



Published in final edited form as:

J Cogn Neurosci. 2018 May ; 30(5): 680–697. doi:10.1162/jocn_a_01228.

The Necessity of the Hippocampus for Statistical Learning

Natalie V. Covington, Sarah Brown-Schmidt, and Melissa C. Duff

Vanderbilt University, Nashville, TN

Abstract

Converging evidence points to a role for the hippocampus in statistical learning, but open questions about its necessity remain. Evidence for necessity comes from Schapiro and colleagues who report that a single patient with damage to hippocampus and broader medial temporal lobe cortex was unable to discriminate new from old sequences in several statistical learning tasks. The aim of the current study was to replicate these methods in a larger group of patients who have either damage localized to hippocampus or a broader medial temporal lobe damage, to ascertain the necessity of the hippocampus in statistical learning. Patients with hippocampal damage consistently showed less learning overall compared with healthy comparison participants, consistent with an emerging consensus for hippocampal contributions to statistical learning. Interestingly, lesion size did not reliably predict performance. However, patients with hippocampal damage were not uniformly at chance and demonstrated above-chance performance in some task variants. These results suggest that hippocampus is necessary for statistical learning levels achieved by most healthy comparison participants but significant hippocampal pathology alone does not abolish such learning.

INTRODUCTION

Moving through the world, we are confronted with a barrage of stimuli: from the sounds of language to the details of visual stimuli. Our brains seek to uncover underlying patterns in the environment by tracking the frequencies with which these stimuli occur as well as the likelihood of co-occurrence between stimuli. These patterns allow us to transform streams of sounds into words and visual details into integrated images. This capacity, called statistical learning, is considered a robust phenomenon observed in humans from infancy through adulthood (Baldwin, Andersson, Saffran, & Meyer, 2008; Conway & Christiansen, 2005; Saffran, Aslin, & Newport, 1996) and has been theorized to underlie a range of abilities including early word segmentation and the acquisition of cognitive maps from visual information (Brady & Oliva, 2008; Saffran et al., 1996).

Statistical learning refers to the process of extracting underlying patterns from the environment, based solely on the statistical regularities in the stimulus input. In an early demonstration of this phenomenon, infants were able to segment “words” from structured continuous stimuli, in which the only cue to word boundaries was the transitional probabilities between syllables (Saffran et al., 1996). Statistical learning has since been

demonstrated in older children and adults and across multiple modalities (e.g., streams of visual stimuli and nonspeech auditory input; Creel, Newport, & Aslin, 2004; Fiser & Aslin, 2001, 2002). Statistical learning has been of particular interest to language researchers, as a potential mechanism supporting language acquisition processes including word segmentation and word learning from continuous speech (Graf Estes, Evans, Alibali, & Saffran, 2007; Saffran et al., 1996), leveraging of adjacent and non-adjacent dependencies for syntax acquisition (Saffran & Wilson, 2003; Gómez, 2002), and tracking of phonotactic, orthographic, and morphological regularities (Pacton, Fayol, & Perruchet, 2005; Chambers, Onishi, & Fisher, 2003).

Although statistical learning has been well characterized behaviorally, its neural basis is less clear. Statistical learning has been characterized as an unconscious, implicit process, and it has been likened to other implicit learning paradigms, such as artificial grammar learning (AGL) and serial reaction time (SRT) tasks (Kim, Seitz, Feenstra, & Shams, 2009; Perruchet & Pacton, 2006). Both statistical learning and traditional implicit learning tasks such as AGL and SRT require learning or extraction of underlying patterns from complex stimuli. These characteristics of statistical learning have led to theoretical accounts of statistical learning that suggest that the nature of learning fits parsimoniously with the learning capacities of the BG (Evans, Saffran, & Robe-Torres, 2009; Kim et al., 2009; Perruchet & Pacton, 2006).

Neuroimaging studies have implicated a widespread and disparate constellation of brain structures in statistical learning. Proposed neural substrates include uni-modal cortical networks, with differences depending on stimulus modality, including visual (Turk-Browne, Scholl, Chun, & Johnson, 2009; Bischoff-Grethe, Proper, Mao, Daniels, & Berns, 2000) and auditory (Karuza et al., 2013; McNealy, Mazziotta, & Dapretto, 2006) networks. Other studies identify networks that contribute to statistical learning regardless of stimulus modality, including BG (Karuza et al., 2013; Turk-Browne et al., 2009). Taken together, these neuroimaging findings suggest that statistical learning is not the purview of a single neural system but rather the product of multiple systems working in parallel. (Note that AGL, the most similar implicit learning task to statistical learning, also appears to rely on a distributed network including frontal, occipital, and parietal cortices as well as subcortical structures; Skosnik et al., 2002; Seger, Prabhakaran, Poldrack, & Gabrieli, 2000; Fletcher, Büchel, Josephs, Friston, & Dolan, 1999).

Of particular interest to the current study is work linking statistical learning to medial temporal lobe (MTL) structures, including hippocampus. During a canonical visual statistical learning task, Turk-Browne and colleagues measured neural responses to structured versus random stimulus input. This study confirmed BG involvement in statistical learning, with significant activation in the striatum, and additionally extended the domain-general network supporting statistical learning to MTL, including hippocampus (Turk-Browne et al., 2009). Given the high degree of overlap between the characteristics of statistical learning and other implicit learning paradigms, a role for the hippocampus is surprising at first glance; however, consideration of the processing capabilities of the hippocampus suggests that it may be a good candidate structure for the demands of statistical learning. In fact, hippocampal involvement has been documented for some versions of the AGL paradigm (e.g., balanced chunk-based designs [Lieberman, Chang,

Chiao, Bookheimer, & Knowlton, 2004] and biconditional grammars [Channon et al., 2002]).

Statistical learning requires the ability to encode and track the relations between individual stimuli, and a hallmark processing feature of the hippocampus is its ability to support the rapid binding of arbitrarily related elements that make up a scene or event and their temporal, spatial, and interactional relations (Eichenbaum & Cohen, 2001). Furthermore, new evidence suggests that the hippocampus contributes to unconscious processing of relational binding and on a rapid time course that would position the hippocampus as a potential contributor to statistical learning (Hannula & Greene, 2012; Hannula, Tranel, & Cohen, 2006).

The most compelling evidence for a link between hippocampus and statistical learning comes from a study demonstrating impaired statistical learning in a single patient with damage to the MTL (Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014). Patient L. S. J. had complete loss of hippocampus bilaterally as well as a more extensive damage to surrounding MTL cortex and left anterior temporal lobe as a result of herpes simplex encephalitis (HSE). This significant MTL damage resulted in profound declarative memory impairment (Wechsler Memory Scale–General Memory Index [WMS GMI] = 47) as well as working memory impairment (Wechsler Memory Scale–Working Memory Index [WMS WMI] = 76), in the context of preserved intelligence (Wechsler Adult Intelligence Scale–Full Scale Intelligence Quotient [WAIS FSIQ] = 92), perception, language, and world knowledge (Gregory, McCloskey, & Landau, 2014). L. S. J. completed four versions of two experiments in which she was passively exposed to continuous sequences of shapes, syllables, scenes, or tones. These continuous streams of stimuli contained an underlying structure of triplets (Experiment 1) or pairs (Experiment 2). In Experiment 1, patient L. S. J. performed at chance and significantly worse than matched healthy comparison participants. However, although comparison participants in Experiment 1 performed better than chance at the group level, they demonstrated significant variability, with a sizeable minority performing at chance levels (this variability in performance in healthy adult participants is consistent with the statistical learning literature more broadly; see (Erickson, Kaschak, Thiessen, & Berry, 2016; Arciuli, Torkildsen, Stevens, & Simpson, 2014; Misyak & Christiansen, 2012; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Therefore, Schapiro et al. implemented Experiment 2, which consisted of repeating pairs of stimuli. The length of exposure to the continuous stimulus stream was kept the same as in Experiment 1. Thus, Experiment 2 consisted of simplified regularities (pairs vs. triplets) as well as increased exposure to each target pair (1.5 times the exposure to pairs compared with target triplets in Experiment 1). These modifications resulted in improved comparison participant performance, whereas patient L. S. J. remained at chance. Given L. S. J.'s severe deficit in statistical learning across both experiments, Schapiro and colleagues conclude that the MTL plays a critical role in statistical learning. Schapiro and colleagues also acknowledge that the role of the hippocampus in producing the observed deficits cannot be dissociated from L. S. J.'s broader MTL damage and that work with patients with more selective damage will be crucial for understanding the role of different regions.

Converging evidence points to a role for the hippocampus in statistical learning, yet open questions remain regarding the necessity of the hippocampus in statistical learning. There are, of course, different standards and interpretations of what it means for a brain structure to be necessary for a particular process. One interpretation of necessity might be that the hippocampus is required to demonstrate any statistical learning at all. For example, in the semantic memory literature, data showing that patients with hippocampal damage can show some new learning, albeit less and more slowly than healthy participants, have been taken as evidence that the hippocampus may not be necessary for all forms of declarative memory (e.g., O’Kane, Kensinger, & Corkin, 2004; see Squire & Zola, 1998). Alternatively, the definition of necessity might dictate that patients with lesion perform normally (i.e., as defined by the performance of matched healthy comparison participants). Under this standard, we might conclude that the hippocampus is necessary for statistical learning even if patients demonstrate learning but when that performance falls well short of the mean performance of the healthy participant group.

There are also different thresholds by which one might judge the significance of any observed statistical learning. Performance of healthy participants in within-group studies is often compared to chance (e.g., Schlichting, Guarino, Schapiro, Turk-Browne, & Preston, 2017; Arciuli et al., 2014; Toro, Sinnott, & Soto-Faraco, 2005; Creel et al., 2004; Saffran, Johnson, Aslin, & Newport, 1999). Comparing performance to chance levels is also a common metric of learning and memory in the amnesia literature (e.g., Dede, Frascino, Wixted, & Squire, 2016; Schapiro et al., 2014; Hannula et al., 2006). In between-group studies of statistical learning of disordered populations, performance is often compared with healthy participant performance (e.g., Evans et al., 2009).

The current study is a replication and extension of the Schapiro et al. study. Through comparison of a group of patients with bilateral hippocampal damage, but who vary in the extent of broader MTL damage, with a large demographically matched sample of healthy comparison participants ($N = 112$), we provide a test of the necessity of the hippocampus for statistical learning in humans. Our data analysis approach is varied, considering both group and individual patient performances across conditions as well as comparison of patient performance with that of the comparison group. The data are also interpreted against distinct definitions of necessity to provide a more nuanced, and complex, understanding of the role of the hippocampus in statistical learning than is possible from neuroimaging or single-case reports alone and to better characterize the scope of preserved or impaired learning abilities in amnesia.

METHODS

Materials were obtained from the authors of Schapiro et al.’s (2014) study. Experimental tasks consisted of a statistical learning task in four modalities (shapes, syllables, scenes, and tones) and across two levels of difficulty (Experiment 1: triplets, Experiment 2: pairs). Following Schapiro and colleagues, patients with amnesia completed all conditions of both experiments, whereas comparison participants completed only a single condition, to ensure naïveté to the test.

Participants

Four individuals with bilateral hippocampal damage and severe declarative memory impairment (hippocampal amnesia) and 112 healthy adults participated in this study. Three participants sustained bilateral hippocampal damage from an anoxic/hypoxic event (e.g., cardiac arrest) resulting in damage thought to be restricted to the hippocampus (1846, 2363, 2563), and one participant developed HSE (1951) resulting in extensive damage to the hippocampus, amygdala, and other MTL cortices bilaterally (similar to the damage in L. S. J. based on the description and visual inspection of scans reported in Shapiro et al.). Structural MRI completed on three of the four patients confirmed bilateral hippocampal damage and significantly reduced hippocampal volumes (studentized residual differences in hippocampal volume relative to a matched comparison group down by at least 2.6 *z* scores; Allen, Tranel, Bruss, & Damasio, 2006; Buchanan, Tranel, & Adolphs, 2005; Figure 1). For 2563 (who wears a pacemaker and was unable to undergo the MRI examination), anatomical analysis was based on computerized tomography and damage confined to the hippocampal region was visible.

All patients with bilateral hippocampal damage had a severe impairment in declarative memory ($M = 65.5$; WMS-III GMI). All participants with hippocampal amnesia performed within normal limits on standardized neuropsychological measures of intelligence ($M = 97.4$; WAIS-III FSIQ). Despite their severe declarative memory impairment, the amnesic patients did not have significant disruptions on standardized neuropsychological measures of language (e.g., Boston Naming Test, Token Test), visual perception (e.g., Complex Figure Test copy), executive function (e.g., Wisconsin Card Sorting), or working memory (e.g., WMS WMI; see Table 1).

One hundred twelve healthy comparison participants (56 women, 56 men) were case-matched on age, years of education, and gender to participants with amnesia. Female patient 1846 was 52 years old and held 14 years of education. Female comparison participants ($N = 56$) were, on average, 53.54 ($SD = 4.45$) years old and held an average of 15.56 ($SD = 1.74$) years of education. Male patients 1951, 2363, and 2563 were, on average 60.67 ($SD = 2.08$) years old and held an average of 16.67 ($SD = 1.15$) years of education. Male comparison participants ($N = 56$) were, on average, 62.38 ($SD = 5.91$) years old and held an average of 16.60 ($SD = 2.09$) years of education. Healthy comparison participants had no history of neurological, psychiatric, or developmental impairments, per self-report.

Stimuli

Stimulus items for all conditions (shapes, syllables, scenes, tones) were identical to those utilized in Shapiro et al. (2014; Figure 2). Shape stimuli consisted of 12 black nonsense shapes drawn from previous SL studies (Fiser & Aslin, 2001, 2002) presented on a white background. Syllable stimuli consisted of 12 CV syllables (/boʊ/, /di/, /nu/, /ru/, /lɑ/, /zɛ/, /mɛ/, /kɑ/, /wu/, /vɑ/, /pi/, /foʊ/) generated using a male voice from the Mbrola speech synthesizer (tctsfpm.ac.be/synthesis/). Scene stimuli consisted of 12 photographs of outdoor scenes presented on a black background. Tone stimuli consisted of 12 pure tone stimuli (frequencies: 262, 278, 294, 311, 330, 349, 370, 392, 415, 440, 466, and 494 Hz). Stimuli were presented using the Psychophysics Toolbox (Brainard, 1997) for MATLAB on

Macintosh computers and a script obtained from Schapiro and colleagues. For shapes, scenes, and tones, stimuli were presented for 0.5 sec with an ISI of 0.5 sec. Syllables were presented as a continuous sequence (no ISI), with a syllable duration of 0.25 sec.

Procedure

Participants with amnesia completed all conditions in both Experiments 1 and 2. At least 4 hr (with intervening distractor tasks) elapsed between conditions, and no more than two conditions were presented in a single day. Comparison participants, in contrast, participated in only a single condition to preserve the surprise of the final test. Experiments 1 and 2 both consisted of two phases: an exposure phase and followed by a surprise test phase. Before the exposure phase, participants were told that they would watch/listen to a series of shapes/syllables/scenes/tones and were asked to pay close attention. They were not informed of any upcoming test. Participants watched/listened to the stimulus sequence for 4.8 min. Shape, scene, and tone sequences were 228 items long. Syllable sequences were 1,152 items long. Because the regularities to be learned in Experiment 2 were simpler (stimulus pairs instead of triplets), but the exposure duration was equivalent (4.8 min for both Experiments 1 and 2), each stimulus pair in Experiment 2 was presented approximately 1.5 times more frequently than stimulus triplets in Experiment 1, consistent with Schapiro and colleagues.

For each modality, stimulus items were randomly assigned without replacement to specific triplets (e.g., /pizewu/, /foonuka/, /boomedi/) for Experiment 1 and specific pairs for Experiment 2. Items within a triplet/pair always maintained their fixed order consecutively, and these triplet/pair assignments were kept constant for all participants. To produce a continuous sequence, in Experiment 1, triplets were presented in pseudorandom order with the following rules: (1) No triplet could be repeated in immediate succession, and (2) no pair of triplets could be repeated in immediate succession. In Experiment 2, pairs were presented in a fully random order, with no constraints of repetition of pairs.

The test phase consisted of a 32-item two-alternative forced-choice (2AFC) test for Experiment 1 and an 18-item 2AFC test for Experiment 2. At the beginning of the test phase, participants were told that they would see/hear two possible sequences and were asked to select the sequence that was most familiar to them, based on what they had just watched/listened to. Participants were encouraged to guess if they were unsure. Healthy comparison participants indicated their choice using the arrow keys on the keyboard. Consistent with Schapiro et al. (2014), patients with amnesia indicated their choice by pointing, with responses recorded on the keyboard by the experimenter. Foils were composed of items from the exposure phase, but in an order that had never occurred during exposure. Target triplets/pairs and foil triplets/pairs occurred equally as often during the test.

For each test trial in the visual modality (shapes, scenes), the first stimulus triplet/pair (either the target or the foil) was presented on either the left or right side of the screen, with the same timing as during exposure. After a 1-sec delay, the second triplet/pair alternative appeared on the opposite side of the screen. Comparison participants used the arrow keys on the keyboard to select whether the “left” or “right” alternative was more familiar to them (patient participants indicated their choice by pointing). Both the order of appearance (first or second) and side of the screen (left or right) of the target triplet/pair were randomized

across test trials. Test trial presentation was identical for the auditory modality (syllables, tones), except that each alternative was played from either the left or right speaker. A black arrow appeared while each alternative was played, indicating which speaker was playing the stimulus.

To rule out a role for task demands unrelated to statistical learning to any impairment in performance in the patients, a third experiment (Experiment 3) was conducted. Experiment 3 began with an exposure phase similar to Experiment 2. However, in the test phase, instead of selecting which of the two sequences was more familiar, patients selected which of two individual items was more familiar (i.e., item recognition test). Better-than-chance performance on the item recognition test would indicate that patients were able to attend to the task during the exposure phase, follow test instructions, and make reliable familiarity judgments. The tone version was omitted from Experiment 3, as patients would require “perfect pitch” to accurately identify single tones during the item recognition test.

General

One final note regarding the procedures used here and the original Schapiro et al. study concerns a slight difference in the construction of the stimuli. All stimulus items and procedures were identical to those used in Schapiro et al. (2014), with the exception that the assignment of stimulus items to triplets and pairs was different. Although the constituent elements that made up the pairs and triplets were drawn from the same stimulus pool as in Schapiro et al., because of a miscommunication, when we constructed our stimulus lists, the composition of the triplets and pairs themselves differed across studies. For example, whereas /piz:ːwu/ was a to-be-learned triplet in the present Experiment 1, /m:ːk*ːwu/ but not /piz:ːwu/ was a to-be-learned triplet in Schapiro et al.’s (2014) study. Note that, in this study, all patients with amnesia and all comparison participants completed tasks with identical triplet assignments. If the paradigms used here are truly tapping into fundamental learning mechanisms that support the acquisition of language and other cognitive abilities, one random stimulus composition should provide an equally good test of the necessity of hippocampus to statistical learning as a different random stimulus composition. However, to ensure that any observed differences between the two studies were not related to this minor difference in stimuli composition, we also reran the patients with amnesia using the identical triplet assignment employed by Schapiro et al. in Experiments 4 and 5.

We should also note that we chose a modified version of the analysis approach used by Schapiro and colleagues (2014). Like Schapiro et al., we analyze performance on each individual experiment separately. Also following Schapiro et al., within each experiment, we first conduct an omnibus model to compare overall performance of patients with that of healthy comparison participants and then conduct planned analyses of each group separately to evaluate whether patients achieved above-chance performance. In contrast to Schapiro et al., we do not use *t* tests over proportion correct within a version. The variance of a sample proportion is determined by the mean probability; thus, *t* tests over proportions are inappropriate and can lead to spurious results (Jaeger, 2008). A better approach is to use logistic regression where the model is used to predict the binary outcome measure—in these

studies, then, we analyze whether the participant was correct at the level of individual trials (Quené & van den Bergh, 2008; Snijders & Bosker, 1999).

RESULTS

Experiment 1

We analyzed the effects of Group (patients vs. comparisons), Task version (shape/syllable/scene/tone), and their interaction on item level accuracy using a logistic model with random intercepts for participants. The fixed effect of participant Group was coded as a mean-centered contrast (patients = 1.56, comparisons = -0.44). The fixed effect of Task version was entered as mean-centered Helmert contrasts. The first contrast compared scene (-0.75) with the average of shape (0.25), syllable (0.25), and tone (0.25). The second contrast compared shape (-0.625) with the average of syllable (0.375) and tone (0.375), ignoring scene (-0.125). The third contrast compared syllable (-0.5) with tone (0.5), ignoring scene (0.0) and shape (0.0). These contrasts were maintained for all subsequent analyses. The fixed effect parameters from this model are summarized in Appendix Table A1. As a group, Experiment 1 participants (patients and comparisons together) demonstrated statistical learning above what would be expected by chance. There was a marginal effect of Group ($b = -0.381$, $p = .08$), with patients performing numerically worse than comparisons. We also observed a significant interaction between participant Group and the second Version contrast (shape task vs. syllable and tone tasks), $b = 0.448$, $p < .05$. Planned comparisons examined these learning effects for the two groups separately.

Analyzed separately, comparison participants demonstrated statistical learning (mean accuracy by version: shape, $M = 72.32\%$; syllable, $M = 70.09\%$; scene, $M = 72.54\%$; tone, $M = 72.54\%$; see Figure 3). The results of this analysis indicated that comparison participants demonstrated statistical learning reliably better than chance ($b = 1.10$, $p < 2e-16$), with no effect of version. The fixed effect parameters from this model are summarized in Appendix Table A2.

Analyzed separately, patients with amnesia demonstrated statistical learning, although this learning effect was inconsistent across the different versions of the task (mean accuracy by version: shape, $M = 46.09\%$; syllable, $M = 54.69\%$; scene, $M = 56.25\%$; tone, $M = 73.44\%$; see Figure 3). The statistical analysis revealed that patients with amnesia demonstrated statistical learning significantly better than chance overall ($b = 0.325$, $p = .0004$). However, a significant effect of the second Version contrast, which compared the shape task with the syllable and tone tasks ($b = 0.759$, $p = .0006$), and a significant effect of the third Version contrast, which directly compared the syllable and tone versions of the task ($b = 0.829$, $p = .002$), indicated that performance was uneven across the versions of the task. In particular, patients were more successful with the tone version of the task than the other three versions (Figure 3). The fixed effect parameters from this model are summarized in Appendix Table A3. Notably, Patient 1951 (with damage to MTL cortex beyond the hippocampus, indicated by triangles in Figure 3) demonstrated the highest mean performance of any patient participant on both the shape and syllable versions as well as performance in line with anoxic patients on the scene and tone versions.

One question is whether poorer performance by patients is due to the fact that patients repeated the task multiple times, perhaps resulting in interference effects. To address this question, we tested for order effects in post hoc analyses and found that task order did not significantly impact patient performance ($z = 1.102$, $p = .271$). Another question regards outliers. All the patient data were within $1.5 \times$ interquartile range (IQR; Tukey, 1977), indicating that there were no patient data points that would be considered an outlier. This is the case for Experiment 1 as well as all other experiments reported below (2–5).

Experiment 2

In line with Schapiro et al. (2014), Experiment 2 contained simplified regularities (pairs), with more frequent repetitions of each stimulus pair. As in Experiment 1, we analyzed the effects of Group (patients vs. comparisons), Task version, and their interaction on item level accuracy using logistic models with random intercepts for participants. Participant group was entered as a mean-centered contrast (patients = 1.56, comparisons = -0.44). Version was entered as mean-centered Helmert contrasts. The first contrast compared scene (-0.75) with the average of shape (0.25), syllable (0.25), and tone (0.25). The second contrast compared shape (-0.625) with the average of syllable (0.375) and tone (0.375), ignoring scene (-0.125). The third contrast compared syllable (-0.5) with Tone (0.5), ignoring scene (0.0) and shape (0.0). These contrasts were maintained for all subsequent analyses. As a group, Experiment 2 participants (patients and comparisons together) demonstrated statistical learning above what would be expected by chance ($b = 1.526$, $p = 1.02e-15$). There was a significant effect of Group ($b = -0.694$, $p = .017$), with patients demonstrating worse accuracy than comparisons. There was also a significant version effect (Contrast 3; $b = -1.183$, $p = .002$), due to overall better performance on the syllables task than the tone task. The model parameters are summarized in Appendix Table B1. Planned comparisons tested these learning effects separately by group.

Analyzed separately, comparison participants demonstrated statistical learning (mean accuracy by version: shape, $M = 86.90\%$; syllable, $M = 85.71\%$; scene, $M = 81.34\%$; tone, $M = 67.86\%$; see Figure 4). Analysis of comparison participants' performance demonstrated statistical learning significantly better than chance ($b = 1.89$, $p < 2e-16$). A significant effect of version Contrast 2 (shape task vs. syllable and tone tasks; $b = -0.968$, $p = .048$) was due to better performance in the shape task compared with the syllable and tone tasks. Furthermore, a significant effect of the third version Contrast (syllable vs. tone tasks; $b = -1.402$, $p = .008$) was due to better performance in the syllable task compared with the tone task. The model parameters are summarized in Appendix Table B2.

Analyzed separately, patients with amnesia demonstrated statistical learning, although this effect was uneven across the tasks (mean accuracy by version: shape, $M = 62.50\%$; syllable, $M = 65.27\%$; scene, $M = 62.50\%$; tone, $M = 51.39\%$; see Figure 4). Patients with amnesia demonstrated statistical learning better than chance ($b = 0.435$, $p = .0196$) with a marginal effect of Contrast 3 ($b = -0.587$, $p = .089$); like healthy comparison participants, the patients performed better in the syllable task than the tone task. The model parameters are summarized in Appendix Table B3. In a post hoc analysis, there was a marginal negative

effect of task order, with poorer patient performance over time (i.e., after completing more versions; $z = -1.785$, $p = .074$).

Experiment 3

Experiment 3 consisted of an item recognition test to rule out difficulty with aspects of the task outside statistical learning. Only patients participated in Experiment 3; their data were analyzed using logistic models as before. The fixed effect of Task version was entered as mean-centered Helmert contrasts. The first contrast compared scene (-0.67) with the average of shape (0.33) and syllable (0.33). The second contrast compared shape (-0.5) with syllable (0.5), ignoring scene. Patients with amnesia demonstrated item familiarity, although this effect was uneven across the tasks (mean accuracy by version: scene, $M = 94.44\%$; shape, $M = 84.72\%$; syllable, $M = 48.61\%$; see Figure 5). Patients with amnesia demonstrated item familiarity better than chance ($b = 1.504$, $p = 1.82e-10$). A significant effect of Contrast 1 ($b = -2.013$, $p = .003$) and Contrast 2 ($b = -1.779$, $p = 1.21e-5$) was due to worse performance in the syllable task compared with the scene and shape tasks. The model parameters are summarized in Appendix Table C1.

Experiment 4: L. S. J. Triplet Assignments

Recall that, for Experiments 1 and 2 reported above, the constituent elements that made up the pairs and triplets were drawn from the same stimulus pool as in Schapiro et al., but the composition of the triplets and pairs themselves differed from those assigned to L. S. J. and her comparisons. To test whether triplet/pair assignment might contribute to differences between our study and that of Schapiro et al., patients with amnesia completed Experiments 1 and 2 a second time, utilizing the exact triple and pair assignments as reported in Schapiro et al. (2014). For Experiment 4, patients with amnesia did not demonstrate statistical learning (mean accuracy by version: shape, $M = 52.34\%$; syllable, $M = 52.34\%$; scene, $M = 51.56\%$; tone, $M = 57.03\%$; see Figure 6). Patient performance on Experiment 4 was analyzed using logistic models as before. The fixed effect of Task version was entered as mean-centered Helmert contrasts. The first contrast compared scene (-0.75) with the average of shape (0.25), syllable (0.25), and tone (0.25). The second contrast compared shape (-0.125) and scene (-0.625) with the average of syllable (0.375) and tone (0.375). The third contrast compared syllable (-0.5) with tone (0.5), ignoring scene and shape. Patients with amnesia did not demonstrate statistical learning ($b = 0.133$, $p = .133$), and there was no effect of version. The model parameters are summarized in Appendix Table D1. In a post hoc analysis, there was no significant effect of task order on patient performance ($z = -0.261$, $p = .794$).

Experiment 5: L. S. J. Pair Assignments

In Experiment 5 with the L. S. J. pair assignments, patients with amnesia demonstrated statistical learning, although performance was uneven across tasks (mean accuracy by version: shape, $M = 55.55\%$; syllable, $M = 70.83\%$; scene, $M = 58.33\%$; tone, $M = 47.22\%$; see Figure 7). Patients with amnesia demonstrated statistical learning ($b = 0.335$, $p = .0096$). A significant effect of Contrast 3 ($b = -1.000$, $p = .0044$) was due to better performance on

the syllable task compared with the tone task. The model parameters are summarized in Appendix Table D2. In a post hoc analysis, there was a significant positive effect of task order, with patient performance improving over time (i.e., after completing more versions; $z = 2.670$, $p = .008$).

Summary of Patient Performance

Across four statistical learning studies (Experiments 1, 2, 4, and 5), four task versions (syllables, tones, shapes, and scenes), and a total of 1,600 2AFC judgments, four patients with bilateral hippocampal damage averaged 57% correct. Was there evidence of new statistical learning in this patient population? Taken together, we can analyze the data set using mixed effects logistic regression from a signal detection framework to ask whether patients were able to correctly identify which of the two sequentially presented options was the target (Fraundorf, Benjamin, & Watson, 2013; DeCarlo, 2012; Wright, Horry, & Skagerberg, 2009), generalizing across experiments and task versions and explicitly modeling random variability across patients in their ability to identify the target (Snijders & Bosker, 1999).

In this post hoc analysis, we modeled the log odds of selecting the second of the two presented options in the 2AFC choice task as a function of whether the correct response was in fact the first or second option. A random slope parameter in the model allows us to model variability across patients in the ability to identify the target. The model parameters are summarized in Appendix Table E1. The intercept parameter was not significant ($z = -0.46$), indicating that patients did not have an overall bias to select one of the two options. The parameter of interest is whether presentation order of the target (first or second option) affected which option the patients selected. The target order effect was significant ($z = 3.804$, $p < .001$, likelihood ratio chi-square = 6.11, $p < .05$). The odds ratio between picking the second over the first of the top options was 1.76 greater when the target was in fact the second option (and vice versa for targets presented first). This finding demonstrates that, as a group and across all measures, patients successfully identified the target more often than chance. This model included random slopes for the target order effect, which is considered a more conservative approach when evaluating significance of fixed effects (see Barr et al., 2013). Inclusion of random slopes also enables us to model the relative ability of each patient to identify the target as a normally distributed random variable. Figure 8 plots the by-participant random effects for the target order effect along with 95% confidence intervals around each participant estimate (an estimate of 0 would indicate a complete inability to identify the target; positive values indicate successful target identification). We see that each patient shows a positive effect, with none of the confidence intervals including zero. Thus, each patient demonstrated evidence of new statistical learning.

DISCUSSION

The current study replicates the methods reported in Schapiro et al. to address the necessity of the hippocampus in statistical learning. We find that, although patients with hippocampal damage demonstrate significant learning, performing at better-than-chance levels that largely fell within the distribution of healthy participant performance, they were not as successful in

statistical learning tasks as healthy comparison participants. Regarding the necessity of the hippocampus for statistical learning, interpreted against the standard of demonstrating any evidence of new learning, we would conclude that the hippocampus contributes to statistical learning but is not strictly necessary for statistical learning to occur. Alternatively, using a definition of necessity that states performance should not differ significantly from healthy comparison participants, we would conclude that the hippocampus is necessary for achieving levels of statistical learning that are possible in healthy, nonimpaired individuals.

Our findings place hippocampus as part of a broader network of structures that work in concert to support statistical learning; this broader learning network provides a route to learning success for patients with hippocampal damage on some occasions but that, without a fully functional hippocampus, falls short of what most healthy individuals can accomplish. Thus, these data are largely consistent with previous work suggesting hippocampal contributions to statistical learning, both in patients (Schapiro et al., 2014) and in healthy adults (Turk-Browne et al., 2009). That said, the patients in the current study including the patient with an etiology and lesion most similar to L. S. J. (1951) performed notably better than L. S. J. Neuroanatomically, in Patient 1951, the entire hippocampal formation is destroyed in the right hemisphere, and in the left hemisphere, the anterior portion of the hippocampal formation is completely destroyed while 1–2 cm of the posterior hippocampus may have been spared (Feinstein et al., 2010). Furthermore, Patients 1846 and 2363 also have significant reductions in hippocampal volume and severe declarative memory impairment. Although significant variability in statistical learning tasks, both within and among participants, is well documented (Erickson et al., 2016; Arciuli et al., 2014; Misyak & Christiansen, 2012; Saffran et al., 1997), it is difficult to discern why the patients in this study performed consistently better than L. S. J.

One possibility is that some degree of residual hippocampal functioning supports the better-than-chance levels of statistical learning observed here. Indeed, we cannot rule out this interpretation as the anoxic patients have some remaining hippocampal volume and Patient 1951 has some remaining tissue in the left posterior hippocampal formation. However, it seems unlikely that such sparse remaining hippocampal tissue, particularly in the case of Patient 1951, is leveraged and recruited in service of statistical learning, when these same patients have such profound and well-documented deficits in other aspects of hippocampal-dependent learning and functioning (e.g., word learning, word list learning, paired associate learning, episodic memory). Another explanation of differences in patient performance across studies relates to differences in working memory ability. L. S. J. might have been particularly disadvantaged because of her additional impairments in working memory (e.g., L. S. J.'s WMS-WMI = 76 compared with the WMS-WMI of Group M in this study = 95 [range = 85–108]). Future work should explore this possibility. In the meantime, the data here clearly suggest that, although an intact hippocampus, working together with a larger network of brain structures outside the MTL, seems critical for statistical learning levels achieved by most healthy comparison participants, significant hippocampal pathology alone does not abolish such learning.

The learning systems engaged during statistical learning tasks may include BG circuits that operate in parallel with hippocampus in healthy participants. This proposal aligns well with

neuroimaging evidence demonstrating both hippocampal and striatal involvement in the same statistical learning task (e.g., Turk-Browne et al., 2009). In healthy individuals, different neural systems may become engaged at different points in time throughout the learning process, with decreased hippocampal involvement and increased frontal and BG activity as learning progresses (Albouy et al., 2008; Opitz & Friederici, 2003, 2004). Paradigms that allow for a continuous measure of learning over time (e.g., Karuza, Farmer, Fine, Smith, & Jaeger, 2014) may offer leverage in identifying changes in hippocampal contributions to statistical learning as the task unfolds. Relatedly, the fact that the patients in our study exhibited better performance in Experiment 2 relative to Experiment 1 may be due to the increased exposure to the stimulus targets (approximately 1.5 times the exposure compared with Experiment 1). Although patient L. S. J. exhibited relatively stable performance across Experiments 1 and 2, an open question is whether patient performance might be lifted to the level of healthy comparison (or moved above chance in the case of patients like L. S. J.) if stimulus exposure was increased even further. Indeed, increased exposure to the stimulus stream has been demonstrated to boost performance in other clinical populations (e.g., Evans et al., 2009).

Although individual patients performed at above-chance levels on some versions of Experiments 1 and 2, as a group, patients consistently performed more poorly than comparison participants. One possibility is that patient performance was poorer than healthy comparisons because they completed multiple versions of similar statistical learning tasks (whereas healthy comparisons completed only one), resulting in interference effects. Analysis of order effects suggests that this is unlikely to be the case, as significant order effects were found only in Experiment 5, and in this case, the effect was positive. Focal hippocampal damage appears to be sufficient to disrupt statistical learning ability, as evidenced by a lack of difference in performance between Patient 1951 (with a more extensive MTL damage) and the anoxic patients. In fact, Patient 1951 demonstrated the highest mean performance of any of the patients on 5 of the 16 experimental versions (Experiment 1: shapes and syllables, Experiment 4: scenes and syllables, Experiment 5: shapes). This performance is particularly striking given Patient 1951's consistently poorer performance on Experiment 3. Patient 1951's poorer performance relative to anoxic patients on this item recognition test is expected, given the known role of MTL structures outside the hippocampus in supporting item memory (Eichenbaum, Yonelinas, & Ranganath, 2007; Mayes, Holdstock, Isaac, Hunkin, & Roberts, 2002). It is also notable that patients as a group performed more poorly on the syllable recognition task compared with shape and scene versions. This may be attributed to the difference in stimulus presentation time, with syllable presentation times (0.25 sec) significantly more transient than shapes/scenes (0.5 sec). In any case, the patients' consistently high performance on the shapes and scenes versions of Experiment 3 suggests that difficulties with task demands such as attending to the exposure phase, following task instructions, and making familiarity judgments are not the cause of impaired statistical learning performance.

We also note that, across all tasks and all participants, performance was highly variable. This performance variability in statistical learning tasks has been noted in recent work (Erickson et al., 2016; Siegelman & Frost, 2015) and points to the importance of testing a variety of stimuli and stimulus combinations. We observed striking differences in patient performance

depending on the assignment of items to triplets and pairs; notably, patient performance was poorer on the triplet assignments that had been tested with patient L. S. J. (Experiment 4). Critically, whereas these caveats point to limits in the utility of these tasks for characterizing individual ability, group level comparisons point to a clear role for the hippocampus in the neural network for statistical learning.

Situating impaired statistical learning performance within the extant literature requires explanation of why patients with amnesia demonstrate intact AGL, a task that is similar to canonical statistical learning paradigms in many ways (Knowlton, Ramus, & Squire, 1992; also see Schapiro et al., 2014, for more discussion). The present task differs from AGL in two key ways. First, stimuli in AGL are constructed using a Markovian finite-state grammar (Reber, 1967). This underlying grammar allows for abstraction of grammatical rules that can be leveraged to identify trained associations and to generalize to novel stimuli (Altmann, Dienes, & Goode, 1995; Mathews et al., 1989; but see Redington & Chater, 1996). This differs from canonical statistical learning tasks, which do not involve generalization. Second, training during AGL involves already segmented input, in contrast to the continuous input provided to patients during typical statistical learning tasks. Despite its name, AGL has rarely been linked to natural language processing, in contrast to statistical learning, which has been linked to language learning and processing across multiple levels of complexity (e.g., Misyak & Christiansen, 2012; Evans et al., 2009; Pacton et al., 2005; Chambers et al., 2003). Impairments in grammatical processing in patients with amnesia may therefore be understudied, based on the assumption that intact performance on AGL tasks carries over to natural language processing.

Links between impairment in statistical learning and hippocampal damage open the door to more careful exploration of hippocampal contributions to grammatical processing and may contribute to the growing body of work demonstrating hippocampal involvement in language use and processing (Duff & Brown-Schmidt, 2012, 2017). For example, individuals with developmental language impairment, a disorder characterized by deficits in morphosyntactic components of language that are thought to be acquired through statistical learning (Evans et al., 2009), demonstrate significant correlations between hippocampal volume and language measures in this population (Lee, Nopoulous, & Tomblin, 2013). These data, together with work by Turk-Browne and colleagues as well as the work presented here, provide a challenge to proposed divisions of labor for hippocampus (vocabulary) and BG (grammar) in language processing (Ullman, 2004). A link between statistical learning and hippocampus also suggests that more careful characterization of grammatical learning and processing in developmental amnesia may be warranted. Although these individuals exhibit average to low-average performance on standardized language measures (Vargha-Khadem et al., 1997), the impact of developmental amnesia on more complex and situated language abilities deserves scrutiny. Taken all together, reconsideration of the role of the hippocampus in grammatical processing more broadly is warranted.

Conclusion

The performances of four individuals with bilateral hippocampal damage and dense amnesia point to a clear role for hippocampus in statistical learning. However, statistical learning was

not uniformly disrupted across all task versions in patients with hippocampal damage, and in some cases, patients with amnesia demonstrated significant learning. The findings here align with prior neuroimaging studies that implicate a network of brain structures in statistical learning (Karuza et al., 2013; Turk-Browne et al., 2009; McNealy et al., 2006; Bischoff-Grethe et al., 2000). This interpretation fits well with a broader move away from attributing a single cognitive ability to a single neural system (e.g., the hippocampus is solely responsible for episodic memory) and toward understanding neural structures situated in larger neural networks and contributing to more than one cognitive ability, based on overlapping processing demands (e.g., hippocampus supports memory but also meets some of the demands of language, empathy, social cognition, creativity, etc.; Rubin, Watson, Duff, & Cohen, 2014). In terms of necessity, the hippocampus appears necessary for reaching statistical learning levels achieved by most healthy comparison participants; however, significant hippocampal pathology alone does not abolish such learning.

Acknowledgments

This work was supported by NIDCD Grant R01 DC011755 to M. C. D. and S. B. S. We thank Sharice Clough for her help during data collection. We thank Anna Schapiro and the Turk-Browne laboratory for their willingness to share experimental materials and for valuable discussion.

References

- Albouy G, Sterpenich V, Baiteau E, Vandewalle G, Desseilles M, Dang-Vu T, et al. Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron*. 2008; 58:261–272. [PubMed: 18439410]
- Allen JS, Tranel D, Bruss J, Damasio H. Correlations between regional brain volumes and memory performance in anoxia. *Journal of Clinical and Experimental Neuropsychology*. 2006; 28:457–476. [PubMed: 16624778]
- Altmann GTM, Dienes Z, Goode A. Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21:899–912.
- Arciuli J, Torkildsen JVK, Stevens DJ, Simpson IC. Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*. 2014; 5:747. [PubMed: 25071692]
- Baldwin D, Andersson A, Saffran J, Meyer M. Segmenting dynamic human action via statistical structure. *Cognition*. 2008; 106:1382–1407. [PubMed: 18035346]
- Bischoff-Grethe A, Proper SM, Mao H, Daniels KA, Berns GS. Conscious and unconscious processing of nonverbal predictability in Wernicke’s area. *Journal of Neuroscience*. 2000; 20:1975–1981. [PubMed: 10684898]
- Brady TF, Oliva A. Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*. 2008; 19:678–685. [PubMed: 18727783]
- Brainard DH. The Psychophysics Toolbox. *Spatial Vision*. 1997; 10:433–436. [PubMed: 9176952]
- Buchanan TW, Tranel D, Adolphs R. Emotional autobiographical memories in amnesic patients with medial temporal lobe damage. *Journal of Neuroscience*. 2005; 25:3151–3160. [PubMed: 15788772]
- Chambers KE, Onishi KH, Fisher C. Infants learn phonotactic regularities from brief auditory experience. *Cognition*. 2003; 87:B69–B77. [PubMed: 12590043]
- Channon S, Shanks D, Johnstone T, Vakili K, Chin J, Sinclair E. Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia*. 2002; 40:2185–2197. [PubMed: 12208014]
- Conway CM, Christiansen MH. Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31:24–39.

- Creel SC, Newport EL, Aslin RN. Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2004; 30:1119–1130.
- DeCarlo LT. On a signal detection approach to m -alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *Journal of Mathematical Psychology*. 2012; 56:196–207.
- Dede AJO, Frascino JC, Wixted JT, Squire LR. Learning and remembering real-world events after medial temporal lobe damage. *Proceedings of the National Academy of Sciences, USA*. 2016; 113:13480–13485.
- Duff MC, Brown-Schmidt S. The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience*. 2012; 6:69. [PubMed: 22493573]
- Duff MC, Brown-Schmidt S. Hippocampal contributions to language use and processing. *The hippocampus from cells to systems*. 2017:503–536.
- Eichenbaum, H., Cohen, NJ. *From conditioning to conscious recollection: Memory systems of the brain*. New York: Oxford University Press; 2001.
- Eichenbaum H, Yonelinas AP, Ranganath C. The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*. 2007; 30:123–152.
- Erickson LC, Kaschak MP, Thiessen ED, Berry CAS. Individual differences in statistical learning: Conceptual and measurement issues. *Collabra*. 2016; 2:14.
- Evans JL, Saffran JR, Robe-Torres K. Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2009; 52:321–335.
- Feinstein JS, Rudrauf D, Khalsa SS, Cassell MD, Bruss J, Grabowski TJ, et al. Bilateral limbic system destruction in man. *Journal of Clinical and Experimental Neuropsychology*. 2010; 32:88–106. [PubMed: 19763994]
- Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*. 2001; 12:499–504. [PubMed: 11760138]
- Fiser J, Aslin RN. Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28:458–467.
- Fletcher P, Büchel C, Josephs O, Friston K, Dolan R. Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex*. 1999; 9:168–178. [PubMed: 10220229]
- Fraundorf SH, Benjamin AS, Watson DG. What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language*. 2013; 69:196–227. [PubMed: 24014934]
- Gómez RL. Variability and detection of invariant structure. *Psychological Science*. 2002; 13:431–436. [PubMed: 12219809]
- Graf Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*. 2007; 18:254–260. [PubMed: 17444923]
- Gregory E, McCloskey M, Landau B. Profound loss of general knowledge in retrograde amnesia: Evidence from an amnesic artist. *Frontiers in Human Neuroscience*. 2014; 8:287. [PubMed: 24834048]
- Hannula DE, Greene AJ. The hippocampus reevaluated in unconscious learning and memory: At a tipping point? *Frontiers in Human Neuroscience*. 2012; 6:80. [PubMed: 22518102]
- Hannula DE, Tranel D, Cohen NJ. The long and the short of it: Relational memory impairments in amnesia, even at short lags. *Journal of Neuroscience*. 2006; 26:8352–8359. [PubMed: 16899730]
- Karuzza EA, Farmer TA, Fine AB, Smith FX, Jaeger TF. On-line measures of prediction in a self-paced statistical learning task. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*; 2014. p. 725-730.
- Karuzza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D. The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*. 2013; 127:46–54. [PubMed: 23312790]
- Kim R, Seitz A, Feenstra H, Shams L. Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*. 2009; 461:145–149. [PubMed: 19539701]

- Knowlton BJ, Ramus SJ, Squire LR. Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*. 1992; 3:172–179.
- Lee JC, Nopoulos PC, Tomblin JB. Abnormal subcortical components of the corticostriatal system in young adults with DLI: A combined structural MRI and DTI study. *Neuropsychologia*. 2013; 51:2154–2161. [PubMed: 23896446]
- Lieberman MD, Chang GY, Chiao J, Bookheimer SY, Knowlton BJ. An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*. 2004; 16:427–438. [PubMed: 15072678]
- Mathews RC, Buss RR, Stanley WB, Blanchard-Fields F, Cho JR, Druhan B. Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1989; 15:1083–1100.
- Mayes AR, Holdstock JS, Isaac CL, Hunkin NM, Roberts N. Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. *Hippocampus*. 2002; 12:325–340. [PubMed: 12099484]
- McNealy K, Mazziotta JC, Dapretto M. Cracking the language code: Neural mechanisms underlying speech parsing. *Journal of Neuroscience*. 2006; 26:7629–7639. [PubMed: 16855090]
- Misyak JB, Christiansen MH. Statistical learning and language: An individual differences study. *Language Learning*. 2012; 62:302–331.
- O’Kane G, Kensinger EA, Corkin S. Evidence for semantic learning in profound amnesia: An investigation with patient H. M. *Hippocampus*. 2004; 14:417–425. [PubMed: 15224979]
- Opitz B, Friederici AD. Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage*. 2003; 19:1730–1737. [PubMed: 12948727]
- Opitz B, Friederici AD. Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *Journal of Neuroscience*. 2004; 24:8436–8440. [PubMed: 15456816]
- Pacton S, Fayol M, Perruchet P. Children’s implicit learning of graphotactic and morphological regularities. *Child Development*. 2005; 76:324–339. [PubMed: 15784085]
- Perruchet P, Pacton S. Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*. 2006; 10:233–238. [PubMed: 16616590]
- Quené H, van den Bergh H. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*. 2008; 59:413–425.
- Reber AS. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*. 1967; 6:855–863.
- Redington M, Chater N. Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*. 1996; 125:123–138.
- Rubin RD, Watson PD, Duff MC, Cohen NJ. The role of the hippocampus in flexible cognition and social behavior. *Frontiers in Human Neuroscience*. 2014; 8:742. [PubMed: 25324753]
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274:1926–1928. [PubMed: 8943209]
- Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. *Cognition*. 1999; 70:27–52. [PubMed: 10193055]
- Saffran JR, Newport EL, Aslin RN, Tunick RA, Barrueco S. Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*. 1997; 8:101–105.
- Saffran JR, Wilson DP. From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*. 2003; 4:273–284.
- Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB. The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*. 2014; 26:1736–1747. [PubMed: 24456393]
- Schlichting ML, Guarino KF, Schapiro AC, Turk-Browne NB, Preston AR. Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience*. 2017; 29:37–51. [PubMed: 27575916]

- Seger CA, Prabhakaran V, Poldrack RA, Gabrieli JDE. Neural activity differs between explicit and implicit learning of artificial grammar strings: An fMRI study. *Psychobiology*. 2000; 28:283–292.
- Siegelman N, Frost R. Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*. 2015; 81:105–120. [PubMed: 25821343]
- Squire LR, Zola SM. Episodic memory, semantic memory, and amnesia. *Hippocampus*. 1998; 8:205–211. [PubMed: 9662135]
- Toro JM, Sinnott S, Soto-Faraco S. Speech segmentation by statistical learning depends on attention. *Cognition*. 2005; 97:B25–B34. [PubMed: 16226557]
- Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*. 2009; 21:1934–1945. [PubMed: 18823241]
- Ullman MT. Contributions of memory circuits to language: The declarative/procedural model. *Cognition*. 2004; 92:231–270. [PubMed: 15037131]
- Vargha-Khadem F, Gadian DG, Watkins KE, Connelly A, Van Paesschen W, Mishkin M. Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*. 1997; 277:376–380. [PubMed: 9219696]
- Wright DB, Horry R, Skagerberg EM. Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods*. 2009; 41:257–267. [PubMed: 19363166]

APPENDIX

Table A1

Experiment 1: Estimated Parameters for the Logistic Regression Model Predicting Item Level Accuracy (Full Data Set)

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	0.92228	0.13540	6.812	9.64e-12
Group (patients = 1.56, comparisons = -0.44)	-0.38095	0.21940	-1.736	.0825
Version contrast 1 (scene = -0.75 vs. shape/syllable/tone = 0.25)	-0.04572	0.23244	-0.197	.8441
Version contrast 2 (shape = -0.625 vs. syllable/tone = 0.375; scene = -0.125)	0.07279	0.24223	0.301	.7638
Version contrast 3 (syllable = -0.5 vs. tone = 0.5; scene/shape = 0)	0.30714	0.27826	1.104	.2697
Group × Contrast 1	0.01068	0.18055	0.059	.9528
Group × Contrast 2	0.44835	0.18908	2.371	.0177
Group × Contrast 3	0.34289	0.22058	1.554	.1201
<i>Random Effects</i>		<i>Variance</i>	<i>SD</i>	
Participant (intercept)	0.6726	0.8201		

Observations = 2,304, Participants = 60

Table A2

Experiment 1: Estimated Parameters for Logistic Regression Model of Item Level Accuracy (Comparison Data Only)

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	1.10339	0.13341	8.271	<2e-16
Version contrast 1 (scene = -0.75 vs. shape/syllable/tone = 0.25)	-0.05447	0.31013	-0.176	.861
Version contrast 2 (shape = -0.625 vs. syllable/tone = 0.375; scene = -0.125)	-0.13204	0.32297	-0.409	.683

Fixed Effects	Estimate	SE	z Value	p Value
Version contrast 3 (syllable = -0.5 vs. tone = 0.5; scene/shape = 0)	0.15761	0.37017	0.426	.670
<i>Random Effects</i>				
	<i>Variance</i>	<i>SD</i>		
Participant	0.7744	0.88		
Observations = 1,792, Participants = 56				

Table A3

Experiment 1: Estimated Parameters for Logistic Regression Model of Item Level Accuracy (Patient Data Only)

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	0.32493	0.09178	3.541	.000399
Version contrast 1 (scene = -0.75 vs. shape/syllable/tone = 0.25)	-0.02835	0.20995	-0.135	.892578
Version contrast 2 (shape = -0.625 vs. syllable/tone = 0.375; scene = -0.125)	0.75906	0.22212	3.417	.000632
Version contrast 3 (syllable = -0.5 vs. tone = 0.5; scene/shape = 0)	0.82888	0.26754	3.098	.001947
<i>Random Effects</i>				
	<i>Variance</i>	<i>SD</i>		
Participant	8.568e-13	9.257e-7		
Observations = 512, Participants = 4				

Table B1

Experiment 2: Estimated Parameters for Logistic Regression Model of Item Level Accuracy (Full Data Set)

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	1.52579	0.19014	8.025	1.02e-15
Group	-0.69432	0.29031	-2.392	.01677
Version contrast 1 (scene = -0.75 vs. shape/syllable/tone = 0.25)	0.03022	0.32754	0.092	.92648
Version contrast 2 (shape = -0.625 vs. syllable/tone = 0.375; scene = -0.125)	-0.75261	0.35263	-2.134	.03282
Version contrast 3 (syllable = -0.5 vs. tone = 0.5; scene/shape = 0)	-1.18341	0.38426	-3.080	.00207
Group × Contrast 1	-0.07563	0.25257	-0.299	.76460
Group × Contrast 2	0.37129	0.26933	1.379	.16802
Group × Contrast 3	0.37423	0.29831	1.254	.20966
<i>Random Effects</i>				
	<i>Variance</i>	<i>SD</i>		
Participant	1.147	1.071		
Observations = 1,296, Participants = 60				

Table B2

Experiment 2: Estimated Parameters for Logistic Regression Model of Item Level Accuracy (Comparison Data Only)

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	1.88528	0.20897	9.022	<2e-16
Version contrast 1 (scene = -0.75 vs. shape/syllable/tone = 0.25)	0.06406	0.45213	0.142	.88733
Version contrast 2 (shape = -0.625 vs. syllable/tone = 0.375; scene = -0.125)	-0.96775	0.48965	-1.976	.04811
Version contrast 3 (syllable = -0.5 vs. tone = 0.5; scene/shape = 0)	-1.40278	0.53000	-2.647	.00813
<i>Random Effects</i>		<i>Variance</i>	<i>SD</i>	
Participant	1.399	1.183		
Observations = 1,008, Participants = 56				

Table B3

Experiment 2: Estimated Parameters for Logistic Regression Model of Item Level Accuracy (Patient Data Only)

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	0.43518	0.18648	2.334	.0196
Version contrast 1 (scene = -0.75 vs. shape/syllable/tone = 0.25)	-0.08534	0.28782	-0.296	.7669
Version contrast 2 (shape = -0.625 vs. syllable/tone = 0.375; scene = -0.125)	-0.17069	0.30014	-0.569	.5696
Version contrast 3 (syllable = -0.5 vs. tone = 0.5; scene/shape = 0)	-0.58663	0.34516	-1.700	.0892
<i>Random Effects</i>		<i>Variance</i>	<i>SD</i>	
Participant	0.07892	0.2809		
Observations = 288, Participants = 4				

Table C1

Experiment 3: Estimated Parameters for Logistic Regression Model of Item Level Accuracy

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	1.5044	0.2359	6.376	1.82e-10
Version contrast 1 (scene = -0.67 vs. shape/syllable = 0.33)	-2.0126	0.5539	-3.633	.00028
Version contrast 2 (syllable = -0.5 vs. shape = 0.5; scene = 0)	-1.7790	0.4065	-4.376	1.21e-5
<i>Random Effects</i>		<i>Variance</i>	<i>SD</i>	
Participant	0.02894	0.1701		
Observations = 216, Participants = 4				

Table D1

Experiment 4 With Exact L. S. J. Triplet Assignments: Estimated Parameters for Logistic Regression Model of Item Level Accuracy

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	0.13332	0.08867	1.504	.133
Version contrast 1 (scene = -0.75 vs. shape/syllable/tone = 0.25)	0.07863	0.20751	0.379	.705
Version contrast 2 (shape = -0.625 vs. syllable/tone = 0.375; scene = -0.125)	0.09465	0.21707	0.436	.663
Version contrast 3 (syllable = -0.5 vs. tone = 0.5; scene/shape = 0)	0.18931	0.25139	0.753	.451
Random Effects				
	<i>Variance</i>	<i>SD</i>		
Participant	0	0		
Observations = 512, Participants = 4				

Table D2

Experiment 5 With Exact L. S. J. Pair Assignments: Estimated Parameters for Logistic Regression Model of Item Level Accuracy

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (performance compared to chance)	0.33452	0.12919	2.589	.00962
Version contrast 1 (scene = -0.75 vs. shape/syllable/tone = 0.25)	-0.03096	0.28113	-0.110	.91229
Version contrast 2 (shape = -0.625 vs. syllable/tone = 0.375; scene = -0.125)	0.16515	0.29521	0.559	.57586
Version contrast 3 (syllable = -0.5 vs. tone = 0.5; scene/shape = 0)	-1.00030	0.35112	-2.849	.00439
Random Effects				
	<i>Variance</i>	<i>SD</i>		
Participant	0.007517	0.0867		
Observations = 288, Participants = 4				

Table E1

All Patient Data, 1,600 Binary Responses From Four Patients, Four Experiments, and Four Task Versions

Fixed Effects	Estimate	SE	z Value	p Value
Intercept (bias to select second over first option)	-0.09035	0.19852	-0.455	.649017
Target order (effect of target order on response)	0.56480	0.14848	3.804	.000142
Random Effects				
	<i>Variance</i>	<i>SD</i>	<i>Corr.</i>	
Participant intercepts	0.08581	0.2929		
Participant slopes for the target order effect	0.04596	0.2144	0.47	
Experiment intercepts	0.03159	0.1777		
Version intercepts	0.02883	0.1698		

Estimated parameters for logistic mixed effects model of the effect of target order (first vs. second option) on order response (the second option is coded as 1; the first option is coded as 0). Corr. = correlation between random intercepts and random slopes.

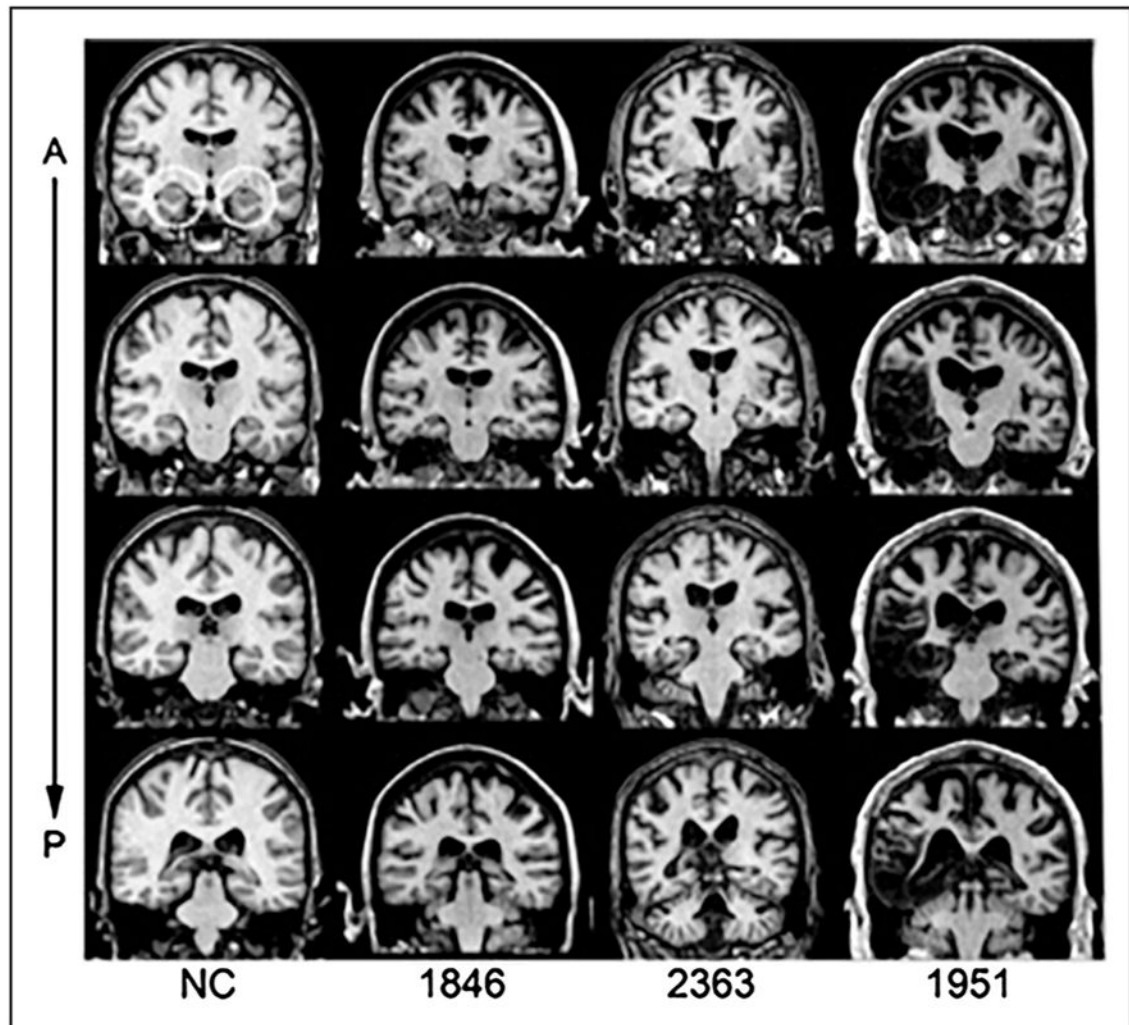


Figure 1. Magnetic resonance scans of hippocampal patients. Images are coronal slices through four points along the hippocampus from T1-weighted scans. Volume changes can be noted in the hippocampal region for Patients 1846 and 2363, and significant bilateral MTL damage including the hippocampus can be noted in Patient 1951. A = anterior; P = posterior; NC = healthy comparison brain.

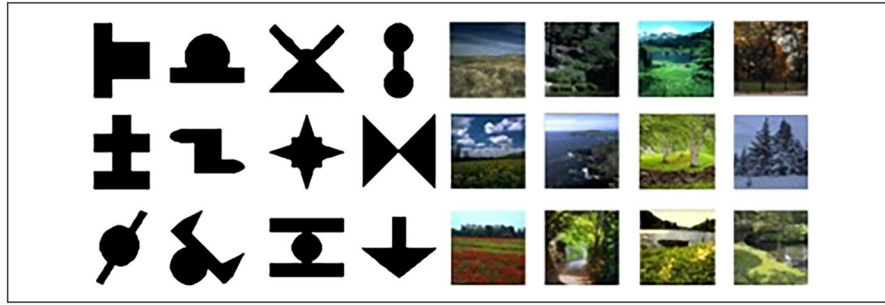


Figure 2.
Shape and scene stimuli (tone and syllable stimuli described in the text).

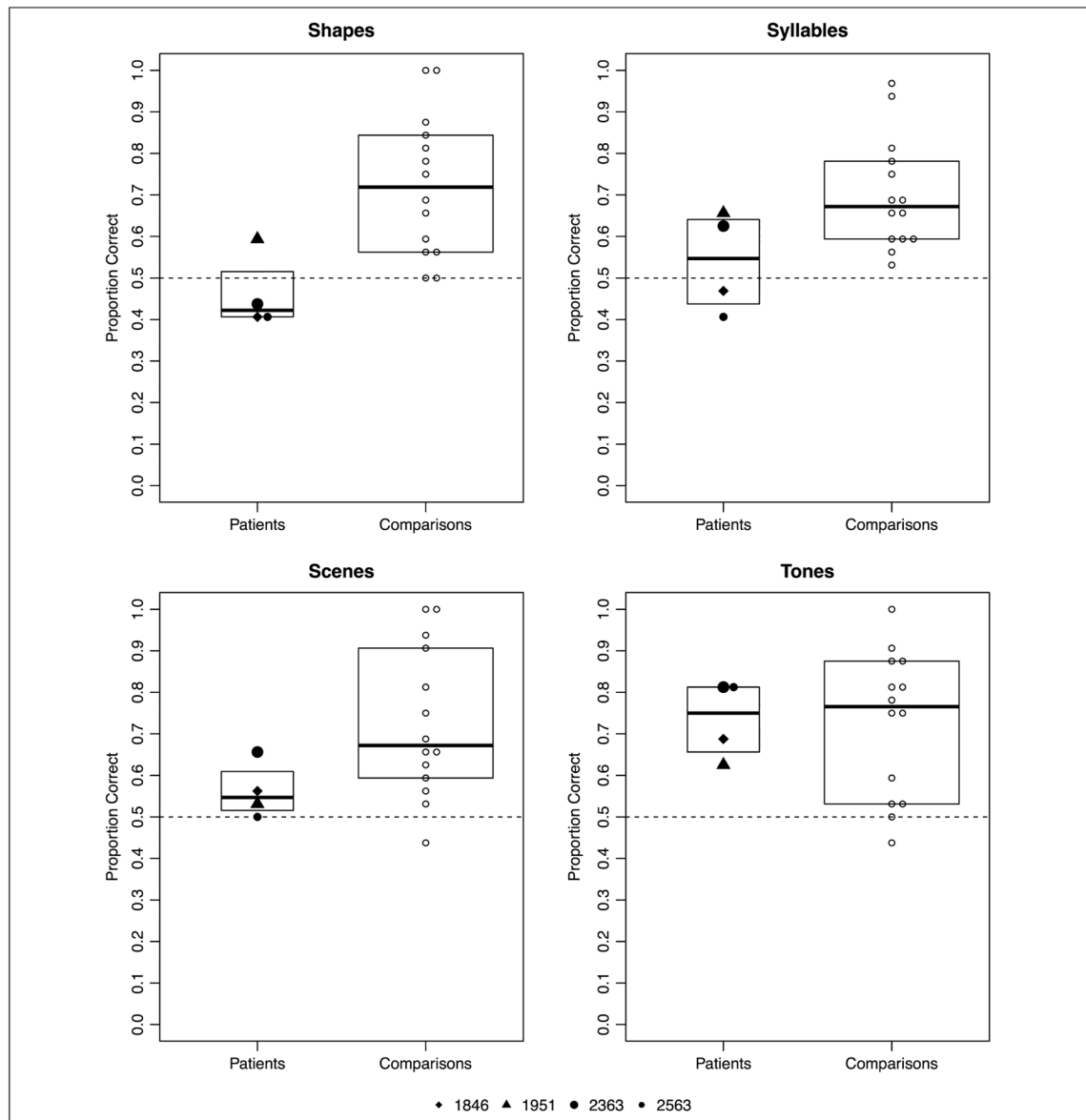


Figure 3. Experiment 1: proportion correct across four versions of the statistical learning task by patients with amnesia and healthy comparison participants. Boxplots indicate the group level IQR.

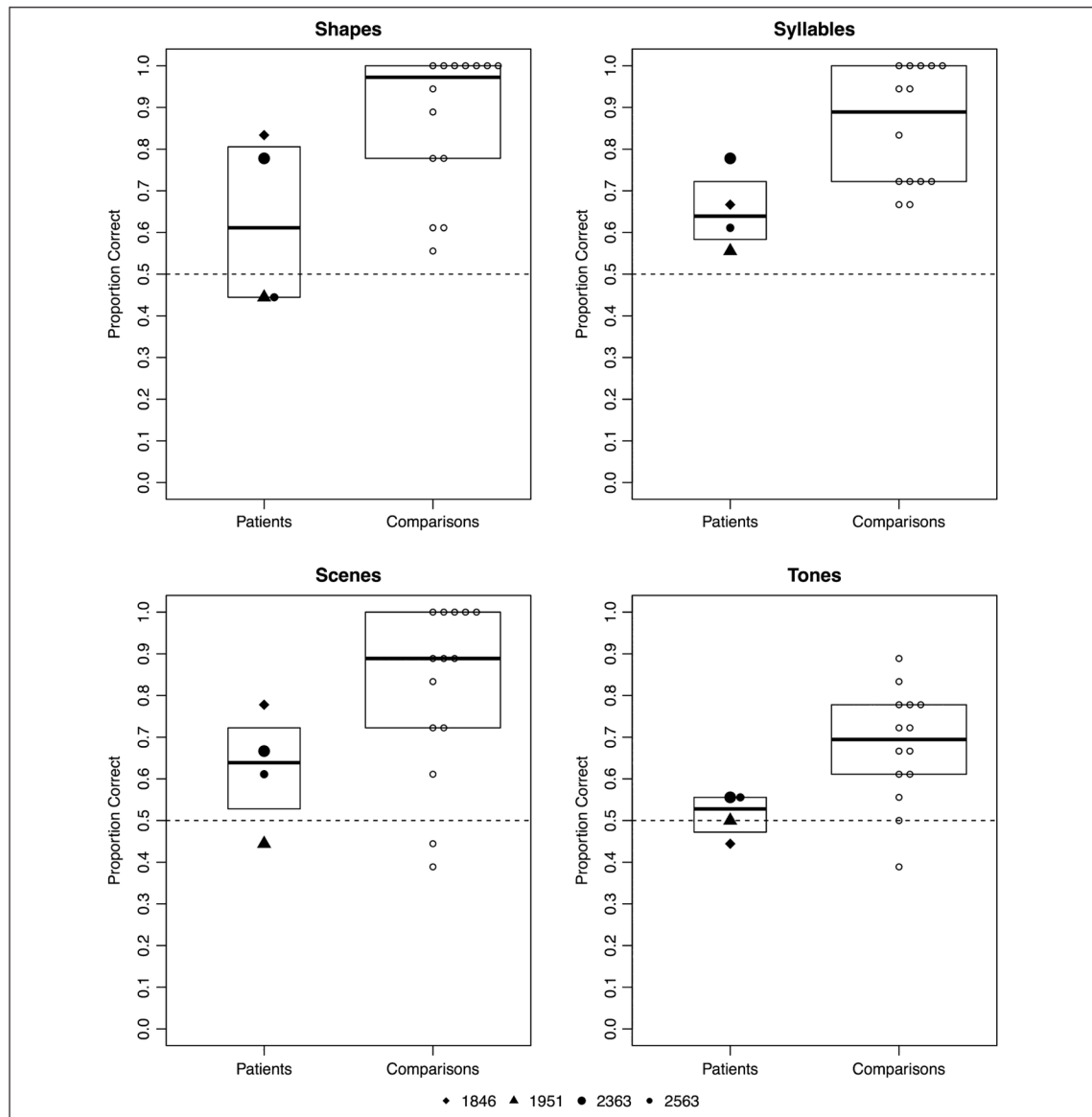


Figure 4. Experiment 2: proportion correct across four versions of the statistical learning task by patients with amnesia and healthy comparison participants. Boxplots indicate the group level IQR.

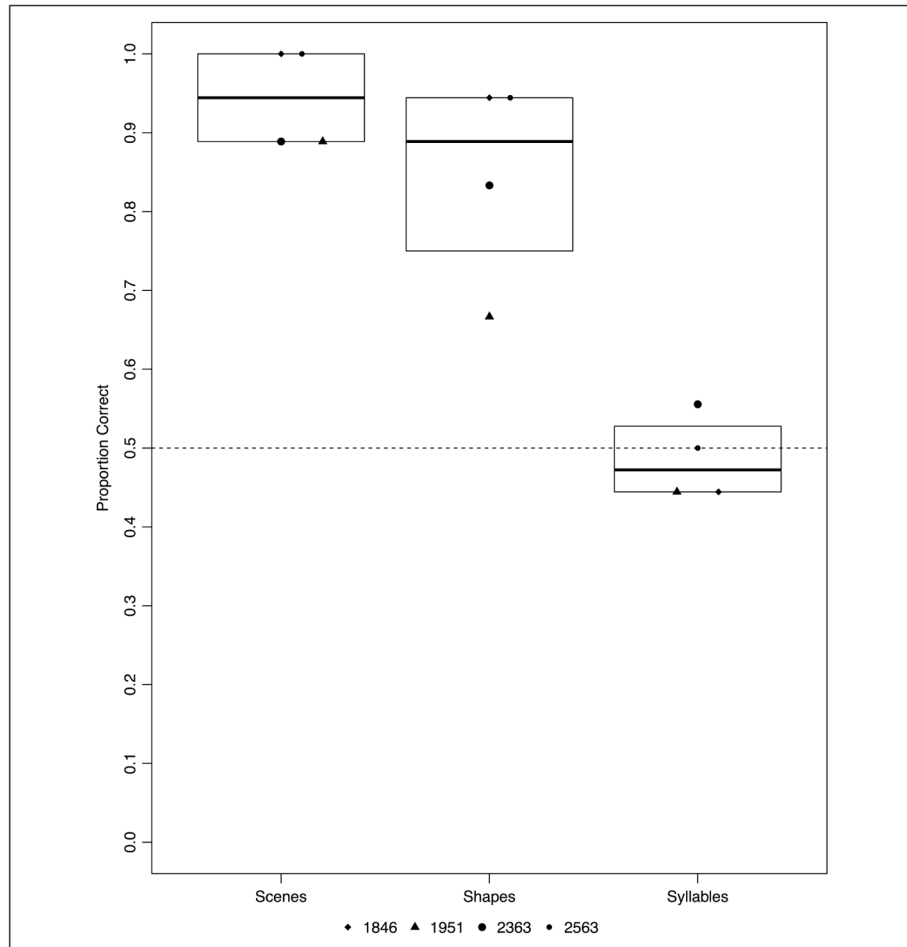


Figure 5. Experiment 3: proportion correct across three versions of the item recognition task by patients with amnesia. Boxplots indicate the group level IQR.

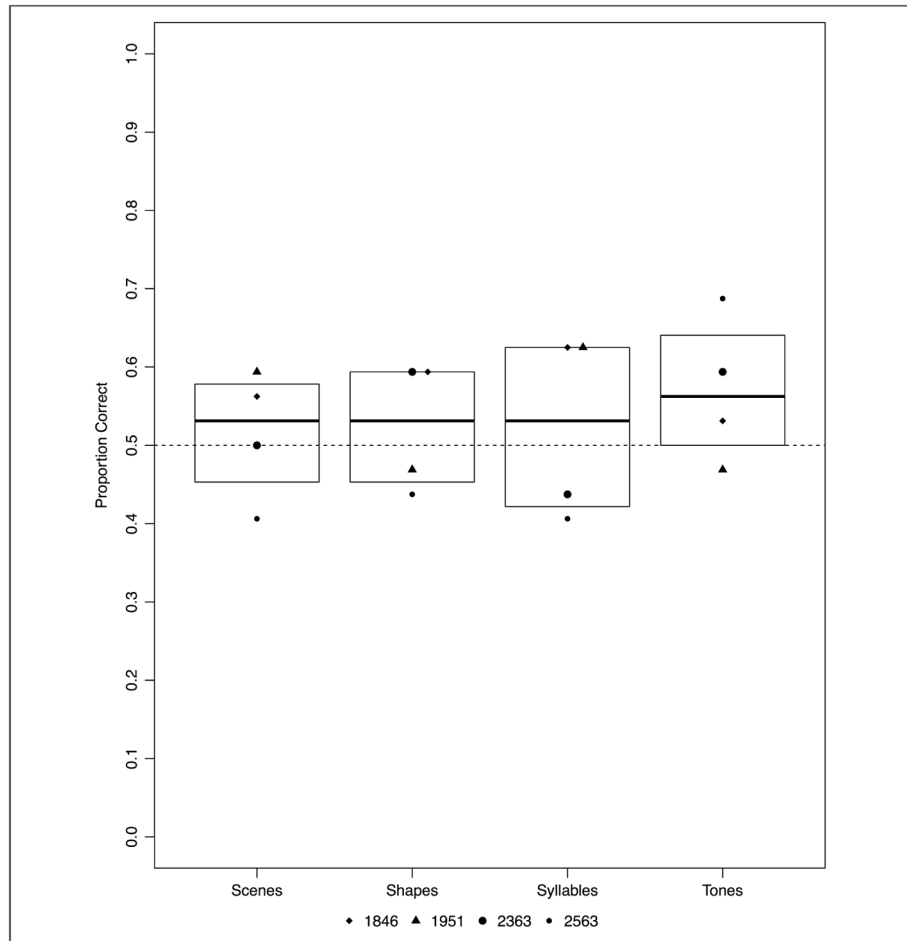


Figure 6. Experiment 4: proportion correct across four versions of the statistical learning task with L. S. J. triplet assignment by patients with amnesia. Boxplots indicate the group level IQR.

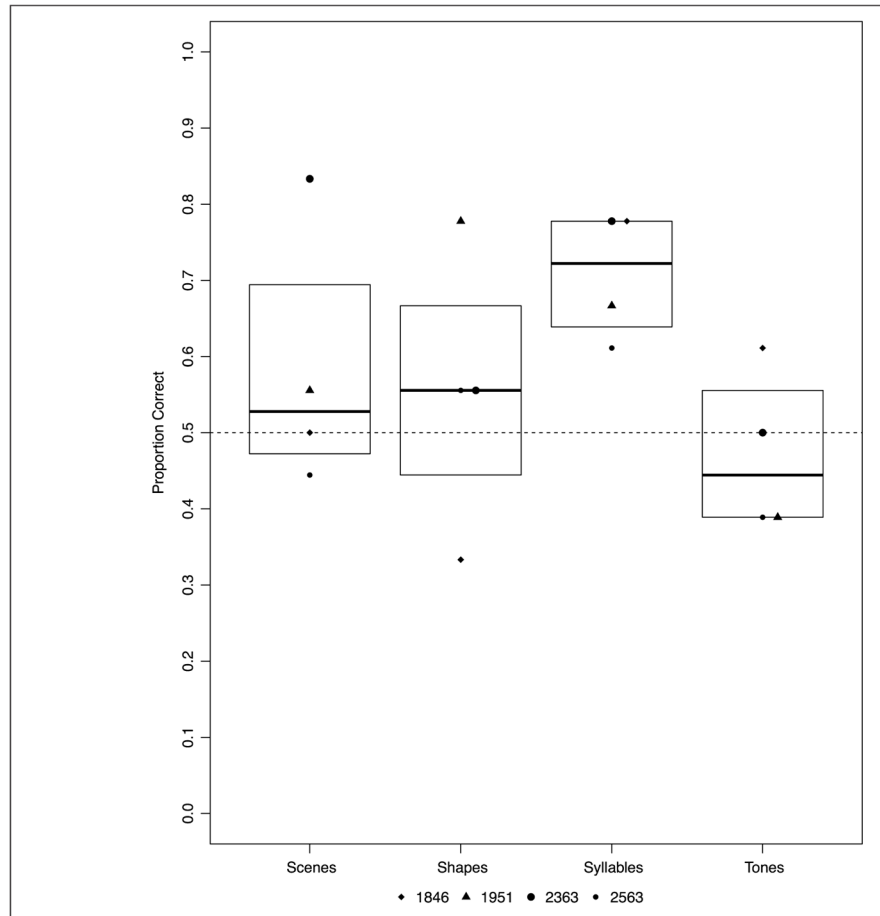


Figure 7. Experiment 5: proportion correct across four versions of the statistical learning task with L. S. J. pair assignment by patients with amnesia. Boxplots indicate the group level IQR.

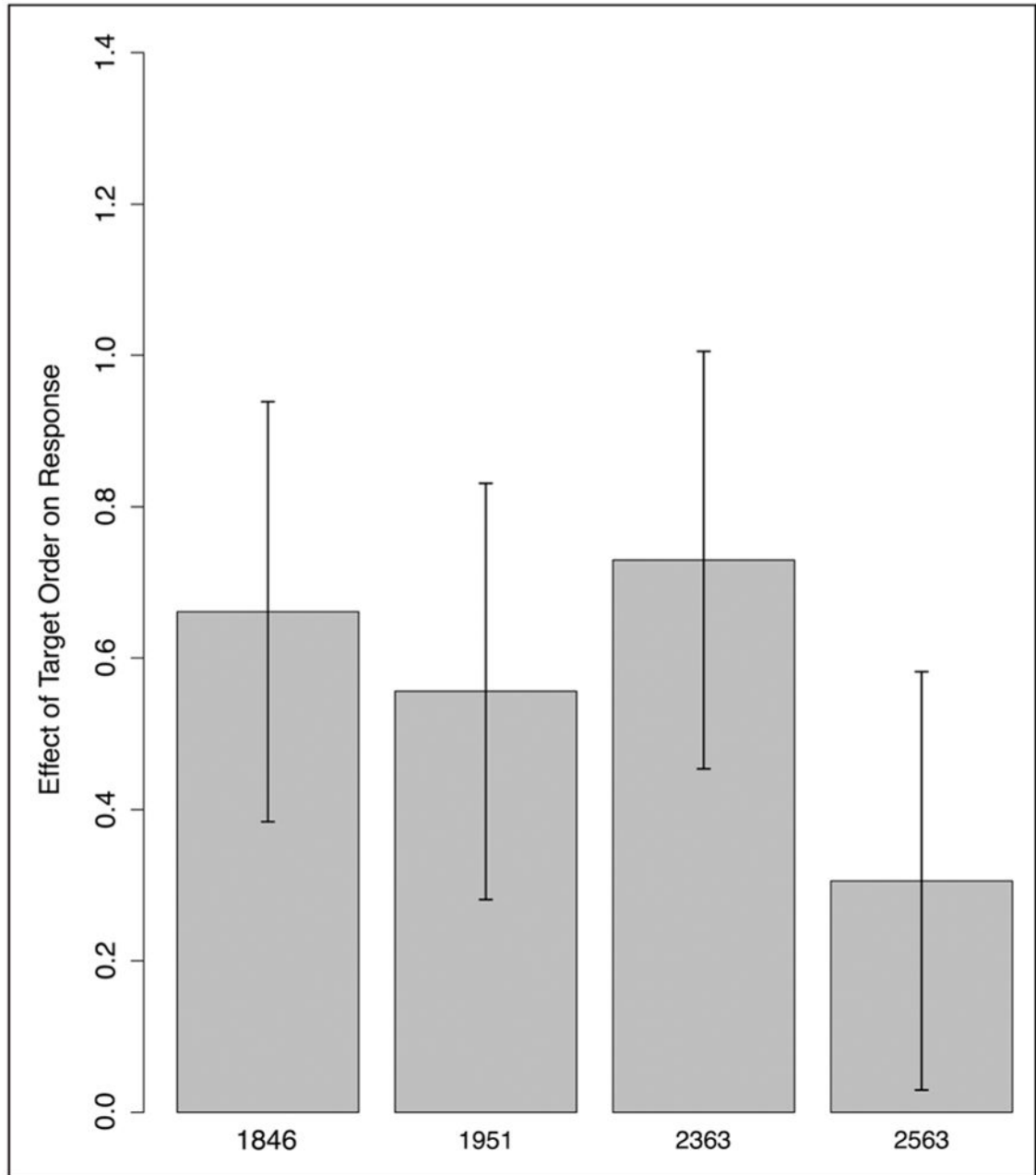


Figure 8. Model-based estimate of by-participant random effects. Patient data across Experiments 1, 2, 4, and 5. Error bars indicate 95% confidence intervals.

Table 1
Demographic, Neuroanatomical, and Neuropsychological Characteristics of Participants With Hippocampal Amnesia

Participant	Demographic					Lesion	HC Volume	Anatomical					Neuropsychological		
	Gender	Age	Hand	Ed	Etiology			WAIS III FSIQ	WMS III GMI	WMS III WMI	BNT	TT	CFT Copy	WCS Cat	
1846	F	52	R	14	Anoxia	Bilateral HC	-4.23	84	57	85	43	41	28	6	
1951	M	63	R	16	HSE	Bilateral HC + MTL	-8.10	106	57	108	49	44	32	6	
2363	M	59	R	18	Anoxia	Bilateral HC	-2.64	98	73	88	58	44	26	6	
1563	M	60	L	16	Anoxia	Bilateral HC	N/A	102	75	99	52	44	36	6	
Mean		58.5		16				97.5	65.5	95.0	50.5	43.3	30.5	6	

Bolded scores indicate 2 *SDs* below population norms. Hand = handedness; Ed = years of education; HC = hippocampus; +MTL = damage extending into the greater MTLs; HC Volume = hippocampal volumetric *z* scores as measured through high-resolution volumetric MRI and compared with a matched healthy comparison group (Allen et al., 2006; Buchanan et al., 2005); BNT = Boston Naming Test; TT = Token Test; CFT = Complex Figure Test; WCT = Wisconsin Card Sorting Task; Cat = number of categories achieved out of six.