aesthetic judgment task first, whereas the other half completed the memorization task first. In each task, 75 scenes were presented for 3.5 sec each, allowing for an average of 12 fixations per trial. Each task was preceded by two practice trials to familiarize participants with the procedure.

In the aesthetic judgment task of the study phase, participants were asked to rate each scene based on how aesthetically pleasing they found it to be. Each trial consisted of the scene presentation, followed by a gray response screen con-
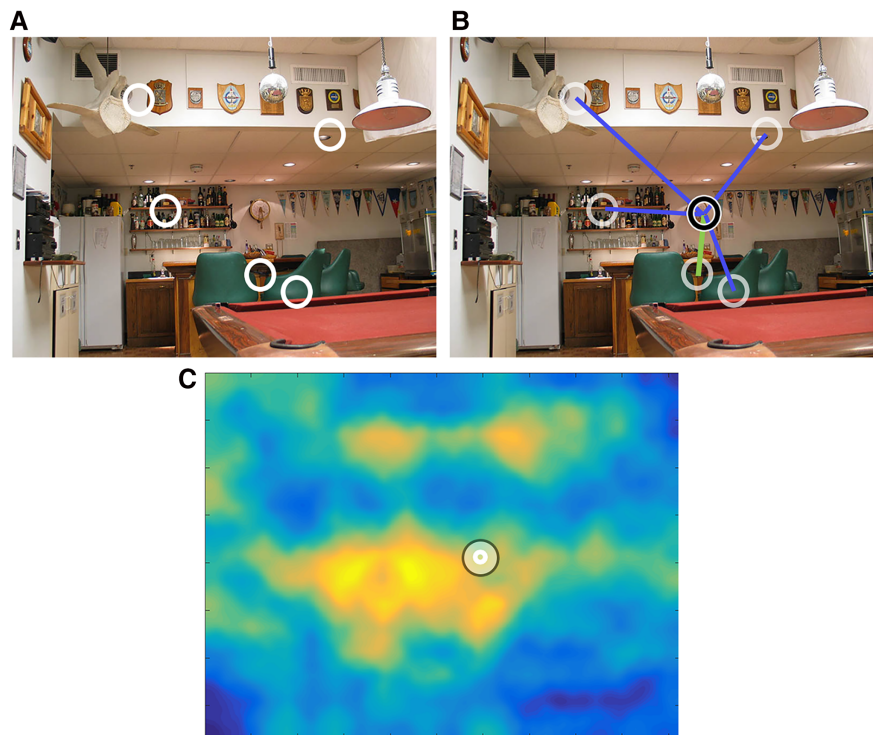
taining the prompt "What is your opinion of the photo?" as well as the key mappings for each response option. Responses were made on the keyboard, had no time limit, and consisted of "dislike," "neutral," and "like"; the response data were not used.

Figure 5. Resampling and meaning measures. (*A*) Fixations (white rings) made while studying the scene. (*B*) One of the fixations made while viewing the scene during test (black ring) along with the study fixations from a) (white rings). To calculate refixation distance for each test fixation, we computed the distance (lines) between the test fixation and every fixation made during study of that scene. The shortest resulting distance was assigned as the refixation distance score for that test fixation (green line). Thus, a lower refixation distance indicates that a test fixation was nearer to a region visited during study. (*C*) The meaning map of the scene. The brighter, yellow regions denote areas of high meaning, whereas the dark blue regions denote areas of low meaning. The meaning score of each test fixation was calculated by taking the average of the density of meaning within a 1° radius (black ring) around the fixation coordinates. The size of the 1° radius black ring is drawn to scale; the fixations, however, are recorded as a single coordinate and are not drawn to scale.

The memorization task of the study phase followed the same general procedure, but participants were instead asked to memorize the scenes. After each scene, they were asked to rate how memorable they found the scene to be. Participants were asked to give this response to ensure that the sequence of events in the memorization task was analogous to the aesthetic judgment task. Responses included "not memorable," "neutral," and "memorable"; again, the response data were not used.

Delay

Between the study and test phases, participants were moved to a computer in a different room to complete a 30 min distractor task that included questionnaires (e.g., personality scales) that were not related to the present study.

Test phase

In the test phase, participants were presented with a series of scenes and asked to rate their recognition memory for each scene. The test phase consisted of 200 trials: 150 old scenes, which had been presented in the study phase, and 50 randomly intermixed new scenes, which had not been presented previously. Each scene was presented for 3.5 sec, as in the study phase, and was subsequently replaced by a recognition judgment screen. Only the old scenes were used in primary analyses; the new scenes served as recognition lures.

For the recognition judgment, participants indicated whether or not they recognized the scene from the study phase. They were given as much time as they needed to select their response. Response options fell on a 1–5 and recollect scale made up of "sure new," "maybe new," "don't know," "maybe old," "sure old," and "recollect old" (Yonelinas 2002; Ramey et al. 2019). Participants were instructed and tested on how to use this scale prior to beginning the test phase.

Data reduction and analysis

Meaning maps

The meaning maps used were those created in Henderson and Hayes (2017), in which participants recruited via Amazon Mechanical Turk rated the meaningfulness of overlapping image patches of varying sizes. Specifically, they rated how informative or recognizable the visual information contained in each patch was. For each scene, the patch ratings were used to construct a map of the spatial distribution of meaning (Figs. 1C, 5C). For more details on how the meaning maps were generated, see Henderson and Hayes (2017). The resulting map for each scene was a 1024 × 768 matrix, with each cell corresponding to a pixel of the scene. The value in each cell represents the intensity of meaning at that point in the scene. The maps were Gaussian smoothed to account for the fall-off in visual acuity from the fovea.

Eye movements

Fixations and saccades were segmented with EyeLink's standard algorithm using velocity and acceleration thresholds (30°/sec and 9500°/sec²; SR Research 2010b). Eye movement data were imported offline into Matlab using the EDFConverter tool. The first fixation was excluded from all analyses because its location was determined by the experiment-defined central fixation point.

Resampling: We computed resampling on a fixation-by-fixation basis in order to determine the extent to which participants visited regions during test that were the same as (or near) those visited during study (Fig. 5A,B). To do this, we considered each fixation made during the test phase individually. For each test phase fixation on each scene by each subject, we computed the distance to each study phase fixation on that same scene by that same subject. The shortest resulting distance was reserved for analysis, and was termed *refixation distance*: the distance from a test fixation to the nearest region that had been viewed during study. Refixation distance thus measures the extent to which a test fixation was far from any previously visited region, such that a lower refixation distance reflects more resampling behavior.

Meaning scores: The extent to which participants attended to meaningful regions was calculated on a fixation-by-fixation basis, similar to refixation distance (Fig. 5C). The meaning map for a given scene was used to compute the average amount of meaning contained in a 1-deg radius around each test fixation on that scene. This yielded a meaning score for each fixation.

Statistical models

Statistical analyses were conducted using linear mixed effects models, which allowed us to harness trial-by-trial (i.e., within-subjects) data while controlling for individual differences and stimulus effects (Nuthmann and Einhäuser 2015). In addition to random intercepts of subject and image for all analyses, we nested fixations within trials for fixation-by-fixation analyses. The dependent variable in each model of resampling was refixation distance. The models were estimated using the lmerTest package in R (Kuznetsova et al. 2017), and were fit using maximum likelihood. The degrees of freedom and t values used were output by the linear mixed effects model for the variables of interest. The degrees of freedom were computed using the Satterthwaite approximation, and were rounded to the nearest integer in the manuscript. Effect sizes were calculated as a standardized regression coefficient (β).

Competing interest statement

Acknowledgments

References

Castelhano MS, Mack ML, Henderson JM. 2009. Viewing task influences eye movement control during active scene perception. *J. Vis* **9:** 6. doi:10.1167/9.3.6

Damiano C, Walther DB. 2015. Content, not context, facilitates memory for real-world scenes. *Vis cogn* **23:** 852–855. doi:10.1080/13506285.2015.1093241

Damiano C, Walther DB. 2019. Distinct roles of eye movements during memory encoding and retrieval. *Cognition* **184:** 119–129. doi:10.1016/j.cognition.2018.12.014

Foulsham T, Kingstone A. 2013. Fixation-dependent memory for natural scenes: an experimental test of scanpath theory. *J Exp Psychol Gen* **142:** 41–56. doi:10.1037/a0028227

Godden DR, Baddeley AD. 1975. Context-dependent memory in two natural environments: on land and underwater. *Br J Psychol* **66:** 325–331. doi:10.1111/j.2044-8295.1975.tb01468.x

Henderson JM. 2003. Human gaze control during real-world scene perception. *Trends Cogn Sci* **7:** 498–504. doi:10.1016/j.tics.2003.09.006

Henderson JM, Hayes TR. 2017. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat Hum Behav* **1:** 743–747. doi:10.1038/s41562-017-0208-0

Henderson JM, Hayes TR. 2018. Meaning guides attention in real-world scene images: evidence from eye movements and meaning maps. *J Vis* **18:** 10. doi:10.1167/18.6.10

Holm L, Mantyla T. 2007. Memory for scenes: refixations reflect retrieval. *Mem Cogn* **35:** 1664–1674. doi:10.3758/BF03193500

Humphrey K, Underwood G. 2010. The potency of people in pictures: evidence from sequences of eye fixations. *J Vis* **10:** 19. doi:10.1167/10.10.19

Kardan O, Berman MG, Yourganov G, Schmidt J, Henderson JM. 2015. Classifying mental states from eye movements during scene viewing. *J Exp Psychol Hum Percept Perform* **41:** 1502–1514. doi:10.1037/a0039673

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest Package: tests in linear mixed effects models. *J Stat Softw* **82:** 1–19. doi:10.18637/jss.v082.i13

Locher PJ, Nodine CF. 1974. The role of scanpaths in the recognition of random shapes. *Percept Psychophys* **15:** 308–314. doi:10.3758/BF03213949

Mancas M, Le Meur O. 2013. Memorability of natural scenes: The role of attention. In 2013 IEEE International Conference on Image Processing, pp. 196–200. IEEE http://ieeexplore.ieee.org/document/6738041/ (Accessed July 30, 2019).

Mantyla T, Holm L. 2006. Gaze control and recollective experience in face recognition. *Vis cogn* **14:** 365–386. doi:10.1080/13506280500347992

Mills M, Hollingworth A, Van der Stigchel S, Hoffman L, Dodd MD. 2011. Examining the influence of task set on eye movements and fixations. *J Vis* **11:** 17. doi:10.1167/11.8.17

Neider MB, Zelinsky GJ. 2006. Scene context guides eye movements during visual search. *Vis Res* **46:** 614–621. doi:10.1016/j.visres.2005.08.025

Noton D, Stark L. 1971. Scanpaths in eye movements during pattern perception. *Science* **171:** 308–311. doi:10.1126/science.171.3968.308

Nuthmann A, Einhäuser W. 2015. A new approach to modeling the influence of image features on fixation selection in scenes. *Ann N Y Acad Sci* **1339:** 82–96. doi:10.1111/nyas.12705

Ramey MM, Yonelinas AP, Henderson JM. 2019. Conscious and unconscious memory differentially impact attention: eye movements, visual search, and recognition processes. *Cognition* **185:** 71–82. doi:10.1016/j.cognition.2019.01.007

Ramey MM, Henderson JM, Yonelinas AP. 2020. The spatial distribution of attention predicts familiarity strength during encoding and retrieval. *J Exp Psychol Gen* doi:10.1037/xge0000758.

Ryan JD, Shen K. 2020. The eyes are a window into memory. *Curr Opin Behav Sci* **32:** 1–6. doi:10.1016/j.cobeha.2019.12.014

Saffran M, Schwartz F. 1994. Of cabbages and things: semantic memory from a neuropsychological perspective–A tutorial review. *Atten Perform XV Conscious nonconscious Inf Process* **25:** 507–536.

Smith SM. 1979. Remembering in and out of context. *J Exp Psychol Hum Learn Mem* **5:** 460–471. doi:10.1037/0278-7393.5.5.460

SR Research. 2010a. *Experiment builder user's manual*. SR Research Ltd., Mississauga, Canada.

SR Research. 2010b. *EyeLink 1000 user's manual (version 1.5.2)*. SR Research Ltd., Mississauga, Canada.

Tatler BW, Hayhoe MM, Land MF, Ballard DH. 2011. Eye guidance in natural vision: reinterpreting salience. *J Vis* **11:** 5–5. doi:10.1167/11.5.5

Tulving E. 1986. Episodic and semantic memory: where should we go from here? *Behav Brain Sci* **9:** 573–577. doi:10.1017/S0140525X00047257

Valuch C, Becker SI, Ansorge U. 2013. Priming of fixations during recognition of natural scenes. *J Vis* **13:** 1–22. doi:10.1167/13.3.3

Van der Stigchel S, Belopolsky A V, Peters JC, Wijnen JG, Meeter M, Theeuwes J. 2009. The limits of top-down control of visual attention. *Acta Psychol (Amst)* **132:** 201–212. doi:10.1016/j.actpsy.2009.07.001

Võ ML-H, Wolfe JM. 2013. The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition* **126:** 198–212. doi:10.1016/j.cognition.2012.09.017

Wynn JS, Bone MB, Dragan MC, Hoffman KL, Buchsbaum BR, Ryan JD. 2016. Selective scanpath repetition during memory-guided visual search. *Vis cogn* **24:** 15–37. doi:10.1080/13506285.2016.1175531

Wynn JS, Olsen RK, Binns MA, Buchsbaum BR, Ryan JD. 2018. Fixation reinstatement supports visuospatial memory in older adults. *J Exp Psychol Hum Percept Perform* **44:** 1119–1127. doi:10.1037/xhp0000522

Wynn JS, Shen K, Ryan JD. 2019. Eye movements actively reinstate spatiotemporal mnemonic content. *Vision* **3:** 21. doi:10.3390/vision3020021

Wynn JS, Ryan JD, Moscovitch M. 2020. Effects of prior knowledge on active vision and memory in younger and older adults. *J Exp Psychol Gen* **149:** 518–529. doi:10.1037/xge0000657

Yonelinas AP. 2002. The nature of recollection and familiarity: a review of 30 years of research. *J Mem Lang* **46:** 441–517. doi:10.1006/jmla.2002.2864

Why do we retrace our visual steps? Semantic and episodic memory in gaze reinstatement

Michelle M. Ramey, Andrew P. Yonelinas and John M. Henderson

Supplemental Material	http://learnmem.cshlp.org/content/suppl/2020/06/09/27.7.275.DC1
References	This article cites 33 articles, 6 of which can be accessed free at: http://learnmem.cshlp.org/content/27/7/275.full.html#ref-list-1
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first 12 months after the full-issue publication date (see http://learnmem.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**