Predictions transform memories: How expected versus unexpected events are integrated or separated in memory

Oded Bein [a,*,1], Camille Gasser [b,*,1], Tarek Amer [c], Anat Maril [d,e,1], Lila Davachi [f,1]

[a] *Princeton Neuroscience Institute, Princeton University, Princeton, NJ, United States*
[b] *Department of Psychology, Columbia University, New York, NY, United States*
[c] *Department of Psychology, University of Victoria, Victoria, Canada*
[d] *Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel*
[e] *Department of Cognitive Science, The Hebrew University of Jerusalem, Jerusalem, Israel*
[f] *Center for Clinical Research, The Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, United States*

ARTICLE INFO

ABSTRACT

Our brains constantly generate predictions about the environment based on prior knowledge. Many of the events we experience are consistent with these predictions, while others might be inconsistent with prior knowledge and thus violate our predictions. To guide future behavior, the memory system must be able to strengthen, transform, or add to existing knowledge based on the accuracy of our predictions. We synthesize recent evidence suggesting that when an event is consistent with our predictions, it leads to neural integration between related memories, which is associated with enhanced associative memory, as well as memory biases. Prediction errors, in turn, can promote both neural integration and separation, and lead to multiple mnemonic outcomes. We review these findings and how they interact with factors such as memory reactivation, prediction error strength, and task goals, to offer insight into what determines memory for events that violate our predictions. In doing so, this review brings together recent neural and behavioral research to advance our understanding of how predictions shape memory, and why.

1. Introduction

Much of what we experience is highly predicted. Our knowledge about the world allows us to predict what will happen when we enter a restaurant, visit a friend, or listen to a familiar song. Occasionally, however, a surprising event occurs that violates our expectations. While it is useful to remember events that are consistent with our predictions, as well as those that violate them, these different types of events pose different demands on our memory system. Consistency with prior knowledge might reinforce our predictions, increasing the strength of predicted information in memory and linking it to existing knowledge structures through *memory integration*. Inconsistency, in turn, might require that we either update our existing knowledge to allow for the integration of novel information, or that we store conflicting information as a separate memory (*memory separation*), in case our existing knowledge becomes relevant again in the future (Bein et al., 2020; Love et al., 2004; McClelland et al., 1995, 2020; Piaget, 1952; van Kesteren

et al., 2012). The question of knowledge updating is even more pressing in the current time of social media and alternative facts. We are constantly exposed to pieces of information that vary both in how truthful and how surprising they are. As such, our memory system must be able to flexibly update our existing knowledge (if new information is believed to be true), dismiss the information altogether (if it is believed to be false), or hold on to the information as a separate memory (if its accuracy is unknown).

We build on prior proposals within the predictive coding framework (Friston, 2005, 2018; Lisman and Redish, 2009), which argue that most of our experiences involve predictions that are based on pre-existing memories. Stimuli in the environment serve as memory cues, leading to the reactivation of related memories (Anderson, 1974; Anderson and Milson, 1989; Collins and Loftus, 1975). This reactivation of related memories encompasses reinstatement of neural activity patterns that represent our learned expectations of what is likely to happen in a particular environment (Ahissar and Hochstein, 2004; Bar, 2009;

Buckner, 2010; Danker and Anderson, 2010). For our purposes, a prediction can be thought of as the reactivation of learned representations that are relevant to what is about to occur. For example, upon entering a kitchen, related memories of objects typically found in a kitchen are reactivated, forming a prediction that we are likely to encounter these objects. Further, we use eye movements to scan our environment in a sequential manner, focusing first on some aspects of the world and then on others (Ahissar and Hochstein, 2004; Bar, 2009; Gronau, 2021; Wynn et al., 2020). Thus, memory-based predictions typically unfold across time, even when we are exposed to both the predictive cue and the predicted (or unexpected) information at the same time. To continue with our kitchen example, one might first focus on one specific object (e. g., a fridge), which in turn leads to the reactivation of related memories that reflect predictions about what additional objects should be found nearby (e.g., a kitchen sink). In this sense, in a familiar context, our brain can generate predictions about what stimuli or events are likely to occur. Events that are consistent with our prior knowledge are consistent with the predictions generated based on this knowledge. Events that are incongruent with our prior knowledge, conversely, violate our predictions (Fig. 1).

Previous research has established that consistency and inconsistency with prior knowledge and predictions impact novel learning, as well as the cortical and hippocampal networks that support learning and memory (reviewed in Alonso et al., 2020; Fernández and Morris, 2018; Gilboa and Marlatte, 2017; Preston and Eichenbaum, 2013; van Kesteren et al., 2012). In particular, recent neuroimaging work in humans has capitalized on multivariate analysis methods, such as representational similarity analysis (RSA; Kriegeskorte et al., 2008), to measure the similarity between activity patterns elicited by different memories. By probing the representations of interrelated memories, research using these approaches has begun to uncover how new experiences and existing knowledge become integrated or separated in the brain (e.g., Audrain and McAndrews, 2022; Bein et al., 2020; Molitor et al., 2021; Schlichting et al., 2015; Sommer et al., 2022; Wu et al., 2023). In this review, we summarize this emerging evidence together with recent behavioral studies (E.g., Antony et al., 2022; Bein et al., 2021; Greve et al., 2017; Popov et al., 2019; Quent et al., 2022; Tompary et al., 2020; Tompary and Thompson-Schill, 2021; Wing et al., 2022) to discuss how expected versus unexpected information becomes represented in the brain and shapes memory formation (Fig. 2).

2. Consistency with prior knowledge promotes neural integration that benefits, but also biases, memory

Intuitively, when new information is consistent with our expectations, that experience should be easily integrated with our existing knowledge (unless the experience is perfectly and fully predicted, which might happen only in theory, at which point no further learning might take place). Indeed, this idea is reflected in prior research showing that consistency with prior knowledge is associated with altered hippocampal and cortical dynamics during learning (Amer et al., 2019; Bein et al., 2014; Brod et al., 2015, 2016; Brod and Shing, 2018; Maril et al., 2011; Reggev et al., 2016; Staresina et al., 2009; Tse et al., 2007, 2011; van Buuren et al., 2014; van der Linden et al., 2017; van Kesteren, Fernández et al., 2010; van Kesteren, Rijpkema et al., 2010; van Kesteren, Beul et al., 2013; Wang et al., 2012; Yacoby et al., 2023), theorized to reflect the rapid integration of novel information (initially represented in the hippocampus) into existing neocortical knowledge stores (Alonso et al., 2020; McClelland, 2013; McClelland et al., 2020; van Kesteren et al., 2012). By focusing on univariate activation and functional connectivity, however, this prior work could not directly probe the neural representations of related memories. More recently, several studies have directly asked whether consistency with prior knowledge promotes integrative memory representations (Fig. 2a). Audrain and McAndrews (2022) presented participants with scene-item pairs that were either consistent or inconsistent with prior knowledge (e.g., beach-starfish or kitchen-elephant; see Fig. 3a for an illustration). After three days of consolidation, participants were cued with the items and tested for their memory of the associated scene. In the medial prefrontal cortex (mPFC), the neural representations of items previously presented with the same scene became more similar to each other compared to items presented with another scene. Importantly, this pattern of increased similarity was observed only when the items were previously presented with a semantically congruent scene, suggesting that consistency with prior knowledge promotes integration between related memories (but see Tompary and Davachi, 2017).

In another study, Sommer et al. (2022) taught participants hierarchical schema knowledge about a group of insects. After this knowledge
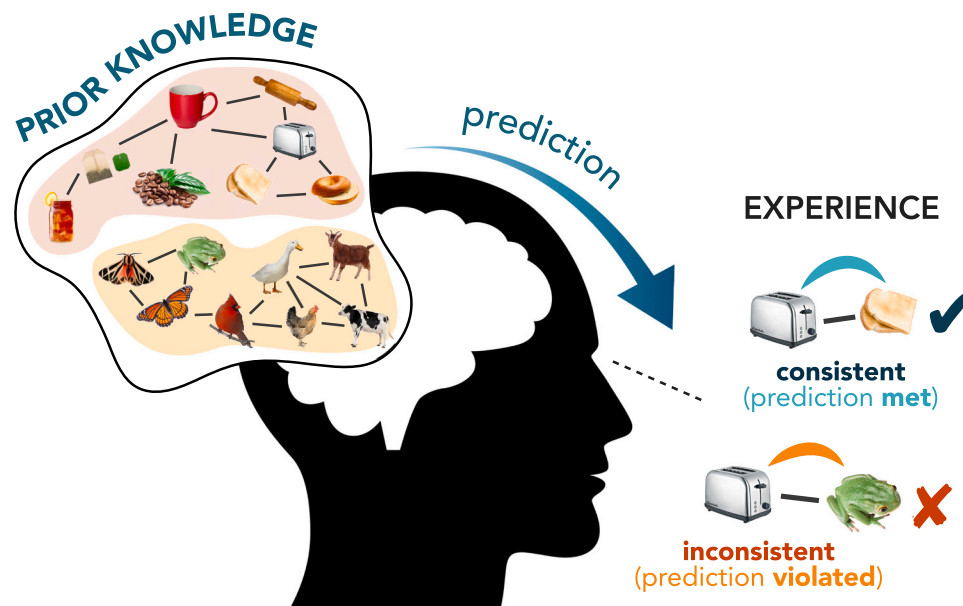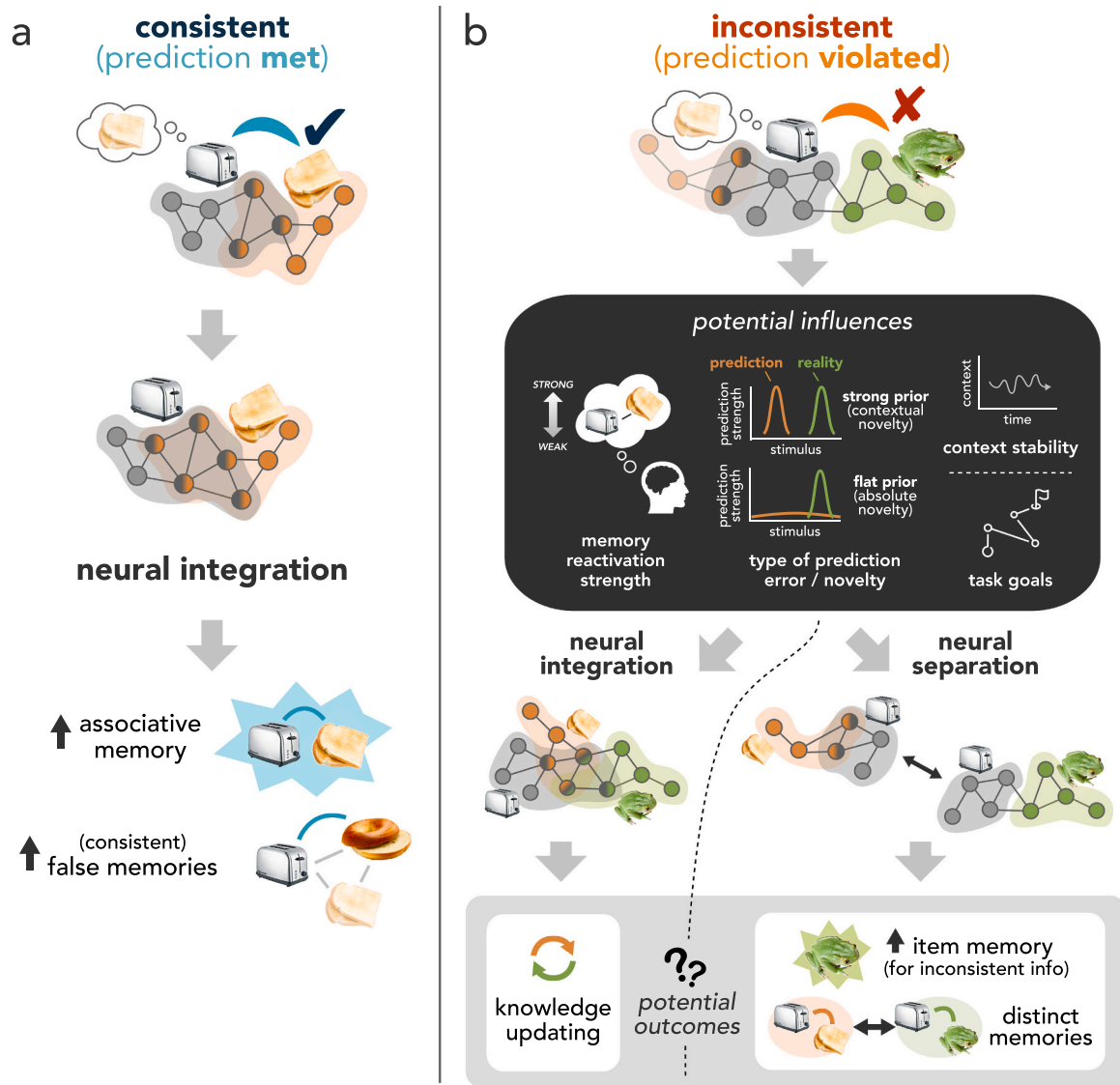

Fig. 1. Prior knowledge allows us to make memory-based predictions about novel events, such as seeing a toaster and expecting that toast will soon follow. These predictions can either be met (i.e., consistent with prior knowledge) or violated (i.e., inconsistent with prior knowledge, such as seeing a frog in your kitchen next to the toaster).

Fig. 2. Hypothesized neural and mnemonic outcomes for events that meet (a) vs. violate (b) our predictions. We conceptualize representations of elements of experiences as nodes with links between them, illustrated here by circles with connecting lines, respectively. These nodes and links can be thought of as belonging to a theoretical neural network or an associative network model, or as neurons (nodes) and synapses (links) in neural ensembles. The top illustrations reflect representations active during encoding. Representational changes can occur already during encoding, or during consolidation or retrieval. (a) When an experience conforms to our predictions, neural integration — that is, increased overlap between memory representations — is likely to occur (resulting in neural similarity). We hypothesize that increased overlap occurs via the strengthening of existing links between nodes representing elements of an experience, or via the creation of new links, potentially leading to more shared nodes. Integration can be either between different elements of the experience (e.g., between novel images of a toaster and toast), or between the novel experience and existing knowledge structures (e.g., between the novel association and your general knowledge about kitchens). Integration can promote both enhanced associative memory as well as increased susceptibility to false memories for other knowledge-consistent details. (b) In contrast, when events violate our predictions, both neural integration and separation — whereby memory representations become *less* similar/overlapping — can occur. Potential factors influencing separation or integration in response to prediction errors are illustrated in the black inset. Separation can happen via multiple mechanisms, such as creating a new representation of the novel experience (toaster-frog association), fully distinct from that of the pre-existing memory prediction (toaster-toast association), or by weakening links between nodes or inactivating nodes initially shared by both elements, resulting in fewer shared nodes (not illustrated here). Integration may promote the updating of prior knowledge (to accommodate the previously unexpected information). Separation may enhance memory for the violating information itself, and lead to distinct memories for both the prior prediction and the unexpected experience. Question marks reflect that these hypotheses remain to be empirically tested. Potential factors that might shape the fate of memories for knowledge-inconsistent information towards either integration or separation are illustrated in the black inset box (see main text for details).

was acquired through several weeks of intensive training, participants learned new schema-related facts about these insects. In both the mPFC and precuneus, the neural representations of new facts about insects in the learned schema were more similar to each other, as compared to the similarity between new facts learned about a different class of organisms (the names of which were familiar to participants, but which were not associated with a hierarchical schema). Further, the neural representations of novel facts about insects were also more similar to the

representations of the schema insects themselves, providing evidence for integration between previously acquired schematic knowledge and new, related facts. Together, these studies suggest that consistency with prior knowledge promotes two types of neural integration: (1) integration between novel events consistent with the same schema, and (2) integration between these memories and prior knowledge representations.

Of note, the interpretation of these neuroimaging studies as memory integration relies on the assumption that increased similarity between
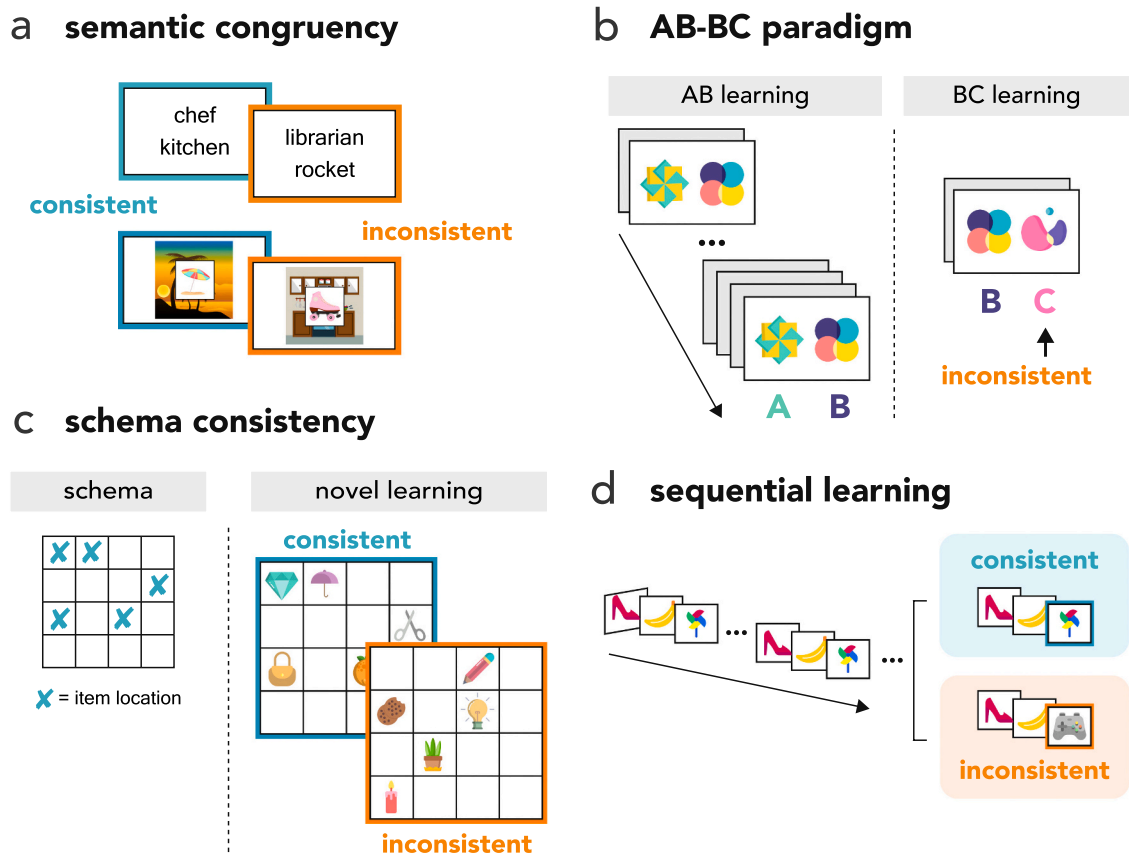
Fig. 3. Illustrative examples of paradigms used to test memory for information that is consistent versus inconsistent with predictions or prior knowledge. (a) Information is either congruent or incongruent with existing semantic knowledge, and therefore is either consistent or inconsistent with predictions arising from semantics. (b) In an AB/BC learning paradigm, participants first learn an association between two items, A-B, and then subsequently see one of those items, B, presented with a new item, C, thereby violating their predictions of having item B associated with item A. (c) In one example of novel information being added to a structured schema, participants first learn a spatial schema (a grid of object locations), then encode grids of novel objects that are either consistent with that spatial schema (learned in the same locations) or inconsistent with it (different locations). This kind of paradigm has also been done using non-spatial schemas, as well as pre-existing schemas learned through lifetime exposure. (d) In a sequential learning task, participants first learn sequences of stimuli through repeated exposure, and are then presented with sequences that either conform to or violate the learned sequence. Note that some previous studies combined elements across paradigms, e.g., sequential learning tasks in which item sequences are consistent or inconsistent with semantic knowledge (e.g., "theater – popcorn – candy" vs. "theater – popcorn – basketball").

neural representations reflects the integration between memories. Although this interpretation is generally accepted, it is plausible that individual memory traces have remained unchanged, but that after learning, the presentation or remembrance of one item triggers the co-activation of associated memories. In this case, increased neural similarity between an item and its associate might reflect the fact that both individual representations are active in the brain at the same time, rather than a fully united and integrated representation. Although it is difficult to disambiguate between these possibilities with tools like fMRI or EEG, both explanations are consistent with the idea that memory integration has occurred. Even if the representations of constituent memories are left separated, the fact that their activation becomes interdependent after learning implies that those representations are linked together in the brain.

Paradigms using stimuli that unfold sequentially (Fig. 3d) also emphasize the role of the hippocampus in the integration of knowledge-consistent events. The hippocampus is known to mediate sequential, memory-based predictions (Kok and Turk-Browne, 2018; Schapiro et al., 2012; Shohamy and Turk-Browne, 2013). Sequential learning, in particular, increases the representational similarity between members of the learned sequence, presumably reflecting prediction of the sequence, as well as neural integration between sequence elements (e.g., Paz et al., 2010; Sakai and Miyashita, 1991; Schapiro et al., 2013). Memory prediction and integration are likely dependent on the process of pattern

completion, whereby a full memory trace (e.g., predicted information, or other related memories) is reinstated from a partial cue. Within the hippocampus, pattern completion is typically associated with subregion CA3 (Hasselmo and Eichenbaum, 2005; Horner and Doeller, 2017; Knierim and Neunuebel, 2016; Marr, 1971; O'Reilly and McClelland, 1994; Treves and Rolls, 1994; Yassa and Stark, 2011). A recent study by Yousuf et al. (2021) first cued participants with a category name (e.g., "furniture"), allowing predictions of possible category members to emerge, and then presented participants with an item that was either a member of the cued category (e.g., "table") or not (e.g., "banana"). Univariate activation in CA3 correlated with memory of category-consistent, but not inconsistent, objects. This result is in line with the idea that CA3 pattern completion promotes memory for events that are consistent with our predictions (see Grande et al., 2019; Horner et al., 2015 for related findings). Pattern completion in CA3 can, in turn, facilitate the full reactivation of related, pre-existing memories in the cortex through hippocampal-cortical connections (Danker et al., 2017; Griffiths et al., 2019; Hindy et al., 2016, 2019; Michelmann et al., 2021; Norman and O'Reilly, 2003; Ritchey et al., 2013; Shohamy and Turk-Browne, 2013).

Thus far, we have discussed evidence that consistency between related experiences might promote neural integration. Do congruency and neural integration impact integrative or associative forms of memory for such experiences? While the study by Yousef et al. (2021) focused

on item memory, numerous other studies provide evidence that consistency with prior knowledge or expectations particularly benefits associative memory (Fig. 2a). For example, memory for the association between two items of a pair is enhanced when items are semantically congruent (e.g., kitchen-microwave; Amer et al., 2018; Bein et al., 2015; Frank et al., 2018; Quent et al., 2022; Tompary et al., 2020; Tompary and Thompson-Schill, 2021; van Kesteren, Beul et al., 2013; Wu et al., 2023). Such semantic congruency also promotes unitization in memory, or the integration of the two components into one representation, which in turn improves speed and accuracy during memory retrieval (Gronau et al., 2008; Tibon et al., 2014; see also Tibon et al., 2018 for related findings). Further, this integration might be directional. In one interesting behavioral study, participants encoded both congruent (e.g., kitchen-microwave) and unrelated (e.g., kitchen-pencil) word pairs (Popov et al., 2019). The authors found that for congruent pairs, associative memory was asymmetric: specifically, cueing participants with `kitchen` during an initial memory test facilitated future retrieval performance when `kitchen` was again used as the memory cue, but *not* when `microwave` was the cue. In contrast, for unrelated pairs, associative memory performance improved across retrieval attempts irrespective of which item in the pair was used as a cue, suggesting that associations were represented symmetrically. An intriguing possibility is that congruency with prior predictions enhances neural integration and associative memory specifically in the direction in which the prediction, or pattern completion, unfolds.

Congruency also enhances the integration of information across different events. For example, Frank et al. (2018) presented participants with sequential pairs of semantically congruent images (e.g., all images related to a "theater" theme). Memory for these pair sequences was compared to sequences in which some of the pairs included a semantically incongruent item, as well as sequences without a coherent theme. Congruency increased memory for pairs, as well as participants' ability to accurately retrieve which item pairs belonged to the same sequence, suggesting that semantic relatedness promoted integration between pairs in the sequence. Additional work has found that semantic relatedness across word pairs (e.g., peace-*razor* and peace-*shave*) can increase the dependency between these memories, such that the more the pairs were related, the higher the likelihood that if one pair was remembered correctly, so would be the other (Antony et al., 2022). This dependency suggests that semantic overlap increases memory integration not only between components of the same experience (e.g., between simultaneously presented items in semantically congruent pairs) but also across congruent experiences from different learning experiences.

In studies that did not address congruency with prior knowledge, the similarity of neural representations in the cortex and the hippocampus has been found to promote associative memory (DuBrow and Davachi, 2014; Ezzyat and Davachi, 2014). Additionally, a recent study showed that neural similarity between exemplars of a familiar semantic category (i.e., 'birds') mediated subsequent memory between the exemplars and a newly learned verbal label (Bruett et al., 2020). While these studies provide some clues, understanding precisely how integrated neural representations promoted by prior knowledge facilitate subsequent memory of that information is an open question.

While prior knowledge provides scaffolding for integrative memory, that same knowledge can also bias memory and lead to mistakes (Fig. 2a). We use our semantic or schematic knowledge to make inferences and fill in gaps in our memory. This reconstruction process usually works, because — by definition — a large proportion of our experiences matches our learned expectations. However, the integration of novel experiences with existing knowledge structures can also lead to false memories, (Alba and Hasher, 1983; Bartlett, 1932; Schacter, 2022). A classic example of this phenomenon is the DRM paradigm, in which participants are presented with a list of semantically related words. At retrieval, they are then likely to incorrectly endorse memory for words that belong to the same semantic context, but were not actually studied (Deese, 1959; Roediger and McDermott, 1995). Such false memories are

associated with increased similarity between neural representations of studied items and those of semantically related but unstudied items, both in the temporal lobe (Chadwick et al., 2016; Zhu et al., 2019), as well as in the left inferior frontal gyrus (LIFG) and parietal cortex (Ye et al., 2016; see also Lee et al., 2019 for related findings, and Dennis et al., 2022 for review).

Using another paradigm, Tompary and Thompson-Schill (2021) presented participants with images of items from different semantic categories in different locations on a screen. Each semantic category was clustered around a location, which allowed the authors to examine how category membership biased memory for item locations. The authors found that location memory for typical category members (e.g., a cardinal is a typical member in the 'birds' category, while a toucan is less so) was more biased towards the cluster center, relative to atypical category members. In another study that used sequential presentation of semantically congruent and incongruent pairs, congruency enhanced memory, but also increased interference from other items in the sequence (Frank et al., 2018). Together, these studies show that integrated representation promoted by prior semantic knowledge can impair memory reconstruction and suggest that the similarity of neural representations might underly these biases. An exciting question is whether the mnemonic benefits of neural integration must necessarily come with the downside of memory biases, or whether these can be dissociated, such that we can benefit from integration, without paying the cost.

2.1. Prior knowledge promotes neural integration regardless of consistency, potentially in different brain regions

Thus far, we have reviewed evidence for the idea that consistency between novel information and prior knowledge or expectations can lead to integrated memory representations. But, is consistency with prior knowledge *required* for cortical integration? Behaviorally, just as prior knowledge benefits memory for congruent information, numerous studies have found enhancements in associative and contextual memory when new information is learned in the presence of prior knowledge, even if the content of that new information is arbitrary with respect to the existing knowledge (Bein et al., 2019; Bellana et al., 2021; DeWitt et al., 2012; Gasser and Davachi, 2023; Liu et al., 2017; Reder et al., 2013, 2016). For example, Gasser and Davachi (2023) found that when participants executed a highly familiar sequence of actions, temporal binding (but not item memory) between novel items studied simultaneously was enhanced, relative to memory for items studied during the execution of an unfamiliar action sequence.

Looking at the brain, Bein et al. (2020) showed that in the LIFG, novel faces associated with famous faces showed increased neural similarity to each other with learning. Critically, this effect was driven by representations of novel faces "moving closer" to those of famous faces, while famous faces' representations remained relatively unchanged, suggesting that the novel faces became assimilated into the representation of their associated famous faces. In another study (Guo and Yang, 2022), participants learned spatial arrangements of objects organized on grids (vs. control objects learned in random locations); see Fig. 3c). On the next day, participants learned associations between these objects and novel scene images. The authors found, first, that items from the same grid (i.e., the same spatial schema) were represented more similarly than those from different grids in the lateral occipital cortex (LOC), suggesting that the items had become integrated based on a shared context. Second, they found greater neural similarity between individual object representations and corresponding object-scene pairs for objects that belonged to a spatial schema, compared to the similarity between objects that did not belong to a schema and their respective scene associates. This latter result suggests that novel learning evokes the reactivation of the schematic or contextual information associated with individual items — consistent with the study by Sommer et al. (2022) mentioned above. Thus, although consistency with prior knowledge

might enhance the reactivation of prior memories (van Kesteren et al., 2020), such reactivation may occur, and in turn facilitate integration, even without consistency between old and new memories (see below for reactivation in the case of knowledge-inconsistent events).

The degree to which consistency is emphasized in a particular paradigm might lead to differences in the cortical regions that construct integrated memory representations. Both Audrain and McAndrews (2022) and Sommer et al. (2022) included explicit manipulations of congruency and found increased similarity in medial cortical regions like the mPFC and precuneus. Other studies without this element of congruency, but which nevertheless engaged prior knowledge, have found similarity effects in lateral brain regions such as LIFG (Bein et al., 2020) and LOC (Guo and Yang, 2022). Prior research has suggested that mPFC detects matches between existing semantic knowledge and new information (Hebscher and Gilboa, 2016; van Kesteren et al., 2012), potentially explaining why integration is found in the mPFC when consistency with prior knowledge is directly manipulated (Audrain and McAndrews, 2022; Sommer et al., 2022). When consistency with prior knowledge is not emphasized, lateral brain regions might be involved in the integration of new information.

Another reason for the medial versus lateral distinction in these studies might be the degree to which prior knowledge is hierarchical. The mPFC in particular has a demonstrated role in representing complex and hierarchical schemas (Baldassano et al., 2018; Masís-Obando et al., 2022; Reagh and Ranganath, 2023), such as that used by Sommer et al. (2022). The task used by Audrain and McAndrews (2022), in turn, was also amenable to a hierarchical structure, in that participants learned scene-item pairs where items either were or were not likely to be found within the higher-level scene context. In contrast, the other aforementioned studies implicated other types of knowledge that are not as strongly hierarchical, such as knowledge about a famous person (Bein et al., 2020), or spatial locations on one grid (Guo and Yang, 2022). This distinction could be important, as recent computational work suggests that hierarchically organized knowledge facilitates integration (McClelland et al., 2020). In sum, how integration promoted by prior knowledge is influenced by the type of prior knowledge, consistency with that knowledge, and the predictions it gives rise to, remains an open question.

3. Prediction errors promote both neural separation and integration

Events that are inconsistent with prior knowledge (i.e., those that elicit prediction errors) should promote learning and memory so that individuals can update inaccurate prior knowledge and generate better predictions in the future. Knowledge updating can be mediated by two processes: On the one hand, memory updating can occur by integrating new information into old knowledge structures, overwriting information that is no longer relevant or accurate. When this process occurs, we would expect to see increased neural similarity (i.e., integration) between old memories and conflicting new information. On the other hand, if pre-existing knowledge might become relevant again in the future, it is more adaptive to create separate representations of old and new memories, such that both pieces of information can be maintained for future use (Gershman et al., 2017; McClelland et al., 1995, 2020). Separate representations can be constructed by the encoding of inconsistent events within the hippocampus (at least initially), separately from existing cortical knowledge stores (Gershman et al., 2014; Love et al., 2004; McClelland et al., 1995, 2020; van Kesteren et al., 2012). Additionally, within the hippocampus, the process of pattern separation, i.e., the allocation of distinct memory representations for overlapping inputs (Treves and Rolls, 1994; Yassa and Stark, 2011), can facilitate the construction of distinct neural representations for events that violate prior knowledge (Frank et al., 2020).

Emerging empirical evidence is consistent with the notion that events that violate prior memories promote both separated and

integrated representations in the hippocampus and the cortex. In one study, Schlichting et al. (2015) presented participants with pairs of items (A-B) during an initial learning phase. Next, previously encoded B items were paired with a new C item (see Fig. 3b for a depiction of the AB/BC learning paradigm). These B-C pairs were likely to elicit prediction errors because they violated the learned expectation that A should appear alongside B. The authors found that in the posterior hippocampus, representations of items A and C were more distinct after relative to before learning, consistent with the idea that these items had become separated in memory. In the anterior hippocampus, however, A-C items became more similar after learning — but only when A-B pairs were strongly learned first, and then A-C pairs were encoded during a second learning phase. When A-B and A-C learning was interleaved, such memory integration did not occur. Outside the hippocampus, Schlichting et al. (2015) found A-C separation and integration in different clusters in the prefrontal cortex.

In another study, Bein et al. (2020) considered predictions borne out of well-learned semantic knowledge. This study presented participants with pairs of one famous and one novel face, and with pairs of two novel faces. The representations of faces in famous-novel pairs became separated in the anterior hippocampus, which was in turn linked to successful associative memory. Conversely, representations in novel-novel face pairs became more similar through learning. One interpretation is that in famous-novel pairs, the famous face reactivated prior knowledge about the relevant celebrity, while the novel face violated these expectations, ultimately promoting separation. In the case of the two novel faces, because there was no prior knowledge, no memory-based predictions were violated, and thus no separation occurred. In contrast to the hippocampus, in the LIFG, novel faces became integrated with famous faces, i.e., with prior knowledge representations.

In a different study, Yacoby et al. (2021) showed that representations of word pairs that were incongruent with prior semantic knowledge (i.e., words that had no meaningful connection to each other, such as "careful" – "tomato"; also see Fig. 3a) were initially more similar to each other in the LIFG, providing evidence of integration between these memories. Interestingly, with learning, their representations became differentiated and reached the same level of distinction as congruent word pairs. Note that Yacoby et al. (2021) did not measure the similarity between the representations of items in a pair (i.e., between components of old and new memories), but rather measured the similarity between different pair representations (that were congruent vs. incongruent). As such, these effects potentially reflect a different process than the previous findings (Bein et al., 2020; Schlichting et al., 2015). Nevertheless, these results show that in the cortex and in the hippocampus, prediction errors evoke neural correlates of separation and integration.

3.1. Influences on memory integration and separation in response to prediction errors

3.1.1. Strength of memory reactivation

Because this work supposes that violations of prior knowledge can trigger distinct kinds of representational changes in resulting memories, it is helpful to consider recent work on mechanisms that push the memory system toward integration or separation (Fig. 2b). Theoretically, the stronger the reactivation of prior memories, the larger the prediction error elicited by an inconsistent event, and the greater the interference between old and new memories. This interference arguably creates a larger need for pattern separation, to avoid confusion between similar but conflicting pieces of information. Separation can also promote the maintenance of both the original and the new memory (Favila et al., 2016; Gershman et al., 2014; Wanjia et al., 2021). Of note, the non-monotonic plasticity hypothesis (NMPH; Norman et al., 2006; Ritvo et al., 2019) argues that when the reactivation of an existing memory is high during novel learning, connections between the two memories will be strengthened, and integration will occur. When reactivation is moderate, these connections will instead be weakened, leading to

separation or differentiation between memories. Low reactivation should have no effects on memory traces. While empirical work using perceptual stimuli found such non-monotonic representational changes (Wammes et al., 2022), research that involved memory-based predictions and prediction errors has generally found a monotonic relationship. Specifically, stronger reactivation of prior memories correlated with more hippocampal pattern separation (Kim et al., 2017; Kim et al., 2020), specifically in the DG/CA3 subregion (Molitor et al., 2021; note that DG and CA3 are collapsed together due to low spatial resolution in human fMRI, and that in CA1, another hippocampal subregion, stronger reactivation led instead to neural integration). Broadly, these studies have interpreted findings of memory separation through the lens of the NMPH, arguing that memory-based predictions evoked in these paradigms led to relatively moderate levels of reactivation in the brain regions examined.

3.1.2. Context stability

One additional possibility is that the relationship between reactivation strength and representational change depends on the stability of the context in which those predictions, or prediction errors, arise (Fig. 2b). Previous work has found strong memory reactivation in response to both knowledge-consistent and knowledge-inconsistent information. In van Kesteren et al.'s (2020) study, participants showed greater memory reactivation of previously-learned A-B associations (measured by self-report and by classification of fMRI multivoxel activation patterns in ventral visual cortex) when novel B-C associations contained information that was semantically congruent (vs. incongruent) with the original A item. Such reactivation, in turn, was associated with better integration between the A and C items. In contrast, Wu et al. (2023) found that neural reactivation (measured by EEG whole-brain pattern similarity) of a just-encoded sequence (e.g., a series of photos depicting someone cooking) was greater when novel information was *incongruent* (e.g., an image of a kite) versus congruent with the preceding sequence. This reactivation was correlated with worse item memory for the novel image. Of note, these studies used different paradigms and reactivation measures. Nevertheless, one speculative reason for the potential discrepancy is the context in which incongruency, or prediction error, occurred. In Wu et al. (2023), prediction errors occurred when incongruent items appeared at the end of a highly stable and coherent event sequence (see also van der Linden et al., 2017). When the unexpected object was encountered, reactivation of the prior event sequence may have been important for disambiguating these conflicting, adjacent experiences, and potentially discarding the novel contradictory information (but see Silva et al., 2019; Sols et al., 2017; Wahlheim et al., 2022; Wahlheim and Zacks, 2019). In van Kesteren et al. (2020), incongruent items encountered across A-B/B-C pairs were separated across time and learning blocks, perhaps lessening the need for disambiguation of these already distinct experiences, and therefore reducing the utility of reactivating existing memories in response to a prediction error (Zeithamova and Preston, 2017). Future work could clarify the consequences of memory reactivation on memory integration and separation across different levels of prediction-(in)consistency and across different learning contexts.

Research on fear learning and extinction further illustrates how the magnitude of prediction error and the context in which they arise can influence memory separation. In these paradigms, animals first complete a learning phase, in which a stimulus is paired with an aversive outcome (e.g., a room or a perceptual cue that predicts the delivery of a disturbing noise). Then, the association is extinguished by no longer pairing the stimulus with the outcome, leading the animal's fear response to diminish or extinguish. If extinction is abrupt — thereby evoking a large prediction error due to the sudden absence of an aversive outcome — animals will later show spontaneous recovery of fear: after some delay, the fear will return (e.g., Bouton, 2004; Rescorla, 2004; Todd et al., 2014). This recovery is presumed to arise because the large prediction error leads the animal to infer that the extinction phase is a

new context (or 'latent cause'; Gershman et al., 2010, 2014, 2017), separate from the original fear context. As a result, the extinction memory is encoded separately from the original fear memory while that fear memory remains intact, which opens the possibility that the original fear will return when the animal is reminded of it. However, it has been shown that if extinction occurs gradually, for example by slowly reducing the frequency of shocks until they are fully eliminated, this should lead to weaker prediction errors. In this case, no new context is inferred, and the original fear memory is instead updated to reflect that the original context is now safe. Indeed, such gradual extinction has been shown to reduce spontaneous recovery (Gershman et al., 2010, 2014; Song et al., 2022; similar notions of context integration vs. separation have also been applied to explain reconsolidation paradigms as in Misanin et al., 1968, Nader et al., 2000, Riccio et al., 2006, and Schiller et al., 2010; see Gershman et al., 2017). Broadly, this work provides additional evidence that prediction errors lead to memory separation, while also adding that the transformation of memories in response to prediction errors depends on both the strength of prediction errors and the inferences made about their source (Gershman et al., 2010, 2017; Yu, 2021).

3.1.3. Types of prediction errors and novelty

The *type of prediction error* generated by an event might also bias memories toward separation versus integration (Fig. 2b). Previous research has used a Bayesian framework to conceptualize the relationship between prediction error and memory (Greve et al., 2017; Quent et al., 2021). In this work, a prediction error is conceptualized as the difference between a 'prior', which encompasses one's expectations about what is likely to occur (based on existing knowledge), and the 'evidence', which reflects knowledge about what actually happened. These prior expectations can take many forms. For example, if a person has a 'flat prior' (indicating that they have no specific prediction about what is likely to occur), and then they encounter new evidence, this mismatch between expectations and reality would elicit a prediction error. However, a prediction error also arises when a person has a very clear prior (indicating that they have strong expectations about what is likely to occur), but then encounters contrasting evidence. While both cases generate prediction errors, and both have been shown to promote associative memory (Greve et al., 2017), these phenomenologically different prediction errors might differentially affect memory separation and integration.

Similarly, previous conceptualizations differentiated between `absolute novelty` and `contextual novelty` and proposed that they might be mediated by different neuromodulatory systems (Kafkas and Montaldi, 2018b). These, in turn, could promote integration versus separation (Fig. 2b). Absolute novelty can be thought of as a prediction error coming from a flat prior: when a situation is fully novel, we should have no prior expectations about what is likely to occur. In contrast, contextual novelty arises when one has specific expectations in a given context, but then learns through evidence that their expectations were wrong (e.g., seeing a pillow case in the kitchen). It has been suggested that contextual novelty might promote dopamine input to the hippocampus, while absolute novelty might enhance acetylcholine input (Hasselmo et al., 1996; Kafkas and Montaldi, 2018b; Lisman and Grace, 2005; Meeter et al., 2004). Acetylcholine has been found to enhance pattern separation and reduce pattern completion in hippocampal subfields (Duncan and Schlichting, 2018; Hasselmo, 2006). Consistent with these findings, task manipulations that have putatively upregulated acetylcholine activity strengthened behavioral markers of pattern separation (Ruiz et al., 2021; Duncan et al., 2012). Interestingly, Sinclair et al. (2021) linked activity in the basal forebrain, a primary source of acetylcholine input to the hippocampus, with increased hippocampal pattern separation for *contextual* novelty. Thus, contextual novelty might promote memory through increasing separation and reducing pattern completion and integration. Intuitively, it makes sense that promoting separation while reducing pattern completion might be

especially important in the case of contextual prediction errors, given that context-based expectations will lead to the completion of the original prediction, and thus could override new, contrasting evidence (Bein et al., 2020). Conversely, when there is no strong prediction, as in the case of a flat prior or absolute novelty, suppression of hippocampal pattern completion might not be necessary to the same extent. In these circumstances, integration of novel information with existing knowledge, rather than separation, may be more likely to occur.

Interestingly, Duszkiewicz et al. (2019) make a similar distinction between types of novelty, but focus on different sources of dopaminergic inputs to the hippocampus and hypothesize different integration and separation outcomes. They define novel experiences with minimal connections to prior knowledge as 'distinct novelty' (similar to absolute novelty), while experiences that share some commonality with prior knowledge are termed 'common novelty'. Common novelty, for example, might involve being in a familiar context that has some distinct, unexpected features (e.g., finding a pillowcase in a kitchen). Distinct novelty is said to activate dopaminergic input to the hippocampus from the locus coeruleus. This input is dense and leads to activation of co-occurring contextual details, which creates a distinct representation of that novel event embedded in its specific episodic context (Duszkiewicz et al., 2019). In contrast, common novelty, much like contextual novelty, is proposed to promote dopaminergic input from the ventral tegmental area (VTA) into the hippocampus, which is more targeted and therefore promotes only the reactivation of other related memories (e.g., memories of other kitchens, or general schematic knowledge about kitchens) and leads to integration of the new experience with existing knowledge (Duszkiewicz et al., 2019). In our view, an integrated representation between experiences that share common features is dependent on the extent to which the unique elements of the novel experience are consistent with prior predictions elicited by the shared context, or whether they trigger a prediction error (e. g., finding a novel piece of cookware in a kitchen vs. finding a pillowcase). Nevertheless, it remains an open question precisely how different kinds of novel experiences (and the predictions or prediction errors that arise) impact neuromodulatory input to the hippocampus to bias the memory system toward integration versus separation (see Rouhani et al., 2023 for a related discussion).

3.1.4. Goals

Ultimately, integration and separation of knowledge-inconsistent events should serve an individual's current goals (e.g., the need to distinguish related events) or future goals (e.g., the need to update knowledge vs. preserve old knowledge that might become relevant again) (Fig. 2b). Several studies show that neural pattern differentiation arises during situations where similar inputs lead to different outcomes, increasing the need for related memories to be stored as non-overlapping representations (Chanales et al., 2017; Favila et al., 2016; Kim et al., 2017; Wanjia et al., 2021). Indeed, in these situations when prediction errors arise, it can be beneficial to retain memory for both the original memory-based prediction and the violating novel information — perhaps even regardless of how strong the original prediction was. In other cases, the strength of memory reactivation can be indicative of future relevance. For example, the knowledge that restaurants are typically found in airports is very well-established and therefore is likely to be strongly reactivated when we enter an airport. This prediction of airports having restaurants is likely to remain relevant in the future, even if we encounter one experience that violates this expectation, such as visiting one small airport that has no restaurants (Yu et al., 2021). In this example, the strong reactivation of prior memories, and the presumably strong prediction error driven by the restaurant-less airport, might lead to the inference that this is another category of airport (e.g., a small, regional airport), or else an aberrant example. As such, an individual may be more likely to create a separate memory for the unexpected event, thereby achieving the goal of preserving their prior memory (Gershman et al., 2014, 2017; Love et al., 2004).

3.1.5. Intrinsic neural dynamics

Interestingly, in some of the studies we reviewed (Bein et al., 2020; Molitor et al., 2021; Schlichting et al., 2015), separation and integration both occurred, in different brain regions, under the same external conditions. An interesting possibility is that neural reactivation and prediction error signals might manifest to different degrees in different brain regions (Bar, 2009; Danker and Anderson, 2010; Kok and Turk-Browne, 2018; Ritchey et al., 2013; Tompary et al., 2016; Xue, 2018), potentially due to variation in intrinsic neural dynamics (e.g., the amount of excitatory versus inhibitory activity in a given brain region; Ritvo et al., 2019) — and that, in turn, could differentially promote the integration versus separation of neural representations. Thus, the brain might simultaneously create multiple representations to promote both updating of prior knowledge via integration and the preservation of prior knowledge alongside new conflicting memories via separation.

3.2. Prediction errors promote a variety of mnemonic outcomes

Does neural separation versus integration of knowledge-inconsistent events lead to different mnemonic outcomes? Theoretically, separated representations can lead to a distinct memory of the violating event, or a reduction in the integration between old and new memories, while integrated representations should promote associative memory between old knowledge and the new, violating event. Several recent behavioral studies have found evidence for both of these phenomena (Antony et al., 2021; Bein et al., 2021; Ben-Yakov et al., 2021; Brod et al., 2018; Greve et al., 2017; Kafkas and Montaldi, 2018a; Wahlheim et al., 2022; Wahlheim and Zacks, 2019). Providing evidence for enhanced integrative or associative memory, Greve et al. (2017) taught participants, through repeated exposures, that different scene categories predict either positive- or negative-valence words. Then, they violated this learned expectation by altering the valence of the words that followed a given scene. Memory for the word-scene association was higher for words that violated previously learned contingencies (Greve et al., 2017). Similarly, Kafkas and Montaldi (2018) taught participants that a specific symbol was followed by either a man-made object or natural object, but then subsequently switched this contingency to violate participants' predictions. Recollection rates, thought to reflect participants' rich, integrated memory for experienced events, were higher for objects that violated prior expectations (Kafkas & Montaldi, 2018). An additional study violated predictions generated by semantic knowledge by presenting participants with objects in highly surprising locations (e.g., a microwave in the kitchen sink) and also found increased recollection for surprising object-scene associations (Quent et al., 2022).

In contrast, supporting the idea that prediction errors can lead to separated memories (i.e., disruption of the association between connected experiences), other work has found that changes in context (termed 'event boundaries') — which can be said to violate expectations of contextual stability — impair associative memory between information from across different contexts (Ben-Yakov et al., 2021; DuBrow and Davachi, 2013; Ezzyat and Davachi, 2014; Heusser et al., 2018; Rouhani et al., 2020). Changes in context also lead to neural separation of event representations (Antony et al., 2021; Baldassano et al., 2017; Bein and Davachi, 2022; Ben-Yakov and Henson, 2018), the strength of which correlates with how separated in time two events are remembered (Ezzyat and Davachi, 2014). These studies are consistent with the idea that prediction errors facilitate the creation of distinct memory traces by promoting neural separation. However, other work has found that neural and behavioral separation between memories also occurs when event boundaries are predictable, in that participants can anticipate precisely when the current event will end and the next will begin, as well as the content of the next event (Bein and Davachi, 2022; Clewett and Davachi, 2017). As such, it is unclear to what degree these effects stem from violated expectations per se (see Shin and DuBrow, 2021 for a related discussion).

While the aforementioned studies generally examined how

prediction errors impact memory for the *associations* between elements of an experience (e.g., links between items across time or space), other work has found evidence that prediction errors can also increase the distinctiveness of memories for *individual* pieces of information. Such work has often used the mnemonic similarity task (MST; Stark et al., 2019), which is typically thought of as a behavioral assay of pattern separation. In this task, participants encode a stream of visual items. Then, during a memory test, they are presented with identical items, visually similar lures, and novel items, and are asked to indicate whether each item they see is exactly the same as a previously studied item ('old'), similar to an encoded item, or completely new. To discriminate old from similar items, participants must have access to fine-grained and detailed item representations, potentially reflecting that the process of pattern separation has taken place. In line with this interpretation, the successful identification of a similar item as 'similar' has been linked with increased univariate BOLD activity in hippocampal subregion DG, as well as with more distinct item representations in DG, as indicated by the successful classification of similar lure versus previously-encoded images using DG multivariate activity patterns (Baker et al., 2016; Bakker et al., 2008; Berron et al., 2016; Knierim and Neunuebel, 2016; Lacy et al., 2011; Leutgeb et al., 2007).

Bein et al. (2021) used this type of MST test (i.e., requiring participants to discriminate between old/similar/new items) for items that either violated predictions acquired during earlier learning, or did not. They found that such prediction violations enhanced item memory: participants correctly identified more identical old items as 'old' when items violated their expectations, relative to when items did not violate prior predictions. However, participants did not perform better at identifying similar lures for violations versus expected items. While these findings are consistent with enhanced detail memories for violations, it does not neatly sit with a pattern separation account, which would have additionally predicted better discrimination between old items and similar lures. One alternate interpretation of these results is that distinct memory representations for violations might be supported by increased processing of perceptual information inputted from the entorhinal cortex to the hippocampus (Colgin, 2016; Hasselmo et al., 1996; Hasselmo and Stern, 2014). In previous work, we have shown that functional connectivity between the hippocampus and entorhinal cortex increases during mnemonic prediction errors (Bein, Duncan, et al., 2020). Thus, prediction errors elicited during Bein et al.'s (2021) study may have upregulated input from the entorhinal cortex and led to the construction of detailed, perceptually rich item representations, which enabled correct identification of an old item as 'old.' However, this perceptual input might not be sufficient to clearly distinguish a similar lure from a previously-seen item — a task that instead requires DG pattern separation. Additionally, the perirhinal cortex, an adjacent brain region, supports recognition memory for items (Brown and Aggleton, 2001; Davachi, 2006; Davachi et al., 2003; Davachi and Wagner, 2002; Eichenbaum et al., 2007; Staresina et al., 2011; Staresina and Davachi, 2008) and is preferentially engaged during mnemonic prediction errors (Chen et al., 2015).

Adding to this complexity is the observation that prediction errors do not always promote memory. First, the most robust finding reported in studies manipulating congruency with semantic knowledge is that events that are consistent with our semantic knowledge are remembered *better* than inconsistent events that presumably elicit prediction errors (Bein et al., 2015; Bonasia et al., 2018; Craik and Tulving, 1975; van Kesteren, Beul et al., 2013; van Kesteren, Rijpkema et al., 2013). In Frank et al.'s study (2018), researchers even used a design that emphasized sequential predictions and included a third baseline (unrelated) condition, and still did not find memory enhancement for incongruent events. Specifically, as described previously, this study presented participants with sequentially ordered quartets of items that either all belonged to the same semantic theme (consistent), contained one item belonging to a different theme (inconsistent), or contained four items from different themes (unrelated). Generally, memory for items

from incongruent sequences was comparable to memory for unrelated sequences — both of which were lower than memory for congruent items (see also van der Linden et al., 2017). In another study, Ortiz--Tudela et al. (2023) taught participants scene-object contingencies, comparing situations where participants have acquired strong prior beliefs (i.e., scenes that have a high probability of being followed by an object from one particular object category, making them highly predicted, and low probabilities of two other categories, making objects from these categories unexpected) versus cases where they have a flat prior (i.e., scenes that are equally likely to be followed by an object from any one of three possible categories). Item memory was highest for objects associated with the flat prior, compared to the strong prior — both when objects belonged to the highly predicted category and when they belonged to a strongly *unexpected* category (see also Ortiz-Tudela et al., 2018). In a similar paradigm, Turan et al. (2023) found better memory for items that met predictions compared to violating them, consistent with a semantic congruency effect.

Importantly, these observed memory benefits for congruent but not incongruent information cannot readily be explained by the use of pre-existing knowledge during retrieval (e.g., due to participants inferring or guessing an item's associate based on semantic knowledge). In particular, researchers have found better memory for congruent over incongruent information when participants performed recognition tests for *individual* items that had belonged to either congruent or incongruent associations at encoding, but had no congruency/incongruency status during retrieval (Bein et al., 2015; Bonasia et al., 2018; van Kesteren, Beul et al., 2013; van Kesteren, Rijpkema et al., 2013). Nevertheless, it could still be argued that semantic associations informed item memory in these studies. Other congruency studies used an associative recognition test that directly controlled for this possibility and found a similar congruency advantage (Amer et al., 2018, 2019; but see Quent et al., 2022). In these studies, participants encoded grocery items presented with prices that were either congruent or incongruent with prior knowledge. At retrieval, participants had to choose the correct price out of two options, both with the same congruency or incongruency status; as such, congruency at retrieval could not explain better memory. Finally, in other studies, expectations were learned within the experiment de novo, such that pre-existing knowledge would not be informative, and a memory advantage was found for events consistent with these learned predictions (Gasser and Davachi, 2023; Ortiz-Tudela et al., 2023; Turan et al., 2023).

The general level of unexpectedness or novelty in the environment might also influence the memory of unexpected events. Broadly, the volatility of an environment is known to influence learning and memory (Daw et al., 2005; Nicholas et al., 2022; Rouhani and Niv, 2021; Yu et al., 2021). In oddball paradigms, rare events (which presumably violate expectations) are typically remembered better than common, expected events (Hunt, 1995; Ranganath and Rainer, 2003; von Restorff, 1933). However, this oddball memory enhancement has been found even if the rare event is the first item in a list, before any predictions can be established or violated, suggesting that other mechanisms beyond prediction error are at play (Hunt, 2006; Waddill & McDaniel, 1998). In addition, although semantically incongruent word pairs are remembered better when they are rare in a learning context (e.g., when there are only a handful of incongruent pairs in a list of otherwise semantically congruent pairs), this effect of rarity is not found for congruent word pairs (Reggev 2018). These results suggest that unexpectedness within a context might benefit memory particularly for events that also violate semantic expectations. Relatedly, hippocampal univariate activity has been shown to decrease when rare events (e.g., oddballs) become more familiar or expected (Bunzeck and Düzel, 2006). We are unaware of studies examining how the rarity of events that violate our expectations, or the general volatility or uncertainty in one's environment, affect the neural integration or separation of such violating events. Nevertheless, the studies surveyed here suggest intriguing avenues for further research.

Overall, the specific details that might contribute to differences in the transformation of memories that are inconsistent with prior knowledge remain to be elucidated. Based on the neural studies and theories surveyed here, we propose that several factors — including the balance between neural separation and integration, the types of prediction errors elicited, the amount of uncertainty in one's environment, the level of memory reactivation, and the type of memory tested — can be jointly considered to understand exactly when and how prediction errors benefit memory.

4. Conclusion

In order to survive and behave adaptively in a changing environment, we need to update our knowledge. This process, however, is not trivial. Sometimes, we might dismiss new information if we deem it as unimportant or inaccurate (e.g., "fake news"). In other cases, when new information is perceived as useful and reliable, knowledge updating is beneficial. In this review, we argued that our brains might flexibly use the integration and separation of memories to answer the different demands of knowledge updating when events are consistent versus inconsistent with our predictions. Many exciting open questions remain, regarding not only how new information is learned, but also how old knowledge is modified or preserved, as well as how these mechanisms serve current and future goals. Further, while past research has focused on brief, individual events that are either broadly consistent or inconsistent with prior memories, our life events often extend across longer periods of time and include multiple components, each with varying degrees of consistency or inconsistency. Making detailed predictions across all of these components and predicting far into the future might be too costly for the brain to perform, leading it to prioritize some kinds of predictions over others (e.g., Brunec and Momennejad, 2022; Lee et al., 2021; Liberman et al., 2002; Niv, 2019; Niv et al., 2015; Trope and Liberman, 2003, 2010). Thus, understanding how predictions and memory for novel information interact in complex and temporally extended events is an open question. Future research could also continue to investigate how integration and separation contribute to the transformation of memories as they undergo consolidation (Audrain and McAndrews, 2022; Gilboa and Moscovitch, 2021; McClelland et al., 1995; Tompary et al., 2020; Tompary and Davachi, 2017). This work could promote a better understanding of the life cycle of memories and how they are interwoven together to promote adaptive learning and behavior.

Acknowledgments

References

Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. 8 (10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011.

Alba, J.W., Hasher, L., 1983. Is memory schematic. Psychol. Bull. 93 (2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203.

Alonso, A., van der Meij, J., Tse, D., Genzel, L., 2020. Na 0elr Meijorg/10.1037/0033-2909.93.2.203ceptual learningn (T32MH065, 239821282094868 Brain Neurosci. Adv. 4. https://doi.org/10.1177/2398212820948686.

Amer, T., Giovanello, K.S., Grady, C.L., Hasher, L., 2018. Age differences in memory for meaningful and arbitrary associations: a memory retrieval account. Psychol. Aging 33 (1), 74–81. https://doi.org/10.1037/pag0000220.

Amer, T., Giovanello, K.S., Nichol, D.R., Hasher, L., Grady, C.L., 2019. Neural correlates of enhanced memory for meaningful associations with age. Cereb. Cortex 29 (11), 4568–4579. https://doi.org/10.1093/cercor/bhy334.

Anderson, J.R., 1974. Retrieval of propositional information from long-term memory. Cogn. Psychol. 6 (4), 451–474. https://doi.org/10.1016/0010-0285(74)90021-8.

Anderson, J.R., Milson, R., 1989. Human memory: an adaptive perspective. *Psychol. Rev.* 96 (4), 703–719. https://doi.org/10.1037/0033-295X.96.4.703.

Antony, J.W., Hartshorne, T.H., Pomeroy, K., Gureckis, T.M., McDougle, S.D., Norman, K.A., 2021. Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. Neuron 109, 1–14.

Antony, J.W., Romero, A., Vierra, A.H., Luenser, R.S., Hawkins, R.D., Bennion, K.A., 2022. Semantic relatedness retroactively boosts memory and promotes memory interdependence across episodes. eLife 11, e72519. https://doi.org/10.7554/eLife.72519.

Audrain, S., McAndrews, M.P., 2022. Schemas provide a scaffold for neocortical integration of new memories over time. Nat. Commun. 13 (1), 5795. https://doi.org/10.1038/s41467-022-33517-0.

Baker, S., Vieweg, P., Gao, F., Gilboa, A., Wolbers, T., Black, S.E., Rosenbaum, R.S., 2016. The Human Dentate Gyrus Plays a Necessary Role in Discriminating New Memories. Curr. Biol. 26 (19), 2629–2634. https://doi.org/10.1016/j.cub.2016.07.081.

Bakker, A., Kirwan, C.B., Miller, M., Stark, C.E.L., 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science 319 (5870), 1640–1642. https://doi.org/10.1126/science.1152882.

Baldassano, C., Hasson, U., Norman, K.A., 2018. Representation of real-world event schemas during narrative perception. J. Neurosci. 38 (45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J.W., Hasson, U., Norman, K.A., 2017. Discovering event structure in continuous narrative perception and memory. e5 Neuron 95 (3), 709–721. https://doi.org/10.1016/j.neuron.2017.06.041.

Bar, M., 2009. The proactive brain: memory for predictions. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364 (1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310.

Bartlett, F.C., 1932. Remembering: A Study in Experimental and Social Psychology. Cambridge University Press.

Bein, O., Reggev, N., Maril, A., 2014. Prior knowledge in fluences on hippocampus and medial prefrontal cortex interactions in subsequent memory. Neuropsychologia 64, 320–330. https://doi.org/10.1016/j.neuropsychologia.2014.09.046.

Bein, O., Trzewik, M., Maril, A., 2019. The role of prior knowledge in incremental associative learning: An empirical and computational approach. J. Mem. Lang. 107, 1–24. https://doi.org/10.1016/j.jml.2019.03.006.

Bein, O., Reggev, N., Maril, A., 2020. Prior knowledge promotes hippocampal separation but cortical assimilation in the left inferior frontal gyrus. Nat. Commun. 11 (1), 4590. https://doi.org/10.1038/s41467-020-18364-1.

Bein, O., Plotkin, N.A., Davachi, L., 2021. Mnemonic prediction errors promote detailed memories. Learn. Mem. 28 (11), 422–434. https://doi.org/10.1101/lm.053410.121.

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., Maril, A., 2015. Delineating the effect of semantic congruency on episodic memory: The role of integration and relatedness. PLoS One 10 (2), e0115624. https://doi.org/10.1371/journal.pone.0115624.

Bein, O., & Davachi, L. (2022). Event integration and temporal pattern separation: How hierarchical knowledge emerges in hippocampal subfields through learning [Preprint]. https://doi.org/10.1101/2022.07.18.500527.

Bellana, B., Mansour, R., Ladyka-Wojcik, N., Grady, C.L., Moscovitch, M., 2021. The influence of prior knowledge on the formation of detailed and durable memories. J. Mem. Lang. 121, 104264 https://doi.org/10.1016/j.jml.2021.104264.

Ben-Yakov, A., Henson, R.N., 2018. The hippocampal film editor: sensitivity and specificity to event boundaries in continuous experience. J. Neurosci. 38 (47), 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018.

Ben-Yakov, A., Smith, V., Henson, R., 2021. The limited reach of surprise: evidence against effects of surprise on memory for preceding elements of an event. Psychon. Bull. Rev. https://doi.org/10.3758/s13423-021-01954-5.

Berron, D., Schütze, H., Maass, A., Cardenas-Blanco, A., Kuijf, H.J., Kumaran, D., Düzel, E., 2016. Strong Evidence for Pattern Separation in Human Dentate Gyrus. J. Neurosci. 36 (29), 7569–7579. https://doi.org/10.1523/JNEUROSCI.0518-16.2016.

Bonasia, K., Sekeres, M.J., Gilboa, A., Grady, C.L., Winocur, G., Moscovitch, M., 2018. Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic events at short and long delays. *Neurobiol. Learn. Mem.* 26–39. https://doi.org/10.1016/j.nlm.2018.02.017.

Bouton, M.E., 2004. Context and behavioral processes in extinction. Learn. Mem. 11 (5), 485–494.

Brod, G., Shing, Y.L., 2018. Specifying the role of the ventromedial prefrontal cortex in memory formation. Neuropsychologia 111, 8–15. https://doi.org/10.1016/j.neuropsychologia.2018.01.005.

Brod, G., Hasselhorn, M., Bunge, S.A., 2018. When generating a prediction boosts learning: the element of surprise. Learn. Instr. 55, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013.

Brod, G., Lindenberger, U., Werkle-Bergner, M., Shing, Y.L., 2015. Differences in the neural signature of remembering schema-congruent and schema-incongruent events. Neuroimage 117, 358–366. https://doi.org/10.1016/j.neuroimage.2015.05.086.

Brod, G., Lindenberger, U., Wagner, A.D., Shing, Y.L., 2016. Knowledge acquisition during exam preparation improves memory and modulates memory formation. J. Neurosci. 36 (31), 8103–8111. https://doi.org/10.1523/JNEUROSCI.0045-16.2016.

Brown, M.W., Aggleton, J.P., 2001. Recognition memory: what are the roles of the perirhinal cortex and hippocampus. Nat. Rev. Neurosci. 2 (1), 51–61. https://doi.org/10.1038/35049064.

Bruett, H., Calloway, R.C., Tokowicz, N., Coutanche, M.N., 2020. Neural pattern similarity across concept exemplars predicts memory after a long delay. NeuroImage 219, 117030. https://doi.org/10.1016/j.neuroimage.2020.117030.

Brunec, I.K., Momennejad, I., 2022. Predictive representations in hippocampal and prefrontal hierarchies. J. Neurosci. 42 (2), 299–312. https://doi.org/10.1523/JNEUROSCI.1327-21.2021.

Buckner, R.L., 2010. The role of the hippocampus in prediction and imagination. Annu. Rev. Psychol. 61 (1), 27–48. https://doi.org/10.1146/annurev. psych.60.110707.163508.

Bunzeck, N., D & e, E., 2006. Absolute coding of stimulus novelty in the human substantia nigra/VTA. Neuron 51 (3), 369–379. https://doi.org/10.1016/j. neuron.2006.06.021.

Chadwick, M.J., Anjum, R.S., Kumaran, D., Schacter, D.L., Spiers, H.J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. Proceedings of the National Academy of Sciences, *113*(36), 10180–10185. https://doi.org/10.1073/pnas.1610686113.

Chanales, A.J.H., Oza, A., Favila, S.E., Kuhl, B.A., 2017. Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. Curr. Biol. 27 (15), 2307–2317. https://doi.org/10.1016/j.cub.2017.06.057.

Chen, J., Cook, P.A., Wagner, A.D., 2015. Prediction strength modulates responses in human area CA1 to sequence violations. *J. Neurophysiol.* 00149, 2015. https://doi.org/10.1152/jn.00149.2015.

Clewett, D., Davachi, L., 2017. The Ebb and flow of experience determines the temporal structure of memory. Curr. Opin. Behav. Sci. 17, 186–193. https://doi.org/10.1016/j.cobeha.2017.08.013.

Colgin, L.L., 2016. Rhythms of the hippocampal network. Nat. Rev. Neurosci. 17 (4), 239–249. https://doi.org/10.1038/nrn.2016.21.

Collins, A.M., Loftus, E.F., 1975. A spreading-activation theory of semantic processing. Psychol. Rev. 82 (6), 407–428. https://doi.org/10.1037/11571-006.

Craik, F.I., Tulving, E., 1975. Depth of processing and the retention of words in episodic memory. J. Exp. Psychol. Gen. 104 (3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268.

Danker, J.F., Anderson, J.R., 2010. The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. Psychol. Bull. 136 (1), 87–102. https://doi.org/10.1037/a0017937.

Danker, J.F., Tompary, A., Davachi, L., 2017. Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. Cereb. Cortex 27 (7), 3515–3524. https://doi.org/10.1093/cercor/bhw146.

Davachi, L., 2006. Item, context and relational episodic encoding in humans. Curr. Opin. Neurobiol. 16 (6), 693–700. https://doi.org/10.1016/j.conb.2006.10.012.

Davachi, L., Wagner, A.D., 2002. Hippocampal contributions to episodic encoding: insights from relational and item-based learning. J. Neurophysiol. 88 (2), 982–990. https://doi.org/10.1152/jn.2002.88.2.982.

Davachi, L., Mitchell, J.P., Wagner, A.D., 2003. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. Proc. Natl. Acad. Sci. USA 100 (4), 2157–2162. https://doi.org/10.1073/pnas.0337195100.

Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8 (12), 1704–1711. https://doi.org/10.1038/nn1560.

Deese, J., 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. J. Exp. Psychol. 58 (1), 17–22. https://doi.org/10.1037/h0046671.

Dennis, N.A., Chamberlain, J., & Carpenter, C.M. (2022). False Memories: What Neuroimaging Tells Us About How We Mis-remember the Past [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/hrj8v.

DeWitt, M.R., Knight, J.B., Hicks, J.L., Ball, B.H., 2012. The effects of prior knowledge on the encoding of episodic contextual details. Psychon. Bull. Rev. 19 (2), 251–257. https://doi.org/10.3758/s13423-011-0196-4.

DuBrow, S., Davachi, L., 2013. The influence of context boundaries on memory for the sequential order of events. J. Exp. Psychol.: Gen. 142 (4), 1277–1286. https://doi.org/10.1037/a0034024.

DuBrow, S., Davachi, L., 2014. Temporal memory is shaped by encoding stability and intervening item reactivation. J. Neurosci. 34 (42), 13998–14005. https://doi.org/10.1523/JNEUROSCI.2535-14.2014.

Duncan, K.D., Schlichting, M.L., 2018. Hippocampal representations as a function of time, subregion, and brain state. Neurobiol. Learn. Mem. 153, 40–56. https://doi.org/10.1016/j.nlm.2018.03.006.

Duszkiewicz, A.J., McNamara, C.G., Takeuchi, T., Genzel, L., 2019. Novelty and dopaminergic modulation of memory persistence: a tale of two systems. Trends Neurosci. 42 (2), 102–114. https://doi.org/10.1016/j.tins.2018.10.002.

Eichenbaum, H., Yonelinas, A.P., Ranganath, C., 2007. The medial temporal lobe and recognition memory. Annu. Rev. Neurosci. 30, 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328.

Ezzyat, Y., Davachi, L., 2014. Similarity breeds proximity: pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. Neuron 81 (5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042.

Favila, S.E., Chanales, A.J.H., Kuhl, B.A., 2016. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. Nat. Commun. 7, 11066. https://doi.org/10.1038/ncomms11066.

Fernández, G., Morris, R.G.M., 2018. Memory, novelty and prior knowledge. Trends Neurosci. 41 (10), 654–659. https://doi.org/10.1016/j.tins.2018.08.006.

Frank, D., Montemurro, M.A., Montaldi, D., 2020. Pattern separation underpins expectation-modulated memory. J. Neurosci. 40 (17), 3455–3464. https://doi.org/10.1523/JNEUROSCI.2047-19.2020.

Frank, D., Montaldi, D., Wittmann, B., Talmi, D., 2018. Beneficial and detrimental effects of schema incongruence on memory for contextual events. Learn. Mem. 25 (8), 352–360. https://doi.org/10.1101/lm.047738.118.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360 (1456), 815–836. https://doi.org/10.1098/rstb.2005.1622.

Friston, K., 2018. Does predictive coding have a future. Nat. Neurosci. 21 (8), 1019–1021. https://doi.org/10.1038/s41593-018-0200-7.

Gasser, C., Davachi, L., 2023. Cross-modal facilitation of episodic memory by sequential action execution, 095679762311582 Psychol. Sci.. https://doi.org/10.1177/09567976231158292.

Gershman, S.J., Blei, D.M., Niv, Y., 2010. Context, learning, and extinction. Psychol. Rev. 117 (1), 197–209. https://doi.org/10.1037/a0017808.

Gershman, S.J., Radulescu, A., Norman, K.A., Niv, Y., 2014. Statistical computations underlying the dynamics of memory updating. PLoS Comput. Biol. 10 (11), e1003939 https://doi.org/10.1371/journal.pcbi.1003939.

Gershman, S.J., Monfils, M.-H., Norman, K.A., Niv, Y., 2017. The computational nature of memory modification. Elife 6. https://doi.org/10.7554/eLife.23763.

Gilboa, A., Marlatte, H., 2017. Neurobiology of schemas and schema-mediated memory. Trends Cogn. Sci. 21 (8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013.

Gilboa, A., Moscovitch, M., 2021. No consolidation without representation: correspondence between neural and psychological representations in recent and remote memory. Neuron 109 (14), 2239–2255. https://doi.org/10.1016/j.neuron.2021.04.025.

Grande, X., Berron, D., Horner, A.J., Bisby, J.A., Düzel, E., Burgess, N., 2019. Holistic recollection via pattern completion involves hippocampal subfield CA3. J. Neurosci. 39 (41), 8100–8111. https://doi.org/10.1523/JNEUROSCI.0722-19.2019.

Greve, A., Cooper, E., Kaula, A., Anderson, M.C., Henson, R., 2017. Does prediction error drive one-shot declarative learning? J. Mem. Lang. 94, 149–165. https://doi.org/10.1016/j.jml.2016.11.001.

Griffiths, B.J., Parish, G., Roux, F., Michelmann, S., van der Plas, M., Kolibius, L.D., Chelvarajah, R., Rollings, D.T., Sawlani, V., Hamer, H., Gollwitzer, S., Kreiselmeyer, G., Staresina, B., Wimber, M., Hanslmayr, S., 2019. Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. Proc. Natl. Acad. Sci. USA 116 (43), 21834–21842. https://doi.org/10.1073/pnas.1914180116.

Gronau, N., 2021. To grasp the world at a glance: the role of attention in visual and semantic associative processing. J. Imaging Sci. Technol. 7 (9) https://doi.org/10.3390/jimaging7090191.

Gronau, N., Neta, M., Bar, M., 2008. Integrated contextual representation for objects' identities and their locations. J. Cogn. Neurosci. 20 (3), 371–388. https://doi.org/10.1162/jocn.2008.20027.

Guo, D., Yang, J., 2022. Reactivation of schema representation in lateral occipital cortex supports successful memory encoding. *Cereb. Cortex*. https://doi.org/10.1093/cercor/bhac475.

Hasselmo, M.E., 2006. The role of acetylcholine in learning and memory. Curr. Opin. Neurobiol. 16 (6), 710–715. https://doi.org/10.1016/j.conb.2006.09.002.

Hasselmo, M.E., Eichenbaum, H., 2005. Hippocampal mechanisms for the context-dependent retrieval of episodes. Neural Netw. 18 (9), 1172–1190. https://doi.org/10.1016/j.neunet.2005.08.007.

Hasselmo, M.E., Stern, C.E., 2014. Theta rhythm and the encoding and retrieval of space and time. Neuroimage 85 (*2*), 656–666. https://doi.org/10.1016/j.neuroimage.2013.06.022.

Hasselmo, M.E., Wyble, B.P., Wallenstein, G.V., 1996. Encoding and retrieval of episodic memories: role of cholinergic and GABAergic modulation in the hippocampus. Hippocampus 6, 693–708.

Hebscher, M., Gilboa, A., 2016. A boost of confidence: the role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. Neuropsychologia 90, 46–58. https://doi.org/10.1016/j.neuropsychologia.2016.05.003.

Heusser, A.C., Ezzyat, Y., Shiff, I., Davachi, L., 2018. Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. J. Exp. Psychol. Learn. Mem. Cogn. 44 (7), 1075–1090. https://doi.org/10.1037/xlm0000503.

Hindy, N.C., Ng, F.Y., Turk-Browne, N.B., 2016. Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nat. Neurosci. 19 (5), 665–667. https://doi.org/10.1038/nn.4284.

Hindy, N.C., Avery, E.W., Turk-Browne, N.B., 2019. Hippocampal-neocortical interactions sharpen over time for predictive actions. Nat. Commun. 10 (1), 3989. https://doi.org/10.1038/s41467-019-12016-9.

Horner, A.J., Doeller, C.F., 2017. Plasticity of hippocampal memories in humans. Curr. Opin. Neurobiol. 43, 102–109. https://doi.org/10.1016/j.conb.2017.02.004.

Horner, A.J., Bisby, J.A., Bush, D., Lin, W.-J., Burgess, N., 2015. Evidence for holistic episodic recollection via hippocampal pattern completion. Nat. Commun. 6, 7462. https://doi.org/10.1038/ncomms8462.

Hunt, R.R., 1995. The subtlety of distinctiveness: what von restorff really did. Psychon. Bull. Rev. 2 (1), 105–112. https://doi.org/10.3758/bf03214414.

Kafkas, A., Montaldi, D., 2018a. Expectation affects learning and modulates memory experience at retrieval. Cognition 180, 123–134. https://doi.org/10.1016/j.cognition.2018.07.010.

Kafkas, A., Montaldi, D., 2018b. How do memory systems detect and respond to novelty. Neurosci. Lett. 680, 60–68. https://doi.org/10.1016/j.neulet.2018.01.053.

Kim, G., Norman, K.A., Turk-Browne, N.B., 2017. Neural differentiation of incorrectly predicted memories. J. Neurosci. 37 (8), 2022–2031. https://doi.org/10.1523/JNEUROSCI.3272-16.2017.

Kim, H., Schlichting, M.L., Preston, A.R., Lewis-Peacock, J.A., 2020. Predictability changes what we remember in familiar temporal contexts. J. Cogn. Neurosci. 32 (1), 124–140. https://doi.org/10.1162/jocn_a_01473.

Knierim, J.J., Neunuebel, J.P., 2016. Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. Neurobiol. Learn. Mem. 129, 38–49. https://doi.org/10.1016/j.nlm.2015.10.008.

Kok, P., Turk-Browne, N.B., 2018. Associative prediction of visual shape in the hippocampus. J. Neurosci. 38 (31), 6888–6899. https://doi.org/10.1523/JNEUROSCI.0163-18.2018.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis—Connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4. https://doi.org/10.3389/neuro.06.004.2008.

Lacy, J.W., Yassa, M.A., Stark, S.M., Muftuler, L.T., Stark, C.E.L., 2011. Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. Learning & Memory 18 (1), 15–18. https://doi.org/10.1101/lm.1971111.

Lee, C.S., Aly, M., Baldassano, C., 2021. Anticipation of temporally structured events in the brain. Elife 10. https://doi.org/10.7554/eLife.64972.

Lee, H., Samide, R., Richter, F.R., Kuhl, B.A., 2019. Decomposing parietal memory reactivation to predict consequences of remembering. Cereb. Cortex 29 (8), 3305–3318. https://doi.org/10.1093/cercor/bhy200.

Leutgeb, J.K., Leutgeb, J.K., Moser, M.-B., Moser, E.I., 2007. Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science 315 (5814), 961–966. https://doi.org/10.1126/science.1135801.

Liberman, N., Sagristano, M.D., Trope, Y., 2002. The effect of temporal distance on level of mental construal. J. Exp. Soc. Psychol. 12.

Lisman, J., Redish, A.D., 2009. Prediction, sequences and the hippocampus. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364 (1521), 1193–1201. https://doi.org/10.1098/rstb.2008.0316.

Lisman, J.E., Grace, A.A., 2005. The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 46 (5), 703–713. https://doi.org/10.1016/j.neuron.2005.05.002.

Liu, Z.-X., Grady, C., Moscovitch, M., 2017. Effects of prior-knowledge on brain activation and connectivity during associative memory encoding. Cereb. Cortex 27 (3), 1991–2009. https://doi.org/10.1093/cercor/bhw047.

Love, B.C., Medin, D.L., Gureckis, T.M., 2004. SUSTAIN: a network model of category learning. Psychol. Rev. 111 (2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309.

Maril, A., Avital, R., Reggev, N., Zuckerman, M., Sadeh, T., Ben Sira, L., Livneh, N., 2011. Event congruency and episodic encoding: a developmental fMRI study. Neuropsychologia 49 (11), 3036–3045. https://doi.org/10.1016/j.neuropsychologia.2011.07.004.

Marr, D., 1971. Simple memory: a theory for archicortex. Philos. Trans. R. Soc. Lond. Ser. B: Biol. Sci. 262 (841), 23–81.

Masís-Obando, R., Norman, K.A., Baldassano, C., 2022. Schema representations in distinct brain networks support narrative memory during encoding and retrieval. eLife 11, e70445. https://doi.org/10.7554/eLife.70445.

McClelland, J.L., 2013. Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. J. Exp. Psychol. Gen. 142 (4), 1190–1210. https://doi.org/10.1037/a0033812.

McClelland, J.L., McNaughton, B.L., O'Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. 102 (3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419.

McClelland, J.L., McNaughton, B.L., Lampinen, A.K., 2020. Integration of new information in memory: new insights from a complementary learning systems perspective. Philos. Trans. R. Soc. Lond. B Biol. Sci. 375 (1799), 20190637. https://doi.org/10.1098/rstb.2019.0637.

Meeter, M., Murre, J.M.J., Talamini, L.M., 2004. Mode shifting between storage and recall based on novelty detection in oscillating hippocampal circuits. Hippocampus 14 (6), 722–741. https://doi.org/10.1002/hipo.10214.

Michelmann, S., Price, A.R., Aubrey, B., Strauss, C.K., Doyle, W.K., Friedman, D., Dugan, P.C., Devinsky, O., Devore, S., Flinker, A., Hasson, U., Norman, K.A., 2021. Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. Nat. Commun. 12 (1), 5394. https://doi.org/10.1038/s41467-021-25376-y.

Misanin, J.R., Miller, R.R., Lewis, D.J., 1968. Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. Science 160 (3827), 554–555. https://doi.org/10.1126/science.160.3827.554.

Molitor, R.J., Sherrill, K.R., Morton, N.W., Miller, A.A., Preston, A.R., 2021. Memory reactivation during learning simultaneously promotes dentate Gyrus/CA2,3 pattern differentiation and CA1 memory integration. J. Neurosci. 41 (4), 726–738. https://doi.org/10.1523/JNEUROSCI.0394-20.2020.

Nader, K., Schafe, G.E., Le Doux, J.E., 2000. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. Nature 406 (6797), 722–726. https://doi.org/10.1038/35021052.

Nicholas, J., Daw, N.D., Shohamy, D., 2022. Uncertainty alters the balance between incremental learning and episodic memory. eLife 11, e81679. https://doi.org/10.7554/eLife.81679.

Niv, Y., 2019. Learning task-state representations. Nat. Neurosci. 22 (10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8.

Niv, Y., Daniel, R., Geana, A., Gershman, S.J., Leong, Y.C., Radulescu, A., Wilson, R.C., 2015. Reinforcement learning in multidimensional environments relies on attention mechanisms. J. Neurosci. 35 (21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015.

Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. Psychol. Rev. 110 (4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611.

Norman, K.A., Newman, E., Detre, G., Polyn, S., 2006. How inhibitory oscillations can train neural networks and punish competitors. Neural Comput. 18 (7), 1577–1610. https://doi.org/10.1162/neco.2006.18.7.1577.

O'Reilly, R.C., McClelland, J.L., 1994. Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. Hippocampus 4 (6), 661–682. https://doi.org/10.1002/hipo.450040605.

Ortiz-Tudela, J., Milliken, B., Jiménez, L., Lupiáñez, J., 2018. Attentional influences on memory formation: a tale of a not-so-simple story. Mem. Cogn. 46 (4), 544–557. https://doi.org/10.3758/s13421-017-0784-2.

Ortiz-Tudela, J., Nolden, S., Pupillo, F., Ehrlich, I., Schommartz, I., Turan, G., Shing, Y.L., 2023. Not what u expect: effects of prediction errors on item memory. J. Exp. Psychol.: Gen. https://doi.org/10.1037/xge0001367.

Paz, R., Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2010). A neural substrate in the human hippocampus for linking successive events. Proc. Natl. Acad. Sci. U. S. A., 107(13), 6046–6051. https://doi.org/10.1073/pnas.0910834107.

Piaget, J. (1952). *The origins of intelligence in children*. International Universities Press.

Popov, V., Zhang, Q., Koch, G.E., Calloway, R.C., Coutanche, M.N., 2019. Semantic knowledge influences whether novel episodic associations are represented symmetrically or asymmetrically. Mem. Cogn. 47 (8), 1567–1581. https://doi.org/10.3758/s13421-019-00950-4.

Preston, A.R., Eichenbaum, H., 2013. Interplay of hippocampus and prefrontal cortex in memory. Curr. Biol. 23 (17), R764–R773. https://doi.org/10.1016/j.cub.2013.05.041.

Quent, J.A., Henson, R.N., Greve, A., 2021. A predictive account of how novelty influences declarative memory. Neurobiol. Learn. Mem. 179, 107382 https://doi.org/10.1016/j.nlm.2021.107382.

Quent, J.A., Greve, A., Henson, R.N., 2022. Shape of U: the nonmonotonic relationship between object–location memory and expectedness. Psychol. Sci. 33 (12), 2084–2097. https://doi.org/10.1177/09567976221109134.

Ranganath, C., Rainer, G., 2003. Neural mechanisms for detecting and remembering novel events. Nat. Rev. Neurosci. 4 (3), 193–202. https://doi.org/10.1038/nrn1052.

Reagh, Z.M., Ranganath, C., 2023. Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. Nat. Commun. 14 (1), 1279. https://doi.org/10.1038/s41467-023-36805-5.

Reder, L.M., Liu, X.L., Keinath, A., Popov, V., 2016. Building knowledge requires bricks, not sand: the critical role of familiar constituents in learning. Psychon. Bull. Rev. 23 (1), 271–277. https://doi.org/10.3758/s13423-015-0889-1.

Reder, L.M., Victoria, L.W., Manelis, A., Oates, J.M., Dutcher, J.M., Bates, J.T., Cook, S., Aizenstein, H.J., Quinlan, J., Gyulai, F., 2013. Why It's easier to remember seeing a face we already know than one we don't: preexisting memory representations facilitate memory formation. Psychol. Sci. 24 (3), 363–372. https://doi.org/10.1177/0956797612457396.

Reggev, N., Bein, O., Maril, A., 2016. Distinct neural suppression and encoding effects for conceptual novelty and familiarity. J. Cogn. Neurosci. 28 (10), 1455–1470. https://doi.org/10.1162/jocn.

Rescorla, R.A., 2004. Spontaneous recovery. In: Learning & Memory, 11. Cold Spring Harbor, *N.Y*, pp. 501–509. https://doi.org/10.1101/lm.77504.

Riccio, D.C., Millin, P.M., Bogart, A.R., 2006. Reconsolidation: a brief history, a retrieval view, and some recent issues. Learn. Mem. 13 (5), 536–544. https://doi.org/10.1101/lm.290706.

Ritchey, M., Wing, E.A., LaBar, K.S., Cabeza, R., 2013. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cereb. Cortex 23 (12), 2818–2828. https://doi.org/10.1093/cercor/bhs258.

Ritvo, V.J.H., Turk-Browne, N.B., Norman, K.A., 2019. Nonmonotonic plasticity: how memory retrieval drives learning. Trends Cogn. Sci. 23 (9), 726–742. https://doi.org/10.1016/j.tics.2019.06.007.

Roediger 3rd, H.L., McDermott, K.B., 1995. Creating false memories: remembering words not presented in lists. J. Exp. Psychol.: Learn. Mem. Cogn. 21 (4), 803–814.

Rouhani, N., Niv, Y., 2021. Signed and unsigned reward prediction errors dynamically enhance learning and memory. Elife 10. https://doi.org/10.7554/eLife.61077.

Rouhani, N., Norman, K.A., Niv, Y., Bornstein, A.M., 2020. Reward prediction errors create event boundaries in memory. Cognition 203, 104269. https://doi.org/10.1016/j.cognition.2020.104269.

Rouhani, N., Niv, Y., Frank, M.J., Schwabe, L., 2023. Multiple routes to enhanced memory for emotionally relevant events. Trends Cogn. Sci. S136466132300150X https://doi.org/10.1016/j.tics.2023.06.006.

Ruiz, N.A., Thieu, M.K., Aly, M., 2021. Cholinergic modulation of hippocampally mediated attention and perception. Behav. Neurosci. 135 (1), 51–70. https://doi.org/10.1037/bne0000405.

Sakai, K., Miyashita, Y., 1991. Neural organization for the long-term memory of paired associates. Nature 354, 152–155.

Schacter, D.L., 2022. The seven sins of memory: An update. Memory 30 (1), 37–42. https://doi.org/10.1080/09658211.2021.1873391.

Schapiro, A.C., Kustner, L.V., Turk-Browne, N.B., 2012. Shaping of object representations in the human medial temporal lobe based on temporal regularities. Curr. Biol. 22 (17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056.

Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B., Botvinick, M.M., 2013. Neural representations of events arise from temporal community structure. Nat. Neurosci. 16 (4), 486–492. https://doi.org/10.1038/nn.3331.

Schiller, D., Monfils, M.-H., Raio, C.M., Johnson, D., C., LeDoux, J.E., Phelps, E.A., 2010. Preventing the return of fear in humans using reconsolidation update mechanisms. Nature 463, 49–54.

Schlichting, M.L., Mumford, J.A., Preston, A.R., 2015. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat. Commun. 6, 8151. https://doi.org/10.1038/ncomms9151.

Shin, Y.S., DuBrow, S., 2021. Structuring memory through inference-based event segmentation. Top. Cogn. Sci. 13 (1), 106–127. https://doi.org/10.1111/tops.12505.

Shohamy, D., Turk-Browne, N.B., 2013. Mechanisms for widespread hippocampal involvement in cognition. J. Exp. Psychol. Gen. 142 (4), 1159–1170. https://doi.org/10.1037/a0034461.

Silva, M., Baldassano, C., Fuentemilla, L., 2019. Rapid memory reactivation at movie event boundaries promotes episodic encoding. J. Neurosci. 39 (43), 8538–8548. https://doi.org/10.1523/JNEUROSCI.0360-19.2019.

Sinclair, A.H., Manalili, G.M., Brunec, I.K., Adcock, R.A., Barense, M.D., 2021. Prediction errors disrupt hippocampal representations and update episodic memories. Proc. Natl. Acad. Sci. USA 118 (51), e2117625118. https://doi.org/10.1073/pnas.2117625118.

Sols, I., DuBrow, S., Davachi, L., Fuentemilla, L., 2017. Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. e4 Curr. Biol. 27 (22), 3499–3504. https://doi.org/10.1016/j.cub.2017.09.057.

Sommer, T., Hennies, N., Lewis, P.A., Alink, A., 2022. The assimilation of novel information into schemata and its efficient consolidation. J. Neurosci. 42 (30), 5916–5929. https://doi.org/10.1523/JNEUROSCI.2373-21.2022.

Song, M., Jones, C.E., Monfils, M.-H., Niv, Y., 2022. Explaining the effectiveness of fear extinction through latent-cause inference. Neurons Behav. Data Anal. Theory (NBDT) 1 (1). ⟨http://arxiv.org/abs/2205.04670⟩.

Staresina, B.P., Davachi, L., 2008. Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. J. Cogn. Neurosci. 20 (8), 1478–1489. https://doi.org/10.1162/jocn.2008.20104.

Staresina, B.P., Gray, J.C., Davachi, L., 2009. Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. Cereb. Cortex 19 (5), 1198–1207. https://doi.org/10.1093/cercor/bhn165.

Staresina, B.P., Duncan, K.D., Davachi, L., 2011. Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. J. Neurosci. 31 (24), 8739–8747. https://doi.org/10.1523/JNEUROSCI.4978-10.2011.

Stark, S.M., Kirwan, C.B., Stark, C.E.L., 2019. Mnemonic similarity task: a tool for assessing hippocampal integrity. Trends Cogn. Sci. 23 (11), 938–951. https://doi.org/10.1016/j.tics.2019.08.003.

Tibon, R., Greve, A., Henson, R., 2018. The missing link? Testing a schema account of unitization. Mem. Cogn. 46 (7), 1023–1040. https://doi.org/10.3758/s13421-018-0819-3.

Tibon, R., Gronau, N., Scheuplein, A.-L., Mecklinger, A., Levy, D.A., 2014. Associative recognition processes are modulated by the semantic unitizability of memoranda. Brain Cogn. 92C, 19–31. https://doi.org/10.1016/j.bandc.2014.09.009.

Todd, T.P., Vurbic, D., Bouton, M.E., 2014. Behavioral and neurobiological mechanisms of extinction in Pavlovian and instrumental learning. Neurobiol. Learn. Mem. 108, 52–64. https://doi.org/10.1016/j.nlm.2013.08.012.

Tompary, A., Davachi, L., 2017. Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. Neuron 96 (1), 228–241. https://doi.org/10.1016/j.neuron.2017.09.005.

Tompary, A., Thompson-Schill, S.L., 2021. Semantic influences on episodic memory distortions. J. Exp. Psychol. Gen. https://doi.org/10.1037/xge0001017.

Tompary, A., Duncan, K., Davachi, L., 2016. High-resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. Hippocampus 26 (8), 995–1007. https://doi.org/10.1002/hipo.22582.

Tompary, A., Zhou, W., Davachi, L., 2020. Schematic memories develop quickly, but are not expressed unless necessary. Sci. Rep. 10 (1), 16968. https://doi.org/10.1038/s41598-020-73952-x.

Treves, A., Rolls, E.T., 1994. Computational analysis of the role of the hippocampus in memory. Hippocampus 4 (3), 374–391. https://doi.org/10.1002/hipo.450040319.

Trope, Y., Liberman, N., 2003. Temporal construal. Psychol. Rev. 110 (3), 403–421. https://doi.org/10.1037/0033-295x.110.3.403.

Trope, Y., Liberman, N., 2010. Construal-level theory of psychological distance. Psychol. Rev. 117 (2), 440–463.

Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M. P., Morris, R.G.M., 2007. Schemas and memory consolidation. Science 316 (5821), 76–82. https://doi.org/10.1126/science.1135935.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., Morris, R.G.M., 2011. Schema-dependent gene activation and memory encoding in neocortex. Science 333 (6044), 891–895. https://doi.org/10.1126/science.1205274.

Turan, G., Ehrlich, I., Shing, Y.L., & Nolden, S. (2023). From generating to violating predictions: The effects of prediction error on episodic memory [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/zm29a.

van Buuren, M., Kroes, M.C.W., Wagner, I.C., Genzel, L., Morris, R.G.M., Fernández, G., 2014. Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. J. Neurosci. 34 (50), 16662–16670. https://doi.org/10.1523/JNEUROSCI.2365-14.2014.

van der Linden, M., Berkers, R.M.W.J., Morris, R.G.M., Fernández, G., 2017. Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. J. Neurosci. 37 (39), 9474–9485. https://doi.org/10.1523/JNEUROSCI.3603-16.2017.

van Kesteren, M.T.R., Rijpkema, M., Ruiter, D.J., Fernández, G., 2010. Retrieval of associative information congruent with prior knowledge is related to increased

medial prefrontal activity and connectivity. J. Neurosci. 30 (47), 15888–15894. https://doi.org/10.1523/JNEUROSCI.2674-10.2010.

van Kesteren, M.T.R., Fernández, G., Norris, D.G., & Hermans, E.J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. Proc. Natl. Acad. Sci. U. S. A., 107(16), 7550–7555. https://doi.org/10.1073/pnas.0914892107.

van Kesteren, M.T.R., Ruiter, D.J., Fernández, G., Henson, R.N., 2012. How schema and novelty augment memory formation. Trends Neurosci. 35 (4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001.

van Kesteren, M.T.R., Rijpkema, M., Ruiter, D.J., Fernández, G., 2013. Consolidation differentially modulates schema effects on memory for items and associations. PLoS One 8 (2), e56155. https://doi.org/10.1371/journal.pone.0056155.

van Kesteren, M.T.R., Rignanese, P., Gianferrara, P.G., Krabbendam, L., Meeter, M., 2020. Congruency and reactivation aid memory integration through reinstatement of prior knowledge. Sci. Rep. 10 (1), 4776. https://doi.org/10.1038/s41598-020-61737-1.

van Kesteren, M.T.R., Beul, S.F., Takashima, A., Henson, R.N., Ruiter, D.J., Fernández, G., 2013. Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. Neuropsychologia 51 (12), 2352–2359. https://doi.org/10.1016/j.neuropsychologia.2013.05.027.

von Restorff, H., 1933. Über die Wirkung von Bereichsbildungen im Spurenfeld. Psychol. Forsch. 18, 299–342.

Wahlheim, C.N., Zacks, J.M., 2019. Memory guides the processing of event changes for older and younger adults. J. Exp. Psychol. Gen. 148 (1), 30–50. https://doi.org/10.1037/xge0000458.

Wahlheim, C.N., Eisenberg, M.L., Stawarczyk, D., Zacks, J.M., 2022. Understanding everyday events: predictive-looking errors drive memory updating. Psychol. Sci. 33 (5), 765–781.

Wammes, J., Norman, K.A., Turk-Browne, N., 2022. Increasing stimulus similarity drives nonmonotonic representational change in hippocampus. eLife 11, e68344. https://doi.org/10.7554/eLife.68344.

Wang, S.-H., Tse, D., Morris, R.G.M., 2012. Anterior cingulate cortex in schema assimilation and expression. Learn. Mem. 19 (8), 315–318. https://doi.org/10.1101/lm.026336.112.

Wanjia, G., Favila, S.E., Kim, G., Molitor, R.J., Kuhl, B.A., 2021. Abrupt hippocampal remapping signals resolution of memory interference. Nat. Commun. 12 (1), 1–11. https://doi.org/10.1038/s41467-021-25126-0.

Wing, E.A., Burles, F., Ryan, J.D., Gilboa, A., 2022. The structure of prior knowledge enhances memory in experts by reducing interference. Proc. Natl. Acad. Sci. 119 (26), e2204172119 https://doi.org/10.1073/pnas.2204172119.

Wu, X., Packard, P.A., García-Arch, J., Bunzeck, N., Fuentemilla, L., 2023. Contextual incongruency triggers memory reinstatement and the disruption of neural stability. NeuroImage 273, 120114. https://doi.org/10.1016/j.neuroimage.2023.120114.

Wynn, J.S., Ryan, J.D., Moscovitch, M., 2020. Effects of prior knowledge on active vision and memory in younger and older adults. J. Exp. Psychol.: Gen. 149 (3), 518–529. https://doi.org/10.1037/xge0000657.

Xue, G., 2018. The neural representations underlying human episodic memory. Trends Cogn. Sci. 22 (6), 544–561. https://doi.org/10.1016/j.tics.2018.03.004.

Yacoby, A., Reggev, N., Maril, A., 2021. Examining the transition of novel information toward familiarity. Neuropsychologia 161, 107993. https://doi.org/10.1016/j.neuropsychologia.2021.107993.

Yacoby, A., Reggev, N., Maril, A., 2023. Lack of source memory as a potential marker of early assimilation of novel items into current knowledge. Neuropsychologia 185, 108569. https://doi.org/10.1016/j.neuropsychologia.2023.108569.

Yassa, M.A., Stark, C.E.L., 2011. Pattern separation in the hippocampus. Trends Neurosci. 34 (10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006.

Ye, Z., Zhu, B., Zhuang, L., Lu, Z., Chen, C., Xue, G., 2016. Neural global pattern similarity underlies true and false memories. J. Neurosci. 36 (25), 6792–6802. https://doi.org/10.1523/JNEUROSCI.0425-16.2016.

Yousuf, M., Packard, P.A., Fuentemilla, L., Bunzeck, N., 2021. Functional coupling between CA3 and laterobasal amygdala supports schema dependent memory formation. Neuroimage, 118563. https://doi.org/10.1016/j.neuroimage.2021.118563.

Yousuf, M., Packard, P.A., Fuentemilla, L., Bunzeck, N., 2021. Functional coupling between CA3 and laterobasal amygdala supports schema dependent memory formation. Neuroimage, 118563, 10.1016/j.neuroimage.2021.118563.

Yu, L.Q., Wilson, R.C., Nassar, M.R., 2021. Adaptive learning is structure learning in time. Neurosci. Biobehav. Rev. 128, 270–281. https://doi.org/10.1016/j.neubiorev.2021.06.024.

Zeithamova, D., Preston, A.R., 2017. Temporal proximity promotes integration of overlapping events. J. Cogn. Neurosci. 29 (8), 1311–1323. https://doi.org/10.1162/jocn_a_01116.

Zhu, B., Chen, C., Shao, X., Liu, W., Ye, Z., Zhuang, L., Zheng, L., Loftus, E.F., Xue, G., 2019. Multiple interactive memory representations underlie the induction of false memory. Proc. Natl. Acad. Sci. USA 116 (9), 3466–3475. https://doi.org/10.1073/pnas.1817925116.