Published in final edited form as:
Neuron. 2008 October 23; 60(2): 378–389. doi:10.1016/j.neuron.2008.09.023.

Integrating Memories in the Human Brain: Hippocampal–Midbrain Encoding of Overlapping Events

Daphna Shohamy[1,3] and **Anthony D. Wagner**[1,2]

1 *Department of Psychology, Stanford University, Jordan Hall Bldg. 420, Stanford, CA 94305-2130*

2 *Neuroscience Program, Stanford University, Jordan Hall Bldg. 420, Stanford, CA 94305-2130*

3 *Department of Psychology, Columbia University, Schermerhorn Hall, NY, NY, 10027*

SUMMARY

Decisions are often guided by generalizing from past experiences. Fundamental questions remain regarding the cognitive and neural mechanisms by which generalization takes place. Prior data suggest that generalization may stem from inference-based processes that occur at the time of generalization. By contrast, it has been hypothesized that generalization may emerge from mnemonic processes that occur while premise events are being encoded. Here, participants engaged in a two-phase learning and generalization task, wherein they initially learned a series of overlapping associations, and were subsequently probed to generalize what they learned to novel stimulus combinations. Functional magnetic resonance imaging (fMRI) revealed that subsequent generalization performance was associated with coupled changes in learning-phase activity in the hippocampus and midbrain (ventral tegmental area/substantia nigra). These findings provide novel evidence for generalization based on integrative encoding, whereby overlapping past events are integrated into a linked mnemonic representation. Hippocampal–midbrain interactions support the dynamic integration of experiences, providing a powerful mechanism for building a rich associative history that extends beyond individually experienced events.

Memory is essential to behavior, enabling organisms to draw on past experience to guide choices and actions. Extensive evidence suggests that the hippocampus encodes experiences (events) into long term memory as separated, discrete representations (Kirwan and Stark, 2007; Leutgeb et al., 2007; McNaughton and Nadel, 1989; Norman and O'Reilly, 2003; O'Reilly and Rudy, 2001). Such discrete encoding provides a mechanism for remembering specific details of single events. However, experiences often overlap in their content, presenting opportunities for generalizing across them. It has been proposed that effective generalization may depend on integrating discrete experiences into a rich, cohesive representation (Eichenbaum, 2000; Gluck and Myers, 1993). A fundamental question concerns whether, and how, such integration takes place.

One approach to examining generalization is to train an organism on separate events that share common elements (e.g., A–B and B–C) and then test whether the organism demonstrates knowledge about the relation between the elements that were not directly experienced together (e.g., A and C) (Dusek and Eichenbaum, 1997; Eichenbaum, 2000; Greene et al., 2006; Heckers

Correspondence should be addressed to: Daphna Shohamy, Department of Psychology, Columbia University, Schermerhorn Hall, NY, NY 10027, shohamy@psych.columbia.edu, Tel: 212-854-7560.

et al., 2004; Preston et al., 2004). Extant data from this type of paradigm indicate that animals and humans generalize, and that this ability depends on the hippocampus (Dusek and Eichenbaum, 1997; Eichenbaum, 2000; Greene et al., 2006; Heckers et al., 2004; Preston et al., 2004). Specification of the hippocampal mechanisms that enable such generalization is central to understanding how decisions are guided by past experience.

One possible mechanism by which knowledge may be generalized across discrete experiences is through logical inference at test (e.g. Dusek and Eichenbaum, 1997; Greene et al., 2006). Indeed, generalization is often thought to depend on transitive or associative *inference* (Greene et al., 2006; Heckers et al., 2004; Preston et al., 2004). On this view, the hippocampus contributes to generalization by supporting the novel and flexible expression of memories on which inferences rest (Cohen and Eichenbaum, 1993; Eichenbaum, 2000; Preston et al., 2004). That is, the hippocampus stores and enables flexible retrieval of discrete memories that afford an inference about the relation between multiple elements, even when this relation is not directly encoded in memory. Consistent with this view, functional imaging studies in humans have demonstrated greater hippocampal activation when subjects make memory judgments about pairs of items (e.g., A–C) whose relationship was mediated through an element (i.e., B) common to two separate associations (i.e., A–B and B–C) (Greene et al., 2006; Heckers et al., 2004; Preston et al., 2004).

While prior observations provide evidence for inference-based generalization, here we report novel evidence for an alternative mechanism—*integrative encoding*—through which generalization can emerge. By this view, the hippocampus contributes to the integration of distinct episodes into a linked network of mnemonic associations. The dynamic construction of integrated memories—posited to occur as overlapping episodes are experienced—enables direct retrieval of knowledge about the relation between multiple elements that, while not directly experienced, are nevertheless encoded in memory (Eichenbaum, 2000). Thus, at test, generalization is not a reconstructive inference-based process based on flexible retrieval of multiple memories, but rather is a direct expression of knowledge encoded in memory as a synthesis of information across multiple experiences.

Motivated by computational theories and neurophysiological data that suggest that the hippocampus dynamically shifts between encoding and retrieval states (Hasselmo and McClelland, 1999; Hasselmo et al., 1995), we hypothesized that encountering an event that has feature overlap with a previously encoded event can trigger retrieval of memory for the past event, and that this, in turn, can lead to encoding of the two discrete events into an integrated representation. Such integrative encoding would allow direct storage in memory of the relation between two elements that were not experienced together. According to this perspective, subsequent responses to generalization probes can be based on the direct retrieval of a stored *integrative* memory, such that generalization is essentially the same as retrieval of a previously experienced event, rather than a slower, more effortful process of inference. Notably, this notion of alternating encoding and retrieval as a mechanism for integrating memories was proposed as early as 1923 by Richard Semon (Schacter, 2001a). However, little is known about whether and how such mnemonic integration occurs.

Neurochemical modulation is thought to be essential in driving dynamic shifts between encoding and retrieval in the hippocampus (Hasselmo et al., 1995). While previous work emphasizes cholinergic modulation from the basal forebrain, here we propose that the midbrain dopamine system plays a key role in modulating hippocampal mechanisms necessary for cross-episode integration. This perspective is guided by three observations. First, the hippocampus is innervated by dopamine projections from the ventral tegmental area (VTA) in the midbrain (Gasbarri et al., 1994; Swanson, 1982). Second, dopamine release in the hippocampus modulates hippocampal plasticity (Frey et al., 1990; Morris et al., 2003; Otmakhova and

Lisman, 1996); indeed, both midbrain activation (Wittmann et al., 2005) and interaction between hippocampus and midbrain (Adcock et al., 2006) have been shown to facilitate encoding of individual episodes. Third, midbrain dopamine neurons are most responsive under circumstances in which predictions are violated (Lisman and Grace, 2005; Schultz et al., 1997); we argue that prediction violation is precisely what happens when an organism encounters an episode that contains elements that overlap with a previously encoded episode (Kumaran and Maguire, 2006). That is, when encountering an overlapping element, this overlap leads to retrieval of prior episodic details that mismatch the details of the present event. Collectively, these observations led us to predict that integrative encoding across experiences is supported by a cooperative interaction between the hippocampus and midbrain dopamine regions.

Notably, when considered from the perspective of the putative hippocampal–VTA loop (Lisman and Grace, 2005), this functional prediction about the relationship between midbrain activation and learning is distinct from that of the established view of midbrain dopamine neurons in modulating cortico-striatal “habit” learning. This latter view rests on extensive findings demonstrating that midbrain dopamine neurons respond when reward predictions are violated (Schultz, 1998). Specifically, a large body of research indicates that reinforcement learning depends on midbrain dopamine neurons and their striatal (caudate and putamen) targets (Aron et al., 2004; Daw and Doya, 2006; Delgado et al., 2005; Delgado et al., 2000; Faure et al., 2005; Frank et al., 2004; Schultz et al., 1997; Shohamy et al., 2006; Shohamy et al., 2004). This type of learning is thought to be independent of the hippocampal declarative memory system (Gabrieli, 1998; Knowlton et al., 1996; Myers et al., 2003; Shohamy et al., 2006; Shohamy et al., 2008; Yin and Knowlton, 2006), and perhaps even antagonistic to it (Packard and McGaugh, 1996; Poldrack et al., 2001; Poldrack and Packard, 2003).

From the perspective of current theories of reinforcement learning, midbrain dopamine regions would not be expected to contribute to cross-event generalization and midbrain activation would not be expected to couple with that in the hippocampus. However, if, as we hypothesize, midbrain dopamine regions interact with the hippocampus to support generalization, then midbrain activation is predicted to couple with that in the hippocampus and to correlate with generalization performance. Such outcomes would add to a growing body of evidence indicating a broader role for this system in learning and memory by modulating the hippocampus, in addition to the striatum (Adcock et al., 2006; Lisman and Grace, 2005; Wittmann et al., 2005).

To summarize, we hypothesized that (a) generalization stems from integrative encoding that occurs while experiencing events that partially overlap with previously encoded events, (b) that such integrative encoding depends on both the hippocampus and midbrain dopamine regions, and (c) that greater hippocampal–midbrain engagement during integrative encoding enables rapid behavioral generalization in the future.

To test these predictions, 24 participants were scanned with functional magnetic resonance imaging (fMRI) while engaged in an associative learning and generalization task (Fig. 1) (Collie et al., 2002;Grice and Davis, 1960;Hall et al., 1993;Myers et al., 2003). The design was conceptually similar to existing transitive and associative inference paradigms (e.g., Greene et al., 2006;Heckers et al., 2004;Preston et al., 2004), with several noteworthy differences. First, in the present study, rather than being blocked (Greene et al., 2006;Heckers et al., 2004;Preston et al., 2004), here the trained associations were intermixed, much as occurs in everyday life where overlapping experiences are temporally intermixed. By intermixing the presentation of associations with overlapping content multiple times we sought to allow more opportunity for integration during learning. Second, while transitive inference paradigms typically entail hierarchically organized stimuli (e.g., A>B, B>C, C>D, D>E; generalize to B?D probe), which

may encourage the articulation of a logical, rational structure of elements at the time of test, here we used a generalization paradigm that involved arbitrary, overlapping associations between elements, which is more characteristic of the overlap in elements that are encountered in daily episodic experiences.

During the initial phase of the task, participants learned a series of face-scene associations that were structured to include partial overlap across associative pairs, providing an opportunity for integrative encoding. On each trial, participants learned to associate a face with a scene by choosing which of two scenes went with the face and receiving feedback (Fig. 1A). While each face-scene association was learned individually, there was partial overlap across events, such that pairs of faces were associated with a common scene (e.g., F_1–S_1; F_2–S_1). In addition to learning the F_1–S_1 and F_2–S_1 associations, participants were concurrently trained on a second association for one of the faces (i.e., F_1–S_2) (Fig. 1A). Thus, the initial learning phase consisted of three different types of stimulus combinations that contained partial overlap (F_1–S_1; F_2–S_1; F_1–S_2). To the extent that the overlap between F_1–S_1 and F_2–S_1 elicits cross-event integration during learning, we expected that the additional learning of the F_1–S_2 association would lead F_2 to also become associated with S_2 (Collie et al., 2002;Grice and Davis, 1960;Hall et al., 1993;Myers et al., 2003).

Over the course of learning, participants were trained with 24 face-scene associations that followed this structure (F_1 through F_{24}; S_1 through S_{24}). Trials were intermixed, each repeated eight times, and distributed across two encoding scans—an initial encoding scan provided an opportunity to learn the presented associations (early learning), followed by a second encoding scan that provided additional opportunities to strengthen these associations and to integrate across them (late learning).

Following the two encoding scans, a test phase probed participants' ability to generalize. Specifically, generalization trials tested whether participants would choose S_2 when presented F_2 even though they had never encountered this pairing at study (Fig. 1B). These generalization trials were tested together with trials that probed retention of knowledge about the associations that had been previously encountered (F_2–S_1; F_1–S_1; F_1–S_2; 'trained'). To maximize power for imaging analyses, each test-phase trial was repeated six times. Feedback was not provided during this phase, to ensure that no new learning occurred across test trials.

RESULTS

Behavioral Performance

All participants were able to learn and retain the trained pairings. During encoding, most of the learning of the trained pairings occurred early in learning, during the initial encoding scan (with accuracy improving from 57% correct to 80%); late in learning, during the second encoding scan, accuracy further incrementally improved (from 86% to 93%; see Supplemental Results). At test, participants remained highly accurate on the trained pairings (93% correct).

Mean performance on the generalization test probes was high (81%), and consistent across the six repeated test presentations (see Supplemental Results). Interestingly, generalization markedly varied across individuals (range 38–100%), indicating that, on average, participants were able to exploit the overlap in encountered associations, but that they differed in their ability to do so. The key question is: What representations and processes support successful generalization (Fig. 1C)? Is generalization based on logical inference at test, during which retrieval of the individually encoded associations is used to infer that F_2 goes with S_2? Or, does integrative encoding of overlapping events take place during learning via hippocampal–midbrain interactions, such that the untrained generalization associations (i.e., F_2–S_2) are encoded in memory and then retrieved at test, as occurs for trained associations (e.g., F_1–S_2)?

As described below, our fMRI results suggest that successful generalization is driven by cross-episode integration during learning, and not by memory-based inference at test.

Generalization performance is not related to activation at test

Given the literature on transitive and associative inference (Greene et al., 2006; Heckers et al., 2004; Preston et al., 2004), we first examined whether test-phase activation differed on generalization vs. trained trials. Specifically, to the extent that logical inference at test supports generalization, this predicts (a) greater hippocampal activation at *test* during correct generalization relative to correct trained trials (Greene et al., 2006; Heckers et al., 2004; Preston et al., 2004), and (b) that the magnitude of hippocampal activation on generalization trials would correlate with generalization performance. Analyses of the test phase data failed to support either of these predictions, providing no evidence that hippocampal retrieval mechanisms were differentially engaged on generalization vs. trained probes (see Supplemental Results). As such, these data are inconsistent with generalization based on logical inference mechanism.

Activation in hippocampus and midbrain during learning predicts generalization

To assess whether integrative encoding occurs and supports generalization, we examined the relationship between hippocampal and midbrain activation during learning and participants' subsequent generalization performance. Because integration across associations depends on having learned the individual associations, we predicted that integration would occur later, rather than earlier, in learning, and therefore that the ability to generalize at test would be associated with increasing activation over the course of learning in the hippocampus and midbrain. Importantly, this novel prediction stands counter to extensive prior evidence indicating that when encoding individual, non-overlapping associations, hippocampal activation markedly declines across learning exposures (Kohler et al., 2005; Zeineh et al., 2003). Moreover, this prediction may also stand counter to expectations about midbrain encoding activation that derive from the reinforcement learning literature. Specifically, prior evidence indicates that striatal activity declines across learning when acquiring individual, non-overlapping associations in a reinforcement learning context (Delgado et al., 2005). Thus, while it is not known how midbrain activation changes as a function of declarative memory encoding over time, to the extent that midbrain activation tracks striatal activation then one might expect a decline in midbrain activation over the course of learning, which is the opposite of our prediction.

As each generalization trial relates to a series of learning-phase events, our encoding-phase analysis did not use the "subsequent memory" approach (e.g., Paller and Wagner, 2002). Rather, to test the prediction that integrative encoding across learning supports generalization, we conducted regression analyses to determine whether the increase in magnitude of activation from early to late learning correlated with subsequent accuracy on the generalization probes.

Consistent with the integrative encoding hypothesis, this regression analysis revealed that the magnitude of activation increase from early to late learning in bilateral hippocampus and in the midbrain (VTA/SN complex) correlated with subsequent accuracy on the generalization trials at test (Ps<0.05, corrected; Fig. 2A–B); this was the case when analyzing the midbrain data using either a 4- or 8-mm smoothing filter (Fig. 3; complete results from analyses of the data using a 4-mm smoothing filter appear in the Supplemental Results). That is, superior subsequent generalization was preceded by a greater increase in hippocampal and midbrain activation from early to late learning. This relationship between subsequent generalization performance and increasing hippocampal and midbrain activation across learning was also apparent when we median split the 24 participants into 'good' and 'poor' generalization groups (96% correct vs. 66% correct, respectively) (Fig. 2C and Fig. 4A). Specifically, Group ×

Learning Phase interactions revealed a difference in the pattern of hippocampal (right hippocampus, $F(1,22)=4.31$, $P<0.05$; left hippocampus, $F(1,22)=3.25$, $P=0.08$) and midbrain ($F(1,22)=4.70$, $P<0.05$) activity across learning in the two groups, with the 'good' generalization group (Ps<0.001), but not the 'poor' generalization group (Ps>0.60), demonstrating a significant increase in activity from early to late learning (Fig. 2C).

Importantly, these subgroup differences in generalization and in learning-phase hippocampal and midbrain activity were present even when equating for differences in performance on trained associations (Fig. 4A and Supplemental Results). Moreover, multiple subsidiary analyses revealed that the correlation between learning-phase hippocampal and midbrain activity with subsequent generalization was not due to differences in retention of trained associations (see Supplemental Results). We also found no evidence for any subjective differences in the learning experience between those participants that generalized well vs. those that did not, based on self-report questionnaires administered after the study. Indeed, although self-reports should be interpreted with caution, it is interesting to note that when asked, only two participants—both poor generalizers—reported any awareness of the appearance of novel pairings during the test phase. Finally, and importantly, we note that the correlations between subsequent generalization performance and the learning-phase increase in hippocampal and midbrain activation were not unduly driven by those participants who showed maximal generalization performance (>95% correct generalization), as these correlations remained significant even when excluding the best generalizers (n=16; left hippocampus, r=0.60; right hippocampus, r=.68; midbrain, r=0.63; Ps<0.05, corrected).

Response latencies support the integrative encoding hypothesis

An additional key prediction of the integrative encoding account is that, at test, performance on the generalization and trained probes involve the same mechanism—retrieval of an encoded association—even though the latter had been repeatedly encountered and retrieved during study whereas the former had never been experienced. Accordingly, this predicts that (a) response latencies to generalization test probes would be in the same range as the latencies to trained probes, and that (b) the degree of similarity between response latencies to generalization and trained probes would relate to generalization performance.

Consistent with these predictions, regression analyses revealed that the difference in response latencies between correct trained and correct generalization trials at test was negatively correlated with generalization performance ($r=-0.69$, $P<0.001$). Similarly, when median splitting the participants, a Group _ Test Trial Type interaction ($F(1,22)=31.32$ $P<0.0001$) revealed that the 'poor' generalization group showed significantly more slowing on the generalization vs. trained trials relative to the 'good' generalization group (Fig. 4B). This effect was significant even when restricting the test phase analysis to the first encounter with the generalization probes ($F(1,22)=4.61$, $P<0.05$). In fact, among the six participants demonstrating the best generalization performance (the top quartile), there was only a 39 ms difference in mean response latency on generalization vs. trained probes—clearly insufficient time to permit mediated retrieval and logical inference—with half of these participants being faster on generalization trials.

Functional interaction between hippocampus and midbrain

Given the hypothesis that midbrain dopaminergic modulation of the hippocampus is central for integrative encoding, we examined whether there was a functional interaction between these regions during learning. To test this hypothesis, we extracted the learning-phase change in activity from the midbrain region that correlated with generalization (Fig. 2 and **Methods**), and regressed this functionally relevant midbrain response against activity elsewhere to determine whether any hippocampal voxels showed changes in activity that correlated with

that in midbrain. Consistent with our hypothesis, the learning-phase increase in midbrain activity was strongly correlated with that in the hippocampus (Fig. 5; *P*s<0.001 in left and right hippocampus), suggesting a cooperative interaction between these two regions during integrative encoding. Importantly, this cooperative interaction with the midbrain region was selective to the hippocampus (see Supplemental Results), and it remained significant even when excluding the two participants demonstrating the strongest and weakest change in VTA/SN learning-phase activity (Fig. 5).

Hippocampal and midbrain contributions to learning different event types

Finally, we asked whether specific types of overlapping events differentially elicit hippocampal and midbrain activation, by examining how learning-phase activation to the three different event types (F_1–S_1; F_1–S_2; F_2–S_1) differed as a function of generalization. Regression analyses revealed that generalization was more tightly—but not selectively—related to learning-phase hippocampal and midbrain activation changes to the F_2–S_1 trials (Fig. 6); this effect was also apparent when comparing the two generalization subgroups, especially within midbrain (Fig. 6). Importantly, this effect is merely suggestive and should be interpreted with caution, as there was no significant interaction between trial types in either the hippocampus or the midbrain (see Supplemental Results). Nonetheless, generalization might be more tightly associated with changes in learning-phase activity on F_2–S_1 events because these events are uniquely expected to evoke retrieval of a chain of two previously encoded events. Specifically, F_2–S_1 trials may lead to retrieval of F_1–S_1 trials (due to the overlap of S_1), which then may evoke retrieval of F_1–S_2 (due to the overlap of F_1). Such retrieval would enable the encoding of these multiple associations as an integrated representation (Fig. 1C).

DISCUSSION

The present results provide novel evidence for an integrative encoding mechanism in which hippocampal–midbrain interactions give rise to learning that bridges across multiple separate events. This mechanism enables rapid generalization that is based on direct retrieval of an encoded association, rather than on an inference-based process. According to this view, many instances of “generalization” may in fact be direct expressions of stored, integrated representations. To the extent that organisms can bridge across multiple integrated representations, this provides a powerful mechanism for building a rich associative history that extends beyond individually experienced events.

Our findings advance understanding of hippocampal and midbrain function in several important ways. First, we demonstrate that the hippocampus may contribute not only to the encoding of individual experiences as separated, discrete representations, but may also contribute to the integration of memories of overlapping events. This observation suggests a possible mechanism for how the hippocampus may create a continuous link across episodes that are experienced individually and at distinct moments in time.

Second, our data reveal correlated activity between midbrain dopamine regions and the hippocampus during learning, which points to a functional role of midbrain regions in modulating hippocampal-dependent cross-event integration. This novel finding may have important implications for understanding the role of midbrain dopamine regions in memory, by providing a link between theories of dopamine in expectation and prediction (e.g., Bayer and Glimcher, 2005; Schultz et al., 1997) and theories of hippocampal contributions to declarative memory (Eichenbaum and Cohen, 2001; Greene et al., 2006; Kumaran and Maguire, 2006; O’Reilly and Rudy, 2001; Squire, 1992).

Extensive data indicate a critical role for midbrain dopamine neurons in reward prediction and learning (Aron et al., 2004; Daw and Doya, 2006; Delgado et al., 2005; Delgado et al., 2000;

Faure et al., 2005; Frank et al., 2004; Schultz, 1998; Schultz et al., 1997; Shohamy et al., 2006; Shohamy et al., 2004). Such studies demonstrate that midbrain dopamine neurons signal a reward-prediction error—increasing firing when an unexpected reward occurs, and decreasing firing when an expected reward fails to occur (Schultz, 1998). More recently, it has been suggested that midbrain dopamine neurons may also play a critical role in modulating hippocampal-dependent episodic memory (Adcock et al., 2006; Lisman and Grace, 2005; Schott et al., 2005; Wittmann et al., 2005). Midbrain dopamine neurons project not only to the striatum, but also to the hippocampus (Gasbarri et al., 1994; Swanson, 1982), where dopamine has been shown to modulate plasticity (Frey et al., 1990; Morris et al., 2003; Otmakhova and Lisman, 1996).

The precise function of dopamine modulation in the hippocampus remains unknown. However, it has been proposed that a functional loop between the midbrain (VTA) and the hippocampus serves to enhance episodic memory for novel events (Lisman and Grace, 2005). This model builds upon the established role of the hippocampus in novelty detection—a process that is thought to involve comparing present events with memory representations of events in the past, and detecting any mismatch between them (Kohler et al., 2005; Strange and Dolan, 2001; Yamaguchi et al., 2004). When novelty is detected by the hippocampus, a signal is thought to project to the VTA, leading to dopamine release and memory enhancement in the hippocampus (Lisman and Grace, 2005).

The present data support this view by demonstrating a cooperative relationship between the midbrain and the hippocampus, with both regions showing a correlated increase in activation over the course of learning, even as errors (and activation in the striatum) decrease (see Supplemental Results for striatal findings). Importantly, the present data extend this view in several ways, indicating a broader role for this loop in learning and memory.

First, our data suggest that a hippocampal–midbrain network may provide a mechanism not only for the enhancement of long-term memory for individual episodes, but also for cross-episode integration. We propose that the underlying mechanism for integrative encoding may be the detection of mismatch when an organism encounters an episode that has partial overlap with a previously experienced event. For example, when encountering an event (e.g., F_2–S_1) that overlaps with a prior event (e.g., F_1–S_1), the presentation of the overlapping element (S_1) may elicit retrieval of the prior event's features (e.g., F_1). This reactivation of features from a prior event that differ from the features of the present event (i.e., F_1 vs. F_2) may trigger a mismatch signal within the hippocampus that upregulates midbrain dopaminergic feedback onto the hippocampus (Lisman and Grace, 2005), the consequence of which is to increase the probability of encoding the present and prior event features into an integrated representation. By this view, midbrain regions are argued to respond not only to violations of expectation about the value of a predicted outcome (a reward prediction error), but also to violations of expectation about the content of an episode (an episodic prediction error). As such, our data extend prior theories focusing on a role for midbrain dopamine in reward prediction (O'Doherty et al., 2003; Schultz, 1998; Schultz et al., 1997) and stimulus-response learning (Faure et al., 2005; Graybiel, 1995; Shohamy et al., 2006; Yin et al., 2004), and suggest a critical contribution of midbrain mechanisms to other forms of learning.

Second, our findings suggest that it is not item (stimulus) novelty per se that drives this hippocampal–midbrain interaction, but *associative* novelty. That is, the present data demonstrate that the hippocampus and midbrain do not show enhanced activation to novel relative to familiar items, but rather to novel stimulus combinations (see Supplemental Results), complementing prior studies that demonstrate that the hippocampus responds preferentially to associative, rather than item, novelty (Kohler et al., 2005).

Midbrain activations have also been reported to respond to novelty that is either associative (Schott et al., 2004) or item-based (Bunzeck and Duzel, 2006). More recent data suggest that the striatum is preferentially sensitive to item-based perceptual novelty, and may support a mechanism for novelty-based choice (Wittmann et al., 2008). Here, we demonstrate that midbrain activation is correlated with the hippocampus, but not the striatum, and that this correlation is directly related to subsequent successful generalization.

Our results additionally suggest that the neural and cognitive processes supporting generalization are at least partially independent of those supporting learning and retention of the trained associations. Even when controlling for differences in learning and retention of the trained associations among the 'good' and 'poor' generalizers, these groups differed markedly in generalization performance. At the neural level, these groups also differed in their pattern of hippocampal and midbrain activation across learning, even when factoring out variance associated with performance on the trained associations (see Supplemental Results). This dissociation complements prior patient studies that demonstrate that generalization—but not feedback-based learning—is impaired in individuals with damage to the hippocampus, whereas feedback-based learning—but not generalization—is impaired in individuals with disrupted striatal function due to Parkinson's disease (Myers et al., 2003; Shohamy et al., 2006). Here, feedback-based learning indeed involved the striatum (bilateral caudate; see Supplemental Results), but we found no relation between striatal activity and generalization.

Previous fMRI studies of generalization using transitive and associative inference paradigms revealed that hippocampal-dependent inference processes at test support generalization (e.g., Heckers et al., 2004; Preston et al., 2004). Our study differed in several ways. First, by using intermixed episodes, the present design was more similar to the kind of intermixed overlap in elements that one would experience in everyday life. Second, the intermixed repetitions of each encountered association in the present design may have fostered cross-event integration as these associations were being learned. Indeed, prior studies of transitive inference have found that inference-based judgments at test only occur if the training follows a block design that "frontloads" the non-overlapping pairs prior to introducing those with overlap (Titone et al., 2004). Thus, the present results are not contradictory to the findings of inference-based generalization reported in previous studies. Rather, they suggest an alternative mechanism for generalization that may complement inference-based processes. According to this view, generalization may derive from inference-based processes at test or integrative encoding during learning; the nature of the learning experience is likely to be an important factor in determining the relative contributions of each of these mechanisms to generalization.

It is also worth noting that our paradigm further differed from transitive inference designs in the nature of the overlap between the episodes. Specifically, rather than building on a hierarchical structure between elements, the present paradigm used associative overlap, or equivalencies, between partial elements of an episode. It seems plausible that the kind of hierarchical organization typically used in transitive inference paradigms may lend itself more to logical inferential processes. By contrast, converging evidence suggests that associative equivalencies between elements tend to lead to generalization without explicit awareness or recognition of the relationship between the elements (see also (Daw and Shohamy, In Press; Greene et al., 2006; Walther, 2002).

The present form of generalization may be thought of as a type of false memory (Schacter, 2001b), in that participants have the subjective sense of having already experienced the pairing of two elements that in fact had never been encountered together. Indeed, it has been suggested that false memories may emerge through associative experiences during encoding, similar to the integrative encoding mechanism proposed here. On this view, 'false memories' may in fact be associatively generated during encoding, despite not being encountered directly; at test, the

generated association would then be misattributed as an external (experienced) event rather than an internally generated experience (Underwood, 1965). This interpretation is consistent with evidence that encoding mechanisms can predict later false memory under some circumstances (Dennis et al., 2007; Garoff et al., 2005).

In summary, the present data demonstrate that interactions between the hippocampus and midbrain support a mechanism by which organisms can integrate across discrete, but overlapping experiences. By forming a thread that connects otherwise separate experiences, integrative encoding permits organisms to generalize across multiple past experiences to guide choices in the present.

METHODS

Participants

Data are reported from 24 healthy adults (13 females; ages 18–24 yrs); all were right handed, native English speakers. Test phase imaging data were lost from one participant due to corrupted files (learning phase behavioral and imaging data, and test phase behavioral data are reported for this participant). In addition, data were collected but excluded from three additional participants due to their failure to show any evidence of learning (never exceeding chance levels of responding throughout training). All participants received $20/hr for participation, with the experiment lasting approximately 2 hrs. Informed written consent was obtained from all participants in accordance with procedures approved by the institutional review board at Stanford University.

Procedure

The learning and generalization task was a modification of the ‘acquired equivalence’ paradigm (Collie et al., 2002; Grice and Davis, 1960; Hall et al., 1993; Myers et al., 2003). The critical stages consisted of a Learning phase, during which participants used feedback to learn face-scene associations, followed by a Test phase, without any feedback, during which participants were tested on previously learned associations (‘trained’) and on generalization probes (‘generalization’) consisting of novel combinations of face-scene pairings (Fig. 1). To examine the potential role of stimulus novelty in modulating learning and generalization, participants also underwent a Pre-exposure phase before Learning, during which half of the to-be-learned stimuli were presented individually. FMRI data were collected during all phases. Here we report the data from the Learning and Test phases, collapsed across the Pre-exposure manipulation because it did not differentially influence learning and generalization responses (see Supplemental Results). After participating in the task, participants were administered a brief post-test questionnaire to asses their awareness of the equivalencies in the associations and of the appearance of novel pairings at test.

In the Pre-exposure phase, each trial consisted of a single stimulus (face or scene) centrally presented for 1.25 s. Participants responded with a left or right keypress to indicate whether the stimulus was a person or a place. In the Learning phase, each trial consisted of the presentation of a face with two scenes for 3 s, during which participants indicated by keypress which of the two scenes was the correct associate for the face (Fig. 1A). Performance-dependent feedback (“Correct”, “Incorrect”, or “Too Late”) was provided after stimulus presentation and response, and remained on the screen for 1 s. In the Test phase, the trial structure was identical to Learning, except that no feedback was provided following the response (Fig. 1B).

For all phases, trials were intermixed with variable duration fixation null events; the total time allotted for null events was equal to 1/3 of the scan time. The duration and distribution of null

events was optimized for estimation of rapid event-related fMRI responses as calculated using Optseq (http://surfer.nmr.mgh.harvard.edu/optseq/).

Materials

Stimuli consisted of 24 pictures of faces, and 24 pictures of scenes. Faces and scenes were structured in sets, such that two faces (F_1, F_2) were paired with two scenes (S_1, S_2), resulting in four associations for each set: F_1–S_1, F_1–S_2, F_2–S_1, F_2–S_2. Three of these associations were trained during the Learning phase (F_1–S_1, F_1–S_2, F_2–S_1). During Test, subjects were tested on the fourth (untrained) association (F_2–S_2), as well as on the previously trained associations (F_1–S_1, F_1–S_2, F_2–S_1) (Fig. 1). The task consisted of 12 such stimulus sets, resulting in 36 different training trials and 48 different test trials. Importantly, a scene that was the correct choice for a certain face was also presented as the incorrect choice for a different face, such that simple stimulus-response learning strategies were not possible.

For counterbalancing purposes, the 12 sets were divided into 2 subsets of 6. During Pre-exposure, individual face and scene stimuli from one of the subsets were presented 15 times each, in random order. Which subset was pre-exposed was counterbalanced across subjects. During Learning, each of the 36 face-scene associations was repeated 8 times in total, with the scenes on each trial counterbalanced for left-right presentation. The 288 learning-phase trials were divided into 2 training blocks, and intermixed randomly with the constraint that each training block contained an equivalent number of presentations of each trial type. During Test, each trained and untrained (generalization) association was tested 6 times to provide increased power for functional imaging analyses of this phase, with trained and untrained associations intermixed and scenes counterbalanced for left-right presentation. All stimulus presentation orders were constrained such that no association appeared consecutively.

fMRI data acquisition

Whole-brain imaging was conducted on a 3.0T Signa MRI system (GE Medical Systems). Structural images were collected using a T2-weighted flow-compensated spin-echo pulse sequence (TR=3 s; TE=70 ms; 24 contiguous 5-mm thick slices parallel to the AC-PC plane). Functional images were collected using a T2*-weighted two-dimensional gradient echo spiral-in/out pulse sequence (TR=1.5 s; TE=30 ms; 1 interleave; flip angle = 70°; FOV= 20 cm; 64 _ 64 voxels) (Glover and Law, 2001).

The Learning phase was scanned in two 14-min functional runs. The Pre-Exposure (12 min) and Test (15 min) phases were each scanned in a single functional run. For each functional scan, 8 discarded volumes were collected prior to the first trial of the task. A bite bar was used to minimize head motion.

fMRI data analysis

Image preprocessing was performed using SPM2 (Wellcome Department for Cognitive Neurology, London). Functional images were corrected for differences in slice acquisition timing and then corrected for head motion. Each participant's structural images were co-registered to their functional images and segmented into grey matter, white matter and cerebrospinal fluid. The grey matter images were then stripped of any remaining skull and normalized to a grey matter MNI template image. This normalized grey matter image was used for normalization of the structural and functional images. Images were resampled to 3-mm cubic voxels and smoothed with a Gaussian kernel (8 mm full-width half-maximum).

Data were analyzed using SPM2, under the assumptions of the general linear model. Trials were modeled as an event, using a canonical hemodynamic response function and its first-order temporal derivative. Each phase (Pre-exposure, Learning, and Test) was analyzed separately.

During Learning, the first scan ('Early Learning') and the second scan ('Late Learning'), as well as correct and incorrect trials, were modeled separately. The resulting functions were entered into a general linear model with motion parameters entered as a covariate. Linear contrasts were used to obtain participant-specific estimates for each effect. These estimates were then entered into a second-level analysis, treating participant as a random effect, using a one-sample t-test against a contrast value of zero at each voxel. All contrasts were restricted to correct trials.

Because the outcomes from the main regression analysis (learning-phase activity regressed with generalization performance) was of central importance, effects within a priori regions of interest (ROIs) were small volume corrected using anatomical masks for these regions (hippocampus and midbrain/VTA). For the hippocampus, anatomical ROIs were drawn from a standard database (Anatomical Automatic Labeling; AAL). The midbrain ROI was created based on previous reports of VTA activity in humans during a declarative memory encoding task (Adcock et al., 2006): the reported coordinates for maximal activity in right and left VTA were entered as a seed region, and a 10-mm sphere was built around this peak. The resulting ROIs were summed and used as a single mask during analyses. These ROIs allowed for relatively conservative small volume correction.

ROI analyses were conducted to investigate effects revealed by voxel-based comparisons. ROIs included all significant voxels within a 6-mm radius of a maximum, or, where noted, all significant voxels within a cluster. Deconvolution of the signal within ROIs was performed using a finite impulse response function implemented with MarsBar (http://marsbar.sourceforge.net), allowing a comparison of the integrated percent signal changes (summed across 3.0 – 7.5 s post-trial onset) associated with conditions.

Cross-region interactions

The functional interaction between midbrain and hippocampus was assessed using a seed region covariate analysis (Poldrack et al., 2001). A seed region in the midbrain was functionally defined based on the main regression analysis revealing voxels that showed learning-related changes in activity that correlated significantly with generalization performance. The learning-phase change in activity in these midbrain voxels (the difference in integrated % signal change from early to late learning) for each subject was then entered as a regressor against the contrast of early vs. late learning, revealing any voxels that showed learning-phase changes that significantly correlated with changes in the functionally defined seed region in the midbrain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Supported by the National Institute of Mental Health (5R01MH080309–02 to A.D.W, and 5F32MH072135-03 to D.S.), National Alliance for Research on Schizophrenia and Depression (A.D.W.), and Alfred P. Sloan Foundation (A.D.W). The authors are grateful to Yair Avgar, Nathan Clement, and Itamar Kahn for assistance with data analysis, and R. Alison Adcock, Nathaniel Daw and Melina Uncapher for insightful discussion.

References

Adcock RA, Thangavel A, Whitfield-Gabrieli S, Knutson B, Gabrieli JD. Reward-motivated learning: mesolimbic activation precedes memory formation. Neuron 2006;50:507–517. [PubMed: 16675403]

Aron AR, Shohamy D, Clark J, Myers C, Gluck MA, Poldrack RA. Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. J Neurophysiol 2004;92:1144–1152. [PubMed: 15014103]

Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 2005;47:129–141. [PubMed: 15996553]

Bunzeck N, Duzel E. Absolute coding of stimulus novelty in the human substantia nigra/VTA. Neuron 2006;51:369–379. [PubMed: 16880131]

Cohen, NJ.; Eichenbaum, H. Memory, amnesia, and the hippocampal system. Cambridge, MA: MIT Press; 1993.

Collie A, Myers C, Schnirman G, Wood S, Maruff P. Selectively impaired associative learning in older people with cognitive decline. J Cogn Neurosci 2002;14:484–492. [PubMed: 11970807]

Daw ND, Doya K. The computational neurobiology of learning and reward. Curr Opin Neurobiol 2006;16:199–204. [PubMed: 16563737]

Daw ND, Shohamy D. The cognitive neuroscience of motivation and learning. Social Cognition. In Press

Delgado MR, Miller MM, Inati S, Phelps EA. An fMRI study of reward-related probability learning. Neuroimage 2005;24:862–873. [PubMed: 15652321]

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol 2000;84:3072–3077. [PubMed: 11110834]

Dennis NA, Kim H, Cabeza R. Effects of aging on true and false memory formation: an fMRI study. Neuropsychologia 2007;45:3157–3166. [PubMed: 17716696]

Dusek JA, Eichenbaum H. The hippocampus and memory for orderly stimulus relations. Proc Natl Acad Sci U S A 1997;94:7109–7114. [PubMed: 9192700]

Eichenbaum H. A cortical-hippocampal system for declarative memory. Nat Rev Neurosci 2000;1:41–50. [PubMed: 11252767]

Eichenbaum, HE.; Cohen, NJ. From Conditioning to Conscious Recollection: Memory Systems of the Brain. New York: Oxford University Press; 2001.

Faure A, Haberland U, Conde F, El Massioui N. Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. J Neurosci 2005;25:2771–2780. [PubMed: 15772337]

Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 2004;306:1940–1943. [PubMed: 15528409]

Frey U, Schroeder H, Matthies H. Dopaminergic antagonists prevent long-term maintenance of posttetanic LTP in the CA1 region of rat hippocampal slices. Brain Res 1990;522:69–75. [PubMed: 1977494]

Gabrieli JD. Cognitive neuroscience of human memory. Annu Rev Psychol 1998;49:87–115. [PubMed: 9496622]

Garoff RJ, Slotnick SD, Schacter DL. The neural origins of specific and general memory: the role of the fusiform cortex. Neuropsychologia 2005;43:847–859. [PubMed: 15716157]

Gasbarri A, Packard MG, Campana E, Pacitti C. Anterograde and retrograde tracing of projections from the ventral tegmental area to the hippocampal formation in the rat. Brain Res Bull 1994;33:445–452. [PubMed: 8124582]

Glover GH, Law CS. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med 2001;46:515–522. [PubMed: 11550244]

Gluck MA, Myers CE. Hippocampal mediation of stimulus representation: a computational theory. Hippocampus 1993;3:491–516. [PubMed: 8269040]

Graybiel AM. Building action repertoires: memory and learning functions of the basal ganglia. Curr Opin Neurobiol 1995;5:733–741. [PubMed: 8805417]

Greene AJ, Gross WL, Elsinger CL, Rao SM. An FMRI analysis of the human hippocampus: inference, context, and task awareness. J Cogn Neurosci 2006;18:1156–1173. [PubMed: 16839289]

Grice GR, Davis JD. Effect of concurrent responses on the evocation and generalization of the conditioned eyeblink. J Exp Psychol 1960;59:391–395. [PubMed: 13829256]

Hall G, Ray E, Bonardi C. Acquired equivalence between cues trained with a common antecedent. J Exp Psychol Anim Behav Process 1993;19:391–399. [PubMed: 8228835]

Hasselmo ME, McClelland JL. Neural models of memory. Curr Opin Neurobiol 1999;9:184–188. [PubMed: 10322183]

Hasselmo ME, Schnell E, Barkai E. Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. J Neurosci 1995;15:5249–5262. [PubMed: 7623149]

Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D. Hippocampal activation during transitive inference in humans. Hippocampus 2004;14:153–162. [PubMed: 15098721]

Kirwan CB, Stark CE. Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. Learn Mem 2007;14:625–633. [PubMed: 17848502]

Knowlton BJ, Mangels JA, Squire LR. A neostriatal habit learning system in humans. Science 1996;273:1399–1402. [PubMed: 8703077]

Kohler S, Danckert S, Gati JS, Menon RS. Novelty responses to relational and non-relational information in the hippocampus and the parahippocampal region: a comparison based on event-related fMRI. Hippocampus 2005;15:763–774. [PubMed: 15999342]

Kumaran D, Maguire EA. An unexpected sequence of events: mismatch detection in the human hippocampus. PLoS Biol 2006;4:e424. [PubMed: 17132050]

Leutgeb JK, Leutgeb S, Moser MB, Moser EI. Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science 2007;315:961–966. [PubMed: 17303747]

Lisman JE, Grace AA. The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 2005;46:703–713. [PubMed: 15924857]

McNaughton, BL.; Nadel, L. Neuroscience and Connectionist Theory. Hillsdale, N. J.: Lawrence Erlbaum; 1989.

Morris RG, Moser EI, Riedel G, Martin SJ, Sandin J, Day M, O'Carroll C. Elements of a neurobiological theory of the hippocampus: the role of activity-dependent synaptic plasticity in memory. Philos Trans R Soc Lond B Biol Sci 2003;358:773–786. [PubMed: 12744273]

Myers CE, Shohamy D, Gluck MA, Grossman S, Kluger A, Ferris S, Golomb J, Schnirman G, Schwartz R. Dissociating hippocampal versus basal ganglia contributions to learning and transfer. J Cogn Neurosci 2003;15:185–193. [PubMed: 12676056]

Norman KA, O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol Rev 2003;110:611–646. [PubMed: 14599236]

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. Neuron 2003;38:329–337. [PubMed: 12718865]

O'Reilly RC, Rudy JW. Conjunctive representations in learning and memory: principles of cortical and hippocampal function. Psychol Rev 2001;108:311–345. [PubMed: 11381832]

Otmakhova NA, Lisman JE. D1/D5 dopamine receptor activation increases the magnitude of early long-term potentiation at CA1 hippocampal synapses. J Neurosci 1996;16:7478–7486. [PubMed: 8922403]

Packard MG, McGaugh JL. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol Learn Mem 1996;65:65–72. [PubMed: 8673408]

Paller KA, Wagner AD. Observing the transformation of experience into memory. Trends Cogn Sci 2002;6:93–102. [PubMed: 15866193]

Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA. Interactive memory systems in the human brain. Nature 2001;414:546–550. [PubMed: 11734855]

Poldrack RA, Packard MG. Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia 2003;41:245–251. [PubMed: 12457750]

Preston AR, Shrager Y, Dudukovic NM, Gabrieli JD. Hippocampal contribution to the novel use of relational information in declarative memory. Hippocampus 2004;14:148–152. [PubMed: 15098720]

Schacter, DL. Forgotten Ideas, Neglected Pioneers: Richard Semon and the Story of Memory. Philadelphia, PA: Psychology Press; 2001a.

Schacter, DL. The Seven Sins of Memory: How the Mind Forgets and Remembers. New York City: Houghton Miffling; 2001b.

Schott BH, Henson RN, Richardson-Klavehn A, Becker C, Thoma V, Heinze HJ, Duzel E. Redefining implicit and explicit memory: the functional neuroanatomy of priming, remembering, and control of retrieval. Proc Natl Acad Sci U S A 2005;102:1257–1262. [PubMed: 15657126]

Schott BH, Sellner DB, Lauer CJ, Habib R, Frey JU, Guderian S, Heinze HJ, Duzel E. Activation of midbrain structures by associative novelty and the formation of explicit memory in humans. Learn Mem 2004;11:383–387. [PubMed: 15254215]

Schultz W. The phasic reward signal of primate dopamine neurons. Adv Pharmacol 1998;42:686–690. [PubMed: 9327992]

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997;275:1593–1599. [PubMed: 9054347]

Shohamy D, Myers CE, Geghman KD, Sage J, Gluck MA. L-dopa impairs learning, but spares generalization, in Parkinson's disease. Neuropsychologia 2006;44:774–784. [PubMed: 16150469]

Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA. Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain 2004;127:851–859. [PubMed: 15013954]

Shohamy D, Myers CE, Kalanithi J, Gluck MA. Basal ganglia and dopamine contributions to probabilistic category learning. Neurosci Biobehav Rev 2008;32:219–236. [PubMed: 18061261]

Squire LR. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev 1992;99:195–231. [PubMed: 1594723]

Strange BA, Dolan RJ. Adaptive anterior hippocampal responses to oddball stimuli. Hippocampus 2001;11:690–698. [PubMed: 11811663]

Swanson LW. The projections of the ventral tegmental area and adjacent regions: a combined fluorescent retrograde tracer and immunofluorescence study in the rat. Brain Res Bull 1982;9:321–353. [PubMed: 6816390]

Titone D, Ditman T, Holzman PS, Eichenbaum H, Levy DL. Transitive inference in schizophrenia: impairments in relational memory organization. Schizophr Res 2004;68:235–247. [PubMed: 15099606]

Walther E. Guilty by mere association: evaluative conditioning and the spreading attitude effect. J Pers Soc Psychol 2002;82:919–934. [PubMed: 12051580]

Wittmann BC, Daw ND, Seymour B, Dolan RJ. Striatal activity underlies novelty-based choice in humans. Neuron 2008;58:967–973. [PubMed: 18579085]

Wittmann BC, Schott BH, Guderian S, Frey JU, Heinze HJ, Duzel E. Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. Neuron 2005;45:459–467. [PubMed: 15694331]

Yamaguchi S, Hale LA, D'Esposito M, Knight RT. Rapid prefrontal-hippocampal habituation to novel events. J Neurosci 2004;24:5356–5363. [PubMed: 15190108]

Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. Nat Rev Neurosci 2006;7:464–476. [PubMed: 16715055]

Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 2004;19:181–189. [PubMed: 14750976]

Zeineh MM, Engel SA, Thompson PM, Bookheimer SY. Dynamics of the hippocampus during encoding and retrieval of face-name pairs. Science 2003;299:577–580. [PubMed: 12543980]

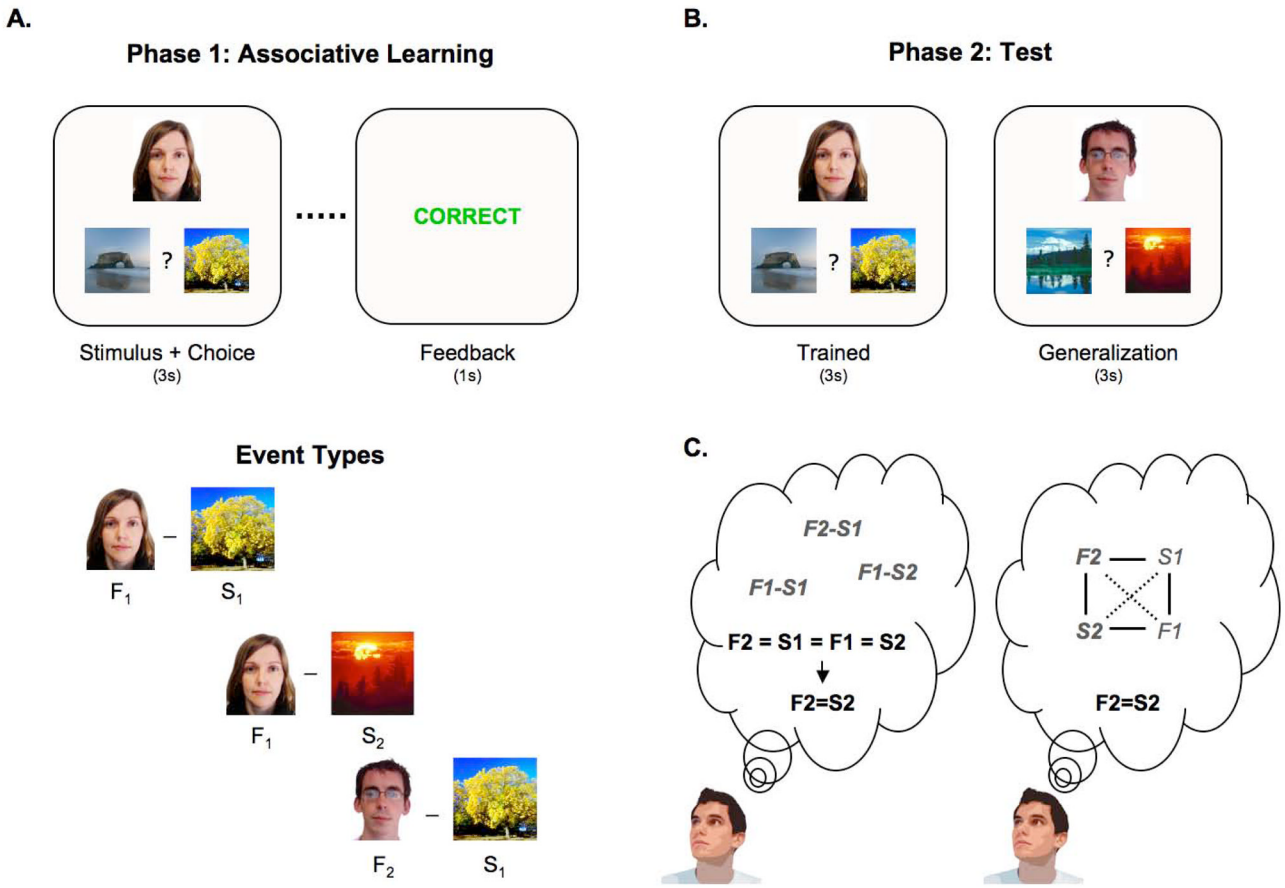

Figure 1.

Representative events and structure of the task. **(A)** Participants learned a series of individual face-scene associations based on feedback (36 individual associations in total). On each trial, the face-scene pair was presented for 3 s, after which performance-dependent feedback was provided for 1 s. There were three learning event types—the individual associations shared overlapping features, with two faces always associated with a common scene, and one of those faces also associated with a second scene. A scene that was the incorrect choice for one face was the correct choice for another face, so that simple stimulus-response learning strategies could not support learning. **(B)** After learning, participants underwent a test phase, where they received no feedback and where they were asked to respond to untrained face-scene associations. These generalization trials were presented together with trials that tested knowledge for previously trained associations. **(C)** The generalization trials can be correctly responded to by way of two different mechanisms: during test, retrieval of the previously trained individual associations may allow participants to draw inferences across them (left); alternatively, the untrained association may have been formed during learning due to retrieval and integrative encoding that is triggered by the overlapping features across individual trained associations.

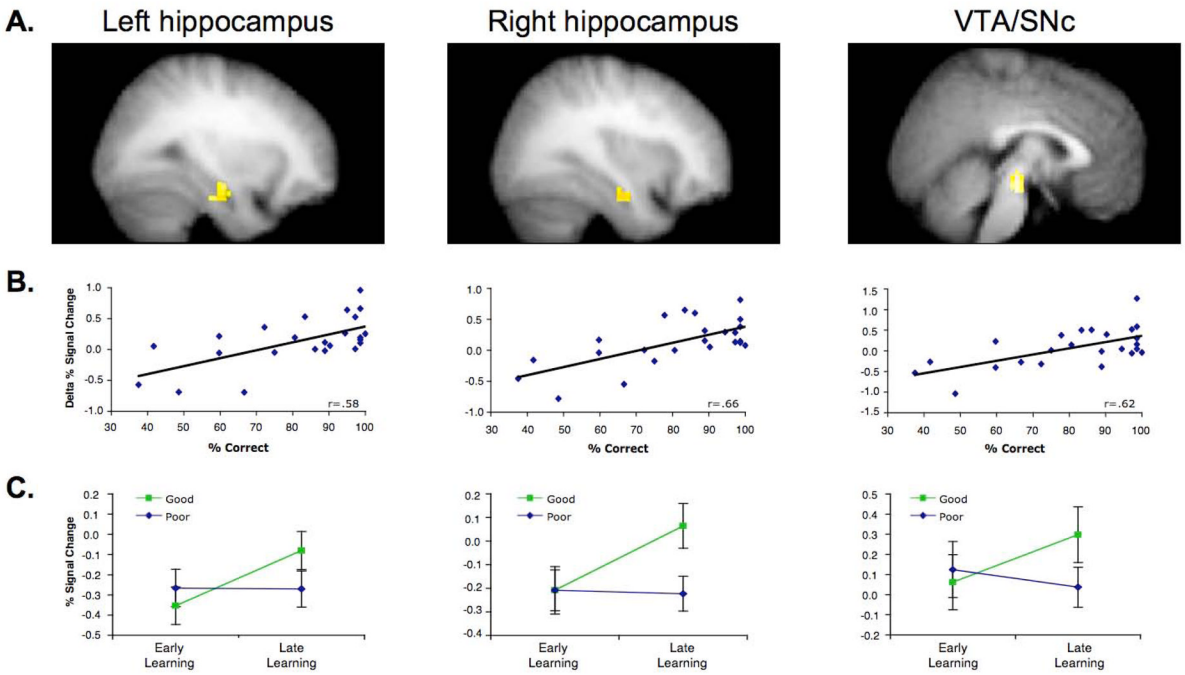

Figure 2.

Hippocampal and midbrain activation during learning predicts correct responding on the generalization trials at test. **(A)** Map-wise regression analyses revealed that the change in activation from early to late learning in left hippocampus (−28,−9,−17; 27 voxels), right hippocampus (31,−5,−20; 22 voxels), and a bilateral midbrain complex (3,−18,−12; 50 voxels) correlated with % correct generalization performance at test (P<0.001, extent threshold 5 voxels; P<0.05, small volume corrected for the hippocampus and midbrain). **(B)** BOLD % signal change data extracted from these hippocampal and midbrain regions (inclusive of all above-threshold voxels within a 6-mm sphere surrounding the peak voxel) confirmed the strong correlation between learning-phase activation increases and generalization performance. **(C)** When participants were median split based on generalization performance at test, an increase in hippocampal and midbrain activation from early to late learning was observed in participants who generalized well (‘good’ group), but not in participants who generalized poorly (‘poor’ group). Error bars +/− S.E.M.

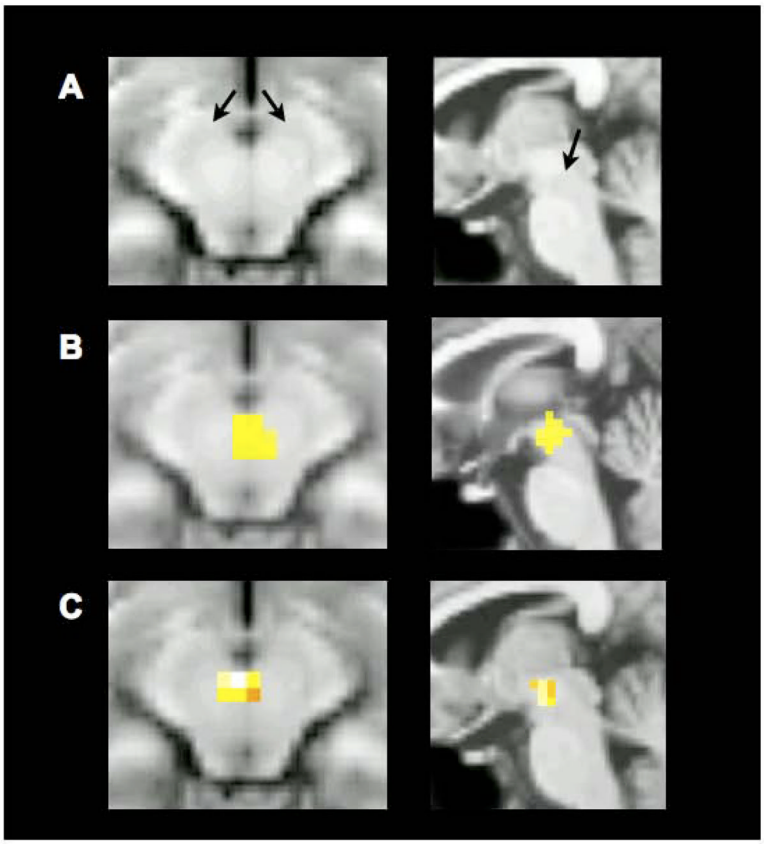

Figure 3.

Localization of midbrain activations, displayed on a canonical T1-weighted image (axial slice, left; sagittal slice, right). The midbrain complex consists of the substantia nigra (SN) and the ventral tegmental area (VTA). **(A)** SN extends lateral and posterior around the oval red nuclei, as indicated by the black arrows. VTA is medial to SN, and borders the interpeduncular cistern. **(B)** Higher magnification of the generalization-related midbrain region-of-interest described in the main findings (data smoothed with an 8-mm filter; $P<0.001$, extent threshold 5 voxels; $P<0.05$, small volume corrected for the hippocampus and midbrain). **(C)** Visualization of generalization-related midbrain activations revealed when using a smaller (4-mm) smoothing filter during functional data preprocessing ($P<0.001$, extent threshold 5 voxels; $P<0.05$, small volume corrected for the hippocampus and midbrain). Full reporting of the data smoothed with a 4-mm filter appear in the Supplemental Results.

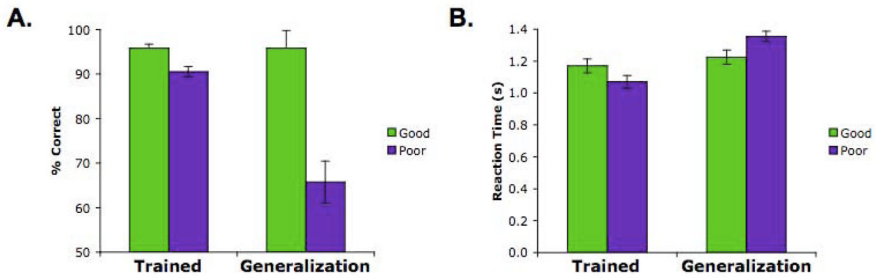

Figure 4.

Behavioral performance at test on trained and generalization trials for the 'good' and 'poor' generalization participants. **(A)** An interaction between group and test trial type revealed a significantly greater difference between the two groups in performance on the generalization trials relative to the trained trials. Importantly, the 'good' group showed no difference in accuracy between trained and generalization trials, whereas the 'poor' group showed superior performance on trained than on generalization trials. **(B)** This pattern was also evident in the speed of responding. The 'poor' group, relative to the 'good' group, showed a marked difference in response latencies to trained vs. generalization trials, consistent with the hypothesis that in the 'good' group the associations necessary to rapidly respond to generalization trials were constructed during learning. Error bars +/− S.E.M.

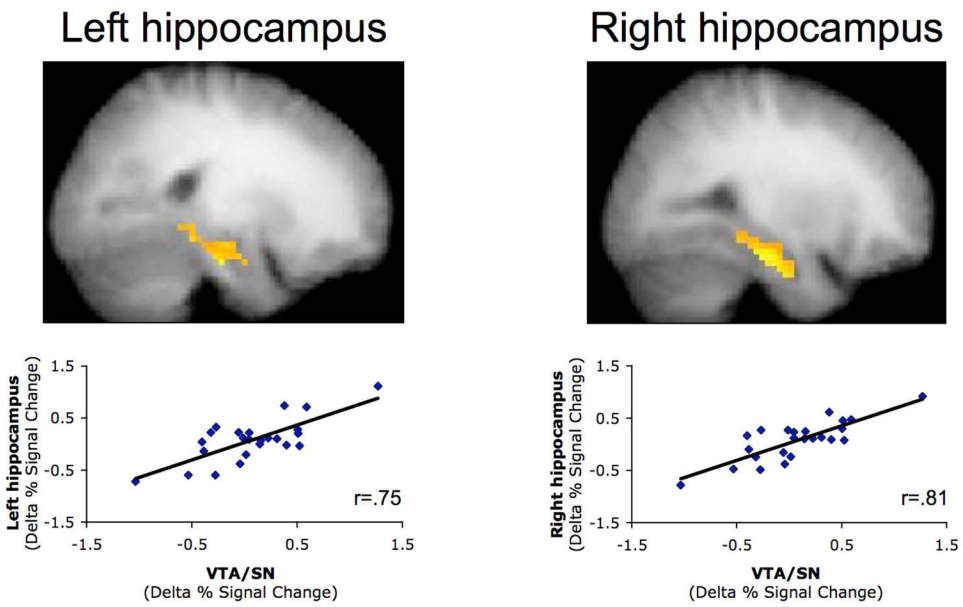

Figure 5.

Learning-phase activation changes in bilateral hippocampus demonstrated a significant correlation with such changes in the midbrain. This analysis regressed the difference in % signal change from early to late learning in a seed region in the midbrain with voxels in the medial temporal lobe ($P<0.001$, extent threshold 5 voxels; small volume corrected, $P<0.05$). Extracting the change in integrated % signal change for all activated voxels in left and right hippocampus (98 and 161 voxels, respectively) identified in this regression confirmed the tight relationship with the change in integrated % signal change in the midbrain. Importantly, this relationship between hippocampal and midbrain activation remained significant even when excluding the two participants demonstrating the strongest and weakest change in midbrain learning-phase activity (left hippocampus–midbrain r=0.59; $P<0.005$; right hippocampus–midbrain, r=0.75, $P<0.001$).

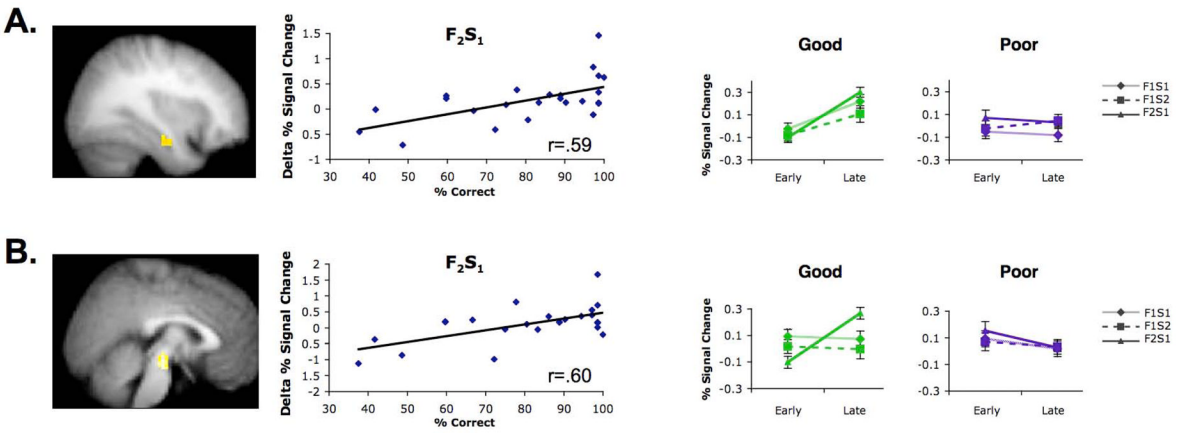

Figure 6.

Hippocampal and midbrain activation to different event types during learning. The relationship between **(A)** hippocampal and **(B)** midbrain activation during learning and generalization performance at test was strongest for the F_2–S_1 learning trials. Regression analyses (left) revealed that subsequent generalization correlated with learning-phase activation increases to F_2–S_1 trials in both bilateral hippocampus (data shown for right hippocampus) and midbrain (*P*s<.05, corrected); no other correlations survived correction for multiple comparisons. Similarly, increased activation during learning of F_2–S_1 events showed the strongest difference across 'good' and 'poor' generalization subgroups. The 'good' (*P*s<0.05), but not the 'poor' (*P*s>0.40), generalization group demonstrated a significant increase in bilateral hippocampal and midbrain activation from early to late learning of F_2–S_1 trials. In the midbrain, this increase was selective to the F_2–S_1 trials, whereas in the hippocampus, a qualitatively similar effect was observed for the F_1–S_1 and F_1–S_2 trials (see Supplemental Results). Error bars +/− S.E.M.