Influences on memory for naturalistic visual episodes:
Sleep, familiarity, and traits differentially affect forms of recall

Marc N. Coutanche\textsuperscript{1,2,3}, Griffin E. Koch\textsuperscript{1,2}, & John P. Paulus\textsuperscript{1,2}

\textsuperscript{1} Department of Psychology, University of Pittsburgh, Pittsburgh PA
\textsuperscript{2} Learning Research and Development Center, University of Pittsburgh, Pittsburgh PA
\textsuperscript{3} Brain Institute, University of Pittsburgh, Pittsburgh PA

PLEASE CITE AS:

Running title: Influences on memory for naturalistic episodes

Keywords: episodic memory, events, recall, sleep, familiarity
Abstract

The memories we form are composed of information that we extract from multifaceted episodes. Static stimuli and paired associations have proven invaluable stimuli for understanding memory, but real-life events feature spatial and temporal dimensions that help form new retrieval paths. We ask how the ability to recall semantic, temporal and spatial aspects (the ‘what, when and where’) of naturalistic episodes is affected by three influences – prior familiarity, post-encoding sleep, and individual differences – by testing their influence on three forms of recall: cued recall, free recall, and the extent that recalled details are recombined to a novel prompt. Naturalistic videos of events with rare animals were presented to 115 participants, randomly assigned to receive a 12- or 24-hour delay with sleep and/or wakefulness. Participants’ immediate and delayed recall was tested and coded by its spatial, temporal and semantic content. We find that prior familiarity with items featured in events improved cued recall, but not free recall, particularly for temporal and spatial details. In contrast, post-encoding sleep, relative to wakefulness, improved free recall, but not cued recall, of all forms of content. Finally, individuals with higher trait scores in the Survey of Autobiographical Memory spontaneously incorporated more spatial details during free recall, and event details (at a trend level) in a novel recombination recall task. These findings show that prior familiarity, post-encoding sleep and memory traits can each enhance a different form of recall. More broadly, this work highlights that recall is heterogeneous in responses to different influences on memory.
1. Introduction

The episodes we experience are complex and multifaceted (Tulving, 2002), with a temporal (Howard & Kahana, 2002; Lehn et al., 2009) and spatial (Robin & Olsen, 2019) context for featured objects or agents. Within dynamic ‘events’, which we operationalize as continuous episodes with definable start- and end-points in a spatial location (Rubin & Umanath, 2015), episodic memories can be formed of item attributes (‘what’), their temporal context (‘when’), and their spatial context (‘where’). Here, we examine how influences on memory affect the recall of dynamic videos.

Why might naturalistic episodes differ from static images or words in how they respond to influences on memory? Information is more likely to be remembered when embedded in dynamic contexts rather than static images (Candan, Cutting, & DeLong, 2015; Matthews, Benjamin, & Osborne, 2007; Sonkusare, Breakspear, & Guo, 2019). Shifts in external input (i.e., spatial changes) and a temporal dimension have been proposed as the way that elements from our experience are successfully distinguished in memory (Ekstrom and Yonelinas, 2020). The presence of these aspects in naturalistic videos suggests that this type of stimulus might contain aspects that influence the distinction of items in ways that are not apparent from static stimuli. Further to this point, naturalistic videos incorporate what has been termed ‘high-dimensional’ facets of memories – namely items, context, and temporal information (Cowell, Barense, & Sadil, 2019), which are not as present in static images. This is particularly the case for the temporal dimension, which provides a unique code for retrieving memories (Ekstrom and Yonelinas, 2020) through temporal drift (Manning et al., 2011), event boundaries (Speer and Zacks, 2005), or prediction errors (Sinclair and Barense, 2019).

The above discussion suggests that naturalistic videos have important aspects not present in static images. On the other hand, if naturalistic videos do not differ from images in how they are encoded and retrieved, this would present a significant challenge to many of the theories discussed above, such as the key role proposed for high-dimensionality (Cowell et al., 2019). It is thus important to establish precisely how the recall of naturalistic videos is affected by influences on memory.

Here, we ask how different factors influence memory for the what, when and where aspects of events, as measured through different tests of recall. We investigate this question for three factors that are known to influence episodic memory for static images and words (e.g.,
Influences on memory

*Item familiarity:* Stronger prior familiarity with an item is associated with a greater likelihood of encoding episodic associations (Bransford & Johnson, 1972; Sala & Gobet, 2017). For instance, familiarity with domains such as chess (Chase & Simon, 1973), academic subjects (Brandt, Cooper, & Dewhurst, 2005), television shows (Long & Prat, 2002), sports (Bruett, Fang, Kamaraj, Haley, & Coutanche, 2018; Rawson & Van Overschelde, 2008) and people (Van Overschelde & Healy, 2001) modulates both retrieval success (Sala & Gobet, 2017) and the nature (Popov, Zhang, Koch, Calloway, & Coutanche, 2019) of episodic associations. Schema theories of memory (Gilboa & Marlatte, 2017; van Kesteren, Ruiter, Fernández, & Henson, 2012) and computational models such as the complementary learning systems model (McClelland, 2013) suggest that existing knowledge can enable the rapid encoding and consolidation of related new information. Prior familiarity might also act by reducing demands on retrieval, ultimately leading to superior memory performance (Trelle, Henson, & Simons, 2019).

*Sleep:* Reactivation of hippocampal representations during post-encoding slow-wave sleep (Born & Wilhelm, 2012) leads to strengthened representations in neocortex (Inostroza & Born, 2013), and the integration of new information with existing knowledge (Tamminen et al., 2010). Sleep has been found to positively affect encoding for new autobiographical memories (Aly & Moscovitch, 2010), contextual details (van der Helm, Gujar, Nishida, & Walker, 2011) and spatial relations (Coutanche, Gianessi, Chanales, Willison, & Thompson-Schill, 2013). It has been proposed that neuronal activity experienced during an episode is later replayed in sequence during sleep to allow its consolidation (Lewis & Durrant, 2011; O’Neill, Pleydell-Bouverie, Dupret, & Csicvari, 2010).

*Memory traits:* Much as individuals differ in cognitive abilities, people vary in how they encode and retrieve information (Kirchhoff, 2009; Palombo, Sheldon, & Levine, 2018). A recently identified individual trait relating to episodic memory is a person’s reported tendency to remember episodes in daily life (Palombo et al., 2018; Palombo, Williams, Abdi, & Levine,
This is reflected in responses to the Survey of Autobiographical Memory (SAM), which quantifies the extent that people report drawing on different forms of memory, through ratings on statements such as “When I remember events, I remember a lot of details” (Palombo et al., 2013). Scores on this scale are associated with differences in resting-state brain networks (Sheldon, Farb, Palombo, & Levine, 2016) and behavioral markers of successful learning (Coutanche & Koch, 2017). Particularly relevant for this investigation, the ‘episodic’ trait predicts individual differences in recollection (but not familiarity) of presented scenes (Palombo et al., 2013).

What, when, and where

It has been suggested that the central binding operation within the formation of episodic memories requires the presence of high dimensionality to be effective (Ekstrom and Yonelinas, 2020). In this context, ‘dimensionality’ describes the associations between complex elements in an episode (Cowell, Barense, & Sadil, 2019). Our examination of what, where and when fits well within this framework, with the presence of all three fitting criteria for high dimensional memories, or “conjunctions of complex elements; e.g., items, context, and temporal information” (p. 5, Cowell et al., 2019). This is not to say that the what, when and where of events will necessarily be affected by the three examined factors in the same way. For instance, computations that are central to memory formation, such as pattern separation, are likely to be more important for some aspects than others. For instance, due to shared spatial contexts across events, interference might be particularly high for the spatial aspects, leading to a greater involvement of pattern separation for this form of information, possibly leading to diverging effects in memory. We outline particular hypotheses below.

The existence of prior knowledge is known to enhance associative binding with an item’s context, sometimes at the cost of perceptual encoding (Gilboa and Marlatte, 2017). As a result, having prior knowledge (such as through a congruent schema) can enhance overall memory while reducing recollection of fine perceptual details, leading to false recognition (e.g., Spalding et al., 2015; though also see Fenn et al., 2009). Based on this distinction between contextual (in our study, spatial and temporal aspects) and perceptual details (in our study, more relevant to attributes), we hypothesized that prior familiarity would particularly benefit the recall of spatial and temporal aspects, while not impacting the recall of attributes to the same degree.
The influence of sleep on memory includes positive effects on spatial (Coutanche et al., 2013; Inostroza, Binder, & Born, 2013) and temporal (Inostroza et al., 2013) information. The effect on item details is, however, less clear. On the one hand, a study of episodic-like memory consolidation in rats revealed improved spatial and temporal, but not visual-object, information after sleep (Inostroza et al., 2013), with no changes observed in the ability to recognize studied items (Fenn et al., 2009). On the other hand, several studies in non-human animals have indicated that sleep deprivation reduces object recognition performance (Palchykova et al., 2006; Pinheiro-da-Silva et al., 2017), leaving open the possibility of item-level improvements. Based on this mixture of results, we predicted that we would observe sleep-induced improvements in memory for spatial and temporal aspects, but did not have a hypothesis for observing differences in attributes.

There is far less evidence available to hypothesize how individual differences in memory traits might differentially predict temporal, spatial or attributes, so we consider this particular examination to be more exploratory.

Memory outcomes

Memory recall can take a number of different forms (Aggleton & Brown, 2006), three of which are examined here. First, recall can be triggered by a broad prompt (‘free recall’) such as “Describe what you remember from the video”. Second, a specific cue can trigger the retrieval of targeted information (‘cued recall’) such as “What happened after the animal attacked the frog?” Finally, the constructive episodic simulation hypothesis (Schacter, & Addis, 2007) suggests a key use of memory is recombining recalled elements to form novel scenarios or constructs (Madore, Szpunar, Addis, & Schacter, 2016; Schacter & Addis, 2007). We test this ‘recombined recall’ by having participants recombine recalled information from events in a new way (e.g., “Describe a day in the life of this animal” after seeing distinct wildlife clips).

Of the three studied factors (familiarity, sleep, trait), we hypothesized that sleep would particularly affect free recall performance because the sequential neuronal replay that is observed during post-encoding sleep (Lewis & Durrant, 2011; O’Neill et al., 2010) is likely to benefit a behavioral measure aided by reactivation of a sequence, such as free recall. The cued recall of temporal aspects might also benefit from sleep for the same reason. However, whereas a cued recall benefit would require both the cue and response to have been reactivated, free recall is flexible enough for a person to specifically retrieve and report strong (and potentially
reactivated) memories, leading us to hypothesize that free recall would be particularly strengthened. There is less basis for predicting how sleep will impact recombined recall of naturalistic videos, though for static images, sleep can help link distinct items (Ellenbogen et al., 2007; Sweegers and Talamini, 2014). Yet, it is unclear whether this will generalize to videos, and to the form of recombination examined here.

Prior familiarity on the other hand is known to benefit associative memory in cases where interference must be overcome (Bein, Reggev, & Maril, 2019). In this study, cued recall is more likely to be susceptible to interference than the other forms of recall because giving a correct response requires selecting from among the many actions and events that occurred in the videos (whereas the other recall tasks are open-ended). Past work has also shown improved cued recall performance when items are known, compared to unknown (e.g., Zhang, Popov, Koch, Calloway, & Coutanche, 2018). This led us to hypothesize that prior familiarity with items in events (here, rare animals) would give superior cued recall performance. An alternative possibility is that novelty (i.e., lower familiarity) will increase salience during encoding, with a positive effect on memory (Bonasia et al., 2018; Tulving & Kroll, 1995; though see Poppenk, Köhler, & Moscovitch, 2010).

Finally, we hypothesized that memory trait scores would positively predict the amount of event information included within recall responses. Specifically, because trait scores probe the degree of spontaneously drawing on different forms of memory (Palombo et al., 2013), we predicted that trait scores (particularly episodic and spatial scales based on their importance to episodic memory) would relate to performance in the open-ended free and recombined recall tasks. This was particularly predicted for the recombined recall task based on recent evidence that an ‘episodic specificity induction’ (Schacter & Madore, 2016) can enhance the amount of episodic (but not object) details recalled by a person (Madore, Gaesser, & Schacter, 2014; Madore & Schacter, 2016). The induction is proposed to heighten a person’s focus on episodic details, which then enhances the construction of mental events (Schacter & Madore, 2016). As this very much parallels the motivation of the episodic memory trait score, we hypothesized that higher episodic SAM scores would be associated with more event aspects being recombined into a novel scenario/mental construction, in the same way that experimentally inducing greater use of episodic memory enhances the recombination of episode details (Madore et al., 2016). An additional reason for this hypothesis is that the creation of the trait measure is based on
autobiographical memory (Palombo et al., 2013), which is similar to our own cue to recombine events into a biographical schema (“describe a day in the life of this animal”).

This investigation

Here, we examine how prior familiarity, sleep, and memory traits affect the recall of temporal (when), spatial (where), and attribute (what) aspects of naturalistic events. Study participants viewed short video clips featuring six rare animals (two fish, two birds, two mammals). Participants’ abilities to recall events in the videos were tested immediately, and after wakefulness or sleep (12 or 24 hours after initial encoding), through different forms of recall.

2. Results

We investigated how the recall of different event aspects (when, where, what) are influenced by prior familiarity with featured items, sleep, and memory traits. The mean rating of prior familiarity for each featured animal (on a 1-7 scale) ranged from 2.58 to 4.75 (mean ($M$) = 3.31, standard deviation ($SD$) = 2.08) with significant differences between animals ($F_{5,660} = 17.51, p < 0.001$). In each of the models presented below (full details in Methods), we test how these three factors affect performance in cued, free, and recombined recall. Memory performance for the conditions is compiled in Tables 1 and 2 for reference.

Table 1. Mean accuracy (cued recall) and amount of information recalled (free and recombined recall) are listed for the immediate questions. Standard deviations are shown in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>12-hour AM</th>
<th>12-hour PM</th>
<th>24-hour AM</th>
<th>24-hour PM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Immediate Cued Recall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal (when)</td>
<td>.296 (.129)</td>
<td>.307 (.184)</td>
<td>.293 (.104)</td>
<td>.277 (.154)</td>
</tr>
<tr>
<td>Spatial (where)</td>
<td>.498 (.187)</td>
<td>.464 (.176)</td>
<td>.500 (.200)</td>
<td>.478 (.134)</td>
</tr>
<tr>
<td>Attributes (what)</td>
<td>.621 (.117)</td>
<td>.584 (.157)</td>
<td>.571 (.123)</td>
<td>.576 (.124)</td>
</tr>
<tr>
<td><strong>Immediate Free Recall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal (when)</td>
<td>5.333 (2.559)</td>
<td>3.793 (1.644)</td>
<td>4.444 (1.917)</td>
<td>3.957 (1.819)</td>
</tr>
<tr>
<td>Spatial (where)</td>
<td>4.257 (2.424)</td>
<td>2.655 (1.233)</td>
<td>2.519 (1.178)</td>
<td>2.570 (1.339)</td>
</tr>
<tr>
<td>Attributes (what)</td>
<td>5.111 (3.298)</td>
<td>4.333 (1.867)</td>
<td>3.556 (1.994)</td>
<td>3.935 (1.920)</td>
</tr>
<tr>
<td><strong>Immediate Recombined Recall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 8 -
<table>
<thead>
<tr>
<th></th>
<th>12-hour AM</th>
<th>12-hour PM</th>
<th>24-hour AM</th>
<th>24-hour PM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Delayed Cued Recall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal (when)</td>
<td>.275 (.109)</td>
<td>.247 (.158)</td>
<td>.222 (.098)</td>
<td>.260 (.127)</td>
</tr>
<tr>
<td>Spatial (where)</td>
<td>.486 (.180)</td>
<td>.391 (.136)</td>
<td>.389 (.168)</td>
<td>.465 (.211)</td>
</tr>
<tr>
<td>Attributes (what)</td>
<td>.565 (.123)</td>
<td>.545 (.135)</td>
<td>.574 (.126)</td>
<td>.545 (.126)</td>
</tr>
<tr>
<td><strong>Delayed Free Recall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal (when)</td>
<td>4.042 (1.723)</td>
<td>3.977 (2.000)</td>
<td>3.938 (2.344)</td>
<td>3.903 (2.028)</td>
</tr>
<tr>
<td>Spatial (where)</td>
<td>3.111 (1.410)</td>
<td>2.966 (1.569)</td>
<td>2.469 (1.665)</td>
<td>2.892 (1.326)</td>
</tr>
<tr>
<td>Attributes (what)</td>
<td>4.542 (2.282)</td>
<td>4.598 (2.040)</td>
<td>3.691 (1.804)</td>
<td>4.220 (2.055)</td>
</tr>
<tr>
<td><strong>Delayed Recombined Recall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal (when)</td>
<td>3.739 (1.197)</td>
<td>2.897 (1.225)</td>
<td>2.691 (1.525)</td>
<td>2.952 (1.716)</td>
</tr>
<tr>
<td>Spatial (where)</td>
<td>2.232 (.901)</td>
<td>2.069 (.888)</td>
<td>2.038 (.568)</td>
<td>2.301 (1.206)</td>
</tr>
<tr>
<td>Attributes (what)</td>
<td>1.942 (.814)</td>
<td>1.954 (1.030)</td>
<td>2.037 (1.083)</td>
<td>2.167 (1.118)</td>
</tr>
</tbody>
</table>

2.1 Cued recall

Cued recall of temporal aspects ($\beta = 0.14, p < .001$), and spatial aspects ($\beta = 0.06, p = .077$) at a trending level, but not attributes ($\beta = 0.05, p = .190$), was significantly greater with increasing prior familiarity with the featured animals (Figure 1).
Figure 1. Cued recall accuracy for different event aspects based on participants’ prior familiarity with the featured animals. Statistically significant positive regression lines are indicated by an asterisk. Trending regression lines are indicated by a cross. Standard error bounds are represented in shaded areas.

An effect of sleep is tested through an interaction between sleep condition (sleep versus wakefulness) and retrieval session (pre or post delay). Sleep did not influence the cued recall of temporal ($\beta = -0.02, p = .662$), spatial ($\beta = -0.02, p = .500$), or attribute ($\beta = 0.06, p = .170$) aspects. Similarly, individual differences in the trait scores did not predict their associated cued recall performance (temporal aspects: $\beta = -0.04, p = .383$; spatial aspects: $\beta = 0.04, p = .275$; attributes: $\beta = -0.10, p = .061$).

Cued recall of temporal aspects was not significantly predicted by any other predictors. Spatial performance was predicted by group ($\beta = -0.12, p = .048$), with greater recall by participants randomly allocated to a sleep group (though because this includes both immediate and delayed recall, this does not reflect an effect of sleep, which requires an interaction between sleep and testing delay). Attribute performance was predicted by session ($\beta = -0.09, p = .026$), with better recall during session 1 ($M = 0.59, SD = 0.13$) than session 2 ($M = 0.56, SD = 0.13$).

To test for possible overlap (redundancy) in cued and free recall responses, we measured whether participants’ cued recall responses had already been included in their prior free recall. The mean proportion of cued recall responses also included in a participant’s prior free recall...
was low at 11.03%. Importantly, the four groups did not differ in the number of cued recall responses having overlap for temporal ($\Gamma^2 = 4.02, p = .259$), spatial ($\Gamma^2 = 4.03, p = .258$) or attribute ($\Gamma^2 = 3.83, p = .281$) questions.

After removing events that overlapped with those mentioned during free recall, prior familiarity with the featured animals continued to predict greater cued recall of temporal aspects ($\beta = 0.14, p < .001$). Spatial aspects ($\beta = 0.06, p = .091$) and attributes ($\beta = 0.07, p = .122$) continued to remain non-significant. The effect of sleep was also not affected by removing overlapping events, as it continued not to influence the cued recall of temporal ($\beta = 0.03, p = .487$), spatial ($\beta = 0.04, p = .245$) or attributes ($\beta = 0.05, p = .191$). Similarly, individual differences in memory trait scores continued to not predict cued recall performance (temporal aspects: $\beta = -0.03, p = .474$; spatial aspects: $\beta = 0.04, p = .358$; attributes: $\beta = -0.07, p = .136$).

Cued recall of temporal and spatial aspects was not significantly predicted by other effects. Attribute performance was again predicted by session ($\beta = -0.10, p = .018$), with greater recall during session 1 ($M = 0.54, SD = 0.14$) than session 2 ($M = 0.50, SD = 0.14$). Overall, removing overlapping events had limited impact on the key results.

2.2. Free recall

The amount of information that was free recalled from events was significantly predicted by the presence of sleep between encoding and retrieval for temporal (interaction between sleep and session: $\beta = 0.10, p = .003$), spatial ($\beta = 0.14, p < .001$) and attribute ($\beta = 0.07, p = .047$) aspects of events (Figure 2). The directionality indicated that sleep reduced the session 1-to-2 decline for temporal aspects (wake session 1: $M = 5.33, SD = 2.56$, session 2: $M = 4.04, SD = 1.72$; sleep session 1: $M = 4.05, SD = 1.79$, session 2: $M = 3.94, SD = 2.10$), spatial aspects (wake session 1: $M = 4.26, SD = 2.42$, session 2: $M = 3.11, SD = 1.41$; sleep session 1: $M = 2.58, SD = 1.24$, session 2: $M = 2.79, SD = 1.52$), and attributes (wake session 1: $M = 5.11, SD = 3.30$, session 2: $M = 4.54, SD = 2.28$; sleep session 1: $M = 3.95, SD = 1.93$, session 2: $M = 4.18, SD = 2.00$). To verify that these effects were not dependent on other group differences, we first verified there were no time-of-day effects (in addition to our inclusion of time-of-day in the model above). The two 24-hour groups—which both received sleep but differed in their encoding/retrieval time-of-day—did not differ in the degree of inter-session change for any of the event aspects (temporal: $\beta = 0.06, p = .107$; spatial: $\beta = 0.06, p = .168$; attributes: $\beta = 0.00, p =$
suggesting that no differences were observed when holding sleep constant. Second, we examined if the wake group's significant decline in recall across sessions depended on its higher starting-point (baseline). To examine this, we progressively removed wake participants in decreasing order of their immediate mean free recall scores for temporal aspects (which had the largest across-session decline in the wake group) until the 12-hour wake and sleep groups did not differ significantly in their immediate recall scores. This new 12-hour wake group ($N = 21$) still showed a significant decrease in free recall performance between sessions (e.g., temporal aspect: $t_{20} = 2.13, p = .046$), supporting the idea that the sleep effects did not depend on any baseline difference.

![Figure 2. Change in mean number of free recalled details between sessions. Error bars reflect the standard error of the mean change. Asterisks reflect a significant difference in the relevant models (sleep x day interaction).](image)

Of the other predictors, session (first > second) significantly predicted temporal ($\beta = -0.14, p < .001$) and spatial ($\beta = -0.10, p = .003$) aspects, as did group when collapsed across sessions (with more spatial aspects in the wake group: $\beta = -0.21, p = .013$). The number of recalled spatial aspects was also positively associated with spatial SAM scores ($\beta = 0.15, p = .008$). Examining the specificity of this relationship showed that spatial aspects were also predicted by episodic SAM scores ($\beta = 0.13, p = .015$) but not semantic SAM scores ($\beta = 0.01, p = .908$).
2.3 Recombined recall

In a model predicting the amount of information that participants extracted from events and recombined into an episodic simulation (“describe a day in the life of the animal”), none of the included predictors reached significance. However, because episodic simulation studies have only tested encoded material within the same session (e.g., Madore et al., 2014; Madore & Schacter, 2016; Madore et al., 2016), we also examined effects only for participants experiencing the shorter 12-hour delay, as well as for immediate retrieval performance. The 12-hour participants showed a trending positive relationship with episodic SAM scores ($\beta = 0.18, p = .062$), and an effect of session ($\beta = -0.10, p = .021; \text{first} > \text{second}$), with no other significant predictors ($ps > .131$). Examining the specificity of the episodic SAM scale showed that semantic ($\beta = 0.22, p = .029$) but not spatial ($\beta = -0.04, p = .690$) scores also positively predicted the amount of information recombined. This was specific to the 12-hour delay, as individuals with the 24-hour delay did not show significant effects for episodic SAM score ($\beta = -0.14, p = .196$), session ($\beta = -0.03, p = .431$), or other variables ($ps > .675$). Focusing on immediate retrieval (across all groups), no predictors reached statistical significance ($ps > .173$).

3. Discussion

This investigation examined how memory recall of naturalistic visual episodes is influenced by prior familiarity with a featured item, post-encoding sleep, and trait differences in episodic memory use. We tested how each of these factors influenced different forms of recall. Familiarity with the animals featured in the videos was positively associated with cued (but not free or recombined recall) of temporal aspects and with spatial aspects (at a trend). Post-encoding sleep protected against a wake-associated decline in free (but not cued or recombined) recall of spatial, temporal and attribute event aspects. Memory trait scores were positively associated with the amount of spatial aspects free recalled, and with the amount of event information recombined into a new scenario (but only within 12 hours of encoding).

Why did prior familiarity improve the cued recall of spatial and temporal information? One possibility is that prior familiarity encourages the integration of new information into cortical networks by reducing interference from other experienced events (Bein et al., 2019). Cued recall is more likely to benefit than is free recall because whereas presented cues can
trigger the retrieval of similar events (leading to interference), the open-ended nature of free recall allows participants to select strong memories for their responses. When the cue is provided by the experiment (cued recall, by definition), a participant is exposed to cues for events that are strongly integrated into memory, and to cues for events that are vulnerable to interference.

The positive relationship between familiarity and the formation of new episodic memories is also consistent with findings that processing familiar items during encoding can enhance new learning (Coutanche & Thompson-Schill, 2015; Lewis & Durrant, 2011; Tse et al., 2007; Zhang et al., 2018). A likely anatomical basis for these effects is the medial prefrontal cortex (mPFC), which is more active when prior knowledge is semantically consistent with observed episodes (Bonasia et al., 2018), possibly through interactions with the medial temporal lobe (Gilboa & Marlatte, 2017; van Kesteren, Ruiter, Fernández, & Henson, 2012). Following their experience with this set of animals, our participants will have new knowledge about these animals incorporated into their neocortex (Coutanche & Koch, 2018; Coutanche & Thompson-Schill, 2019). In addition to effects on encoding and consolidation, prior familiarity might also aid the recall process. For instance, retrieving information that is related to existing knowledge reduces retrieval demands, leading to superior performance (Trelle et al., 2019). Future studies might test the contributions of familiarity for stages of encoding, consolidation, and recall of dynamic episodes (e.g., Anderson, Pichert, & Shirey, 1983). Although it might at first seem surprising that the animals’ attributes were not better remembered for familiar items, we selected the tested attributes to not be typically known, such as an individual animal’s unique patterning or eye color (when visible during the clip).

Post-encoding sleep was found to improve participants’ free recall of spatial, temporal, and attribute aspects. Sleep-induced improvements in episodic memory have previously been observed in tests of free recall (Aly & Moscovitch, 2010; Benson & Feinberg, 1975; Ficca, Lombardo, Rossi, & Salzarulo, 2000; Lahl, Wispel, Willigens, & Pietrowsky, 2008) and cued recall (Grosvenor & Lack, 1984), though (to our knowledge) this is the first time that cued and free recall have been directly compared after naturalistic events. The beneficial effect of sleep on free recall is consistent with free recall reflecting relational memories (Hunt & Einstein, 1981; Long & Kahana, 2015), which are consolidated during sleep (Coutanche et al., 2013; Ellenbogen, Hu, Payne, Titone, & Walker, 2007). In this study, these relational associations can connect spatiotemporal elements of an event. Free recall is likely to benefit most from
consolidated relationships between elements because any resulting new retrieval paths can support open-ended free recall. Neuronal replay observed in the hippocampus during sleep mimics the spatial and temporal order of neuronal activation from an original event (Euston, Tatsuno, & McNaughton, 2007; O’Neill et al., 2010), which bind spatially and temporally adjacent elements in memory (Long & Kahana, 2015), allowing for superior free recall.

Participants with higher spatial memory trait scores free recalled more spatial information, while participants with higher episodic memory trait scores recombined more information from the observed events into a novel scenario. These findings warrant caution because the scales were not specific to the type of information that was recalled (e.g., spatial free recall was associated with spatial and episodic SAM scores). That said, a positive relationship is consistent with the episodic simulation hypothesis (Schacter & Addis, 2007), and findings that inducing episodic encoding can increase the number of episodic details incorporated into constructed mental events (Madore et al., 2014). Further work will be needed to determine the contexts and time delays in which memory trait differences are observed (Palombo et al., 2018).

The consistent modulation of the temporal aspects across the three memory influences supports the idea that naturalistic dynamic events have features (not apparent in static images or words) that are important for memory. This is consistent with the previously discussed proposal that high-dimensional facets of events are key components of memories (Cowell et al., 2019), and that temporal markers might help distinguish elements of our experience (Ekstrom and Yonelinas, 2020). Recently, results from an innovative neurofeedback fMRI paradigm provided further evidence for this. deBettencourt and colleagues demonstrated that neurally reinstating contextual information predicts improvement in subsequent free recall, unlike neurally reinstating item information (deBettencourt, Turk-Browne, & Norman, 2019). Static images often include only item information (or at least have very weak contextual information), suggesting that contextual (temporal and spatial) aspects might provide a route to consolidation that is less available to static stimuli. Reinforcing the significant impact of having a temporal component, this aspect is particularly meaningful for which neural systems become engaged during recall. The brain’s systems are organized in a manner that can be related to the time-scale of their respective representations (Hasson, Chen, & Honey, 2015), with ‘long time-scale’ regions supporting cross-model representations (Baldassano et al., 2017) and shorter time-scale regions supporting sensory representations.
One limitation of this study was that the pre-delay (baseline) performance in the wakefulness group was higher than in the sleep participants. We accounted for this in three ways: 1) Testing the significance of a session/sleep interaction (to account for differences); 2) Including time-of-day in regression models; 3) Repeating the key analysis in a subset of participants with more similar baselines. Although these checks suggested that baseline differences are not required for the sleep effects we observed, we nonetheless suggest caution in interpreting the magnitude of the effect. Interestingly, baseline differences across groups (which might have a circadian source) have been observed across explicit and implicit memory encoding and retrieval (Delpouve et al., 2014; Himmer, Müller, Gais, & Schönauer, 2017; May et al., 2005). The source of such differences is as yet unknown, but potential influences include fluctuations in attention and executive functioning (Schmidt et al., 2007). We support recent calls for the triggers and circumstances of this phenomenon to be examined further (Himmer et al., 2017).

To summarize the results of this study, we report evidence that familiarity, sleep, and individual traits, support memory for the spatial, temporal and attribute aspects of dynamic visual events. These influences are observed in different forms of recall, which highlights the importance of systematically examining different recall tests to fully understand predictors of episodic memory.

4. Materials and Methods

4.1. Participants

115 participants (51 males; $M_{age} = 20.8, SD = 2.9$) were recruited from the Pittsburgh community. Participants (all English speakers without a learning or attentional disorder) were randomly assigned to one of four conditions, and data collection continued until all groups contained at least 25 participants. The groups differed in the delay between the first and final session (12 or 24 hours) and time-of-day of training/testing (AM or PM), creating four groups: 12-hour morning/night (“12 hour AM”; $n = 25$), 12-hour night/morning (“12 hour PM”; $n = 29$), 24-hour morning/night (“24 hour AM”; $n = 29$), 24-hour night/night (“24 hour PM”; $n = 32$). These four conditions allowed us to compare performance for participants who had slept and not slept between sessions, while controlling for time-of-day. Four participants failed to complete the experiment’s surveys so were removed before analysis. Informed consent was obtained from
participants prior to the study. After the study, participants received compensation through course credit or payment for their time. The University of Pittsburgh Institutional Review Board approved all procedures.

4.2. Design

4.2.1 Encoding

During the initial encoding session, participants viewed 24 silent nature video clips that were each 45 seconds in duration. Participants were instructed to watch the videos and were asked to track their current interest with a movable slider to ensure attention and to act as a cover task. Practice trials were administered to ensure participants became comfortable with indicating their interest using the slider. At no time were participants told that they would be asked to retrieve information about the videos.

Six rare animals were featured in the videos (two fish, two birds, two mammals), selected based on (un)familiarity evaluations by an independent norming group that rated familiarity through Amazon Mechanical Turk (residing in the United States) on a scale of 1 ("not familiar") to 5 ("very familiar"). These animals received a mean familiarity of 1.9 (half the familiarity ($M = 4.1$) of typically familiar species such as ‘lion’). Each animal was featured in four videos, which were shown in a pseudo-randomized order so that a video of every animal was presented before the second video of each animal, and so on.

4.2.2 Immediate retrieval

After participants viewed the videos, the immediate retrieval session asked questions about three of the six animals (one fish, one bird, one mammal). The other three animals were held-out to be retrieved during the second session (described below). The selection of the three animals for the first (or second) session was counterbalanced across participants. For each animal, three recall probes were presented in order of increasing specificity - to ensure the cued recall questions and answers (given last) could not influence free recall responses. The experimenter started by asking the participants to give a short description of one of the six animals in the videos until the experimenter recognized the animal being referenced. If a second-session animal was referenced by the participant in the first session, the experimenter asked the participant to describe another animal. This approach ensured that the experimenter did not provide information other than the participant's own words to identify the animal.
Free recall: When the participant referenced one of the three animals that was assigned to immediate retrieval, they were first asked to verbally free recall as much information and detail as they could with the probe: "Can you tell me what you saw in the video?" No time limits were imposed on responses.

Recombined recall: Next, participants’ ability to recombine aspects from events into new scenarios was tested. This question required participants to recall from the observed events, and recombine this information in a new way, to answer: “If you had to describe a day in the life of this animal, what would you say?” Additional questions were included that did not have a temporal focus or require recombination: with a spatial focus (“Can you tell me about the places where this animal was?”) and attribute focus (“What do you now know about this animal after watching this video?”). Including these questions also removed any undue influence of the temporal focus of the recombination question on subsequent responses. As these two questions did not test recombined recall, we focus on the "day in the life" question here. No time limits were imposed on responses.

Cued recall: Finally, participants’ ability to recall specific information was tested through cued recall questions. These questions were temporally specific (e.g., “What did the animal do while it yawned?”; 4 questions), spatially specific (e.g., “In which direction did the animal climb in the tree?”; 4 questions) and attributes (e.g., “What pattern is on the body of this animal?”; 8 questions). Importantly, the attributes were selected to require episodic memory of the videos rather than prior knowledge, such as the pattern on an individual animal’s coat or eye color. These attributes can vary across individual animals, drawing on episodic memory of the events. Due to an experimental error, questions were not asked in 0.48% of cases. These were not used to calculate accuracies. No time limits were imposed on responses.

4.2.3 Delayed retrieval

After immediate retrieval, participants left the laboratory and returned either 12 or 24 hours later (depending on their condition assignment). Upon their return, participants answered (free, recombined, cued) recall questions for the remaining three animals that were not tested in immediate retrieval. Participants then completed the SAM (Palombo et al., 2013), a self-report measure of the extent that a participant employs episodic, semantic, and spatial memory, as well as future prospection in their daily life, followed by the Pittsburgh Sleep Quality Index (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989) and Verbalizer-Visualizer Questionnaire.
(Richardson, 1977). SAM responses were coded based on a key provided by the survey’s authors, producing trait scores for episodic, semantic, and spatial memory for each participant. SAM scores were mean-centered prior to inclusion in the analyses. Finally, participants reported their prior familiarity (i.e., before the study) with the six featured animals on a 7-point scale (1 = not at all familiar; 4 = somewhat familiar; 7 = very familiar).

4.3. Coding responses

All recall responses were coded by researchers who were blind to condition. Participants’ free recall responses were partitioned into the smallest possible units of information about the videos. A trained researcher (parser) separated participants’ typed responses into fragments according to a strict set of rules. First, sentences were segmented into fragments at each verb, prepositional phrase, adjective + noun phrase, and noun + verb (+ object) phrase (Figure 3). After parsing, two additional researchers (coders) independently coded each fragment according to the presence of aspects of events that are temporal, spatial, and attributes. Fragments were classified as temporal when they referred to a temporally-specific action or event that occurred in the video (such as “the animal then ate a bug”). Fragments were classified as spatial when they contained information about location or direction (e.g., “under the log”), and as attributes when they referred to features of the animals (e.g., “the animal has yellow eyes”). Other fragments were labeled ‘N/A’ (Figure 3). After independently coding fragments, the coders discussed and reached agreement on any inter-coder discrepancies. The parser and coders were all blind to participants’ conditions. Responses to the recombined recall question were coded in the same manner to measure the number of fragments referring to an action or event (e.g., “spends the day searching for food”). The numbers of free recalled and recombined recalled fragments of each type were calculated and used as the dependent variable in their respective models.

The video began/ with black birds/ circling/ in the sky.

<table>
<thead>
<tr>
<th>N/A</th>
<th>Attribute</th>
<th>Temporal</th>
<th>Spatial</th>
</tr>
</thead>
</table>

Figure 3. Example of coded fragments from free recall. Fragments defined by the parser are bounded by “/”. Labels (underneath fragments) denote how the information is classified: attribute (color), temporal (an action occurring at a particular time), and spatial (location).

For cued recall, answer keys were first constructed by three independent coders. These individual answer keys were combined to create a final key that was used to categorize responses.
as correct or incorrect (while blind to participant condition). We also measured the amount of overlap between the free and cued recall responses. For every cued recall question, two independent coders rated whether each participant’s response had been included in their preceding free recall response. When the two raters made conflicting judgments, a third rater broke the tie.

4.4. Statistical Analysis

For each form of recall probe (free, recombined, cued), we employed a separate linear mixed effects regression model for each of the three aspects (temporal, spatial, attribute). Within each of these models, we included fixed effects terms for prior familiarity ratings given to the featured animals, the individual’s corresponding memory trait SAM score (episodic score for temporal, spatial score for spatial, semantic score for attributes), participant gender, delay between sessions (12 or 24 hours), time-of-day (morning or night), and an interaction term that tested an effect of sleep (wake or sleep x immediate or delayed; as the sleep variable alone includes performance from both before and after the delay). Random effects were included for subject and featured animal. The free recall and recombined recall models were used to predict the total number of recalled aspects, and the cued recall model predicted accuracy (%) of responses, resulting in the subjects contributing six data points to each model.

Acknowledgements

We thank Heather Bruett for providing comments on an earlier version of the manuscript. We also thank Uma Balaji, Angeleah Carreras, Sharran Chakravorty, Jasmine Issa, Ankitha Iyer, Claire Kolhoff, Mariah Mascara, Alex Pang, Leslie Rajendran, Amber Shergill, Abigail Swire Carlo Vignali, and Aarya Wadke for assistance with data collection and/or coding.

References


