Statistical Learning

A Schapiro and N Turk-Browne, Princeton University, Princeton, NJ, USA

Glossary

Entropy Measure of uncertainty or unpredictability.
Grammar A set of rules about how to generate a sequence of items.
Medial temporal lobe Brain structure consisting of the hippocampus and entorhinal, perirhinal, and parahippocampal cortices.

Regularities Recurring patterns in sensory input over time.
Segmentation Process of breaking up continuous information into components.
Striatum Brain structure in the basal ganglia that includes the caudate nucleus and putamen.
Transition probability Likelihood that one item occurs immediately after another in a sequence.

Statistical learning refers to the ability to extract regularities from the environment over time. Sensitivity to environmental statistics is a pervasive property of the brain that operates over ontogenetic (Berkes, Orban, Lengyel, & Fiser, 2011; Li, Piech, & Gilbert, 2004) and phylogenetic (Purves, Wojtach, & Lotto, 2011; Simoncelli & Olshausen, 2001) timescales. However, the term 'statistical learning' is often used in contexts in which learning is rapid – on the order of minutes (Aslin & Newport, 2012; Saffran, Aslin, & Newport, 1996; Thiessen, Kronstein, & Hufnagle, 2013). In addition, we will use this term to refer to automatic and implicit forms of learning, as opposed to conscious or intentional discovery of regularities (Destrebecqz & Cleeremans, 2001). There are other uses of this term, such as in computer science (Vapnik, 1998), which are related but not the focus of this article.

As implied by the broad meaning of this term, statistical learning is a ubiquitous process that manifests in multiple aspects of behavior. Studies of statistical learning have relied on many different types of dependent measures, such as recognition of regularities (e.g., Fiser & Aslin, 2001; Reber, 1967), facilitated response times for predictable stimuli (e.g., Kim, Seitz, Feenstra, & Shams, 2009; Nissen & Bullemer, 1987), and guidance of attention based on learned associations (e.g., Chun & Jiang, 1998; Kidd, Piantadosi, & Aslin, 2012; Zhao, Al-Aidroos, & Turk-Browne, 2013). In addition, statistical learning can operate over many different modalities and types of stimuli, including syllables (Saffran et al., 1996), nonlinguistic sounds (Gebhart, Newport, & Aslin, 2009), shapes (Fiser & Aslin, 2001), scenes (Brady & Oliva, 2008), tactile stimuli (Conway & Christiansen, 2005), and spatial locations (Mayr, 1996). Finally, such learning takes place throughout the life span, from infants (Kirkham, Slemmer, & Johnson, 2002; Saffran et al., 1996) to older adults (Campbell, Zimerman, Healey, Lee, & Hasher, 2012; Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014). This article will review the current understanding of how such forms of behavior are supported by the brain, as well as the insights that these neural investigations have provided about the nature of statistical learning.

Paired-Associate Learning

Some of the first evidence for how the brain might learn new regularities came from monkey physiology studies of visual associative learning. In a seminal study, Miyashita (1988) found that after viewing a sequence of fractal images presented many times in the same order, individual neurons in the macaque temporal lobe came to respond similarly to images presented nearby in the sequence. This was one of the first demonstrations that neural selectivity can reflect learned associations (as opposed to stimulus features). This led to more focused studies on how objects become associated, using the paired-associate task (Figure 1).

In this task, the subject is presented with a cue stimulus, followed by a delay of a few seconds, and then two choice stimuli. One of the choice stimuli has been preselected as the rewarded associate of the cue, and the subject learns by trial and error to choose this associate. This training has a similar effect as in Miyashita (1988), leading neurons in the temporal lobe to respond selectively to both members of a pair (Sakai & Miyashita, 1991). A later study established a more direct link between exposure to regularities and neural changes by examining the timecourse, rather than just the end point, of learning (Messinger, Squire, Zola, & Albright, 2001). These increases in neural similarity can occur even if one of the paired stimuli is task-irrelevant and the learning is thus incidental (Erickson & Desimone, 1999). In fact, such associative learning can occur merely as a result of eye movements between stimuli (Li & DiCarlo, 2008).

Initially, the neurons undergoing such change were thought to be located in area TE of the macaque anterior inferior temporal (IT) cortex, and indeed, this area does seem to participate in learning (e.g., Li & DiCarlo, 2008). However, an adjacent area of the medial temporal lobe (MTL) – perirhinal cortex (PRC) – contains a higher proportion of pair-coding neurons (Naya, Yoshida, & Miyashita, 2003). Indeed, learning occurs more rapidly in PRC than in IT (Naya, Yoshida, & Miyashita, 2001) and PRC is necessary for effects to appear in IT (Buckley & Gaffan, 1998; Higuchi & Miyashita, 1996).
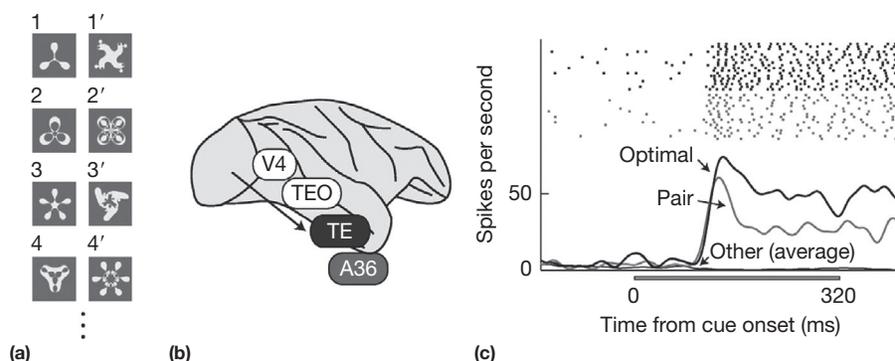
Figure 1   Paired-associate learning. (a) Examples of paired fractal stimuli. (b) Lateral view of macaque brain displaying the ventral visual processing stream. A36 is part of PRC. (c) An example neuron that responds optimally to a particular fractal and very strongly to that fractal's pair but not to any other fractals, demonstrating associative learning. The raster plot, top, shows action potentials on separate trials, and the lines, below, are average neural firing rates across trials. Adapted with permission from Osada, T., Adachi, Y., Kimura, H. M., & Miyashita, Y. (2008). Towards understanding of the cortical network underlying associative memory. *Philosophical Transactions of the Royal Society of London*, *Series B: Biological Sciences*, *363*, 2187–2199.

Studies of associative learning in nonhuman primates have guided research on statistical learning in the human brain by showing that exposure to temporally structured information leads to rapid changes in neural selectivity (e.g., Messinger et al., 2001) and by suggesting that the MTL may be a particularly relevant brain system (e.g., Osada, Adachi, Kimura, & Miyashita, 2008; Wirth et al., 2003).

Segmentation of Continuous Streams

Unlike the paired-associate task, in naturalistic settings, it is often unclear which objects should be associated. The world is noisy and continuous – there is rarely temporal separation between groups of objects or other obvious cues that indicate boundaries. For example, we learn which syllables go together to form words, despite the fact that there are often no clear pauses between adjacent syllables from different words (Kuhl, 2004). In order to extract meaningful chunks of structure from a continuous stream, it is therefore necessary to pick up on the statistical relationships between objects.

In an initial demonstration of this ability, Saffran et al. (1996) exposed infants to a continuous stream of syllables that lasted only two minutes. Critically, the order of the syllables was generated from four groups of three syllables, or 'words.' The syllables within a word always appeared in the same sequence, but the order of words was random. As a result, the transition probabilities between syllables within words (1.0) were higher than between words (0.33). In a subsequent test using a preferential attention procedure, infants showed that they were sensitive to these statistics by discriminating between words and recombinations of the same syllables. This canonical demonstration of statistical learning provided a possible mechanism for identifying words during language acquisition.

A handful of functional magnetic resonance imaging (fMRI) studies have investigated the neural basis of this kind of auditory statistical learning. One key region is the superior temporal gyrus (STG), whose activation distinguishes between syllable sequences with regularities relative to random sequences (Karuza et al., 2013; McNealy, Mazziotta, & Dapretto, 2006).

The STG, which is involved in auditory perception and speech processing (Hickok & Poeppel, 2007), has been implicated more broadly in the extraction of auditory regularities (Cunillera et al., 2009; Overath et al., 2007). Another important region is the inferior frontal gyrus (IFG), whose activation seems to track the expression of auditory statistical learning in behavior (Karuza et al., 2013; see also Abla & Okanoya, 2008). The precise functions of these regions and how they interact to support statistical learning are active areas of research.

The need to segment continuous input into meaningful chunks extends to the visual domain. Visual information is highly structured over space and time, allowing us to construct representations of scenes and events. Such visual statistical learning has been investigated with variants of the Saffran et al. (1996) task, in which regularities are constructed from stimuli such as shapes (Fiser & Aslin, 2002), colors (Turk-Browne, Isola, Scholl, & Treat, 2008), scenes (Brady & Oliva, 2008), and actions (Baldwin, Andersson, Saffran, & Meyer, 2008).

Initial fMRI studies showed that visual statistical learning occurs quickly within a few repetitions of regularities and recruits known memory systems, including the MTL and striatum (Turk-Browne, Scholl, Chun, & Johnson, 2009). The MTL, and hippocampus in particular, is especially engaged by regularities when they provide a predictive cue about what will occur next (Turk-Browne, Scholl, Johnson, & Chun, 2010).

Though these studies indicated some involvement of the MTL, it was not clear exactly what kind of learning might be occurring there. The paired-associate studies in nonhuman primates reviewed earlier suggested that the MTL might represent regularities by increasing the similarity of neural representations of associated objects. Schapiro, Kustner, and Turk-Browne (2012) investigated such representational changes in the human brain using a statistical learning paradigm in which pairs of abstract objects were embedded in a continuous stream (Figure 2). The pattern of activity evoked by each object was measured before and after exposure to these regularities with high-resolution fMRI, and the similarity of these patterns was examined in the MTL. Both cortical and hippocampal subregions showed increased similarity for objects that had been
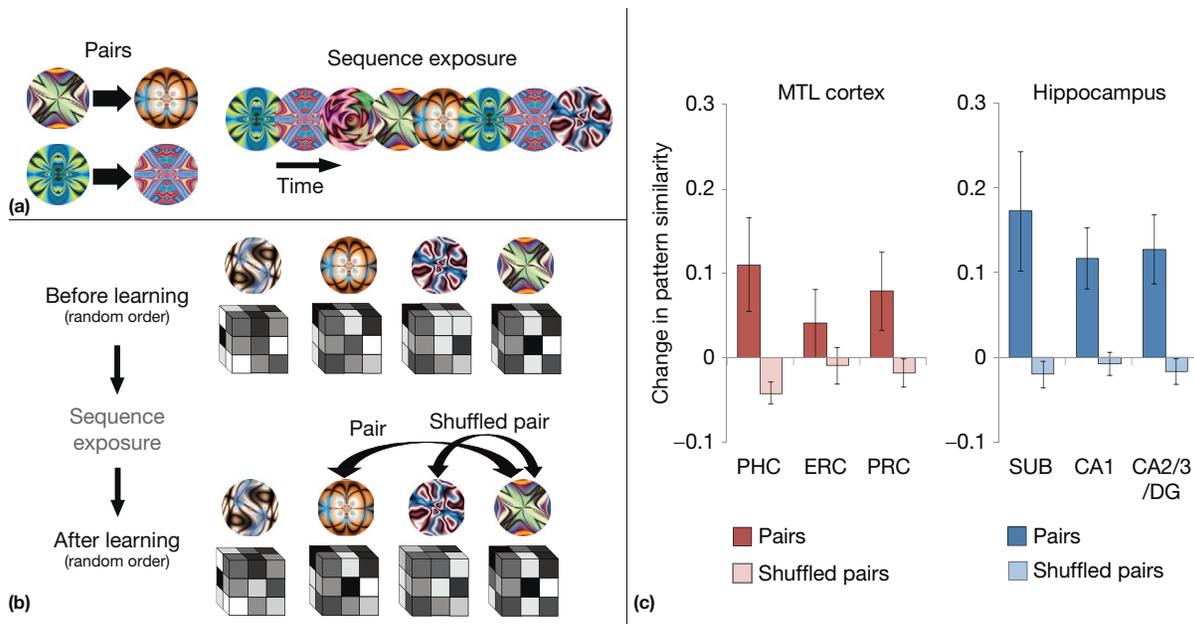
Figure 2   Design and results from Schapiro et al. (2012). (a) Participants were exposed to a continuous sequence generated from pairs of abstract stimuli. (b) Before and after this exposure, stimuli were presented in a random order, and the pattern of activity evoked by each stimulus was measured within regions of interest in the MTL. Correlations were calculated for the patterns corresponding to the members of a pair and for recombinations of the same stimuli (shuffled pairs). (c) From before to after learning, pattern similarity increased throughout much of the MTL cortex and hippocampus for pairs, but not shuffled pairs. PHC, parahippocampal cortex; ERC, entorhinal cortex; PRC, perirhinal cortex; SUB, subiculum; CA1, cornu ammonis field 1; CA2/3/DG, cornu ammonis fields 2 and 3 and dentate gyrus. Adapted with permission from Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*.

paired. Moreover, in the CA3 subfield, the first object in a pair changed more than the second, consistent with this region playing a role in prediction.

Beyond fMRI studies, converging support for the role of the MTL in statistical learning comes from patient work. In particular, a case study of a patient with complete bilateral hippocampal loss revealed widespread deficits in statistical learning across a range of visual and auditory stimuli (Schapiro et al., 2014). This suggests that although prior studies of auditory statistical learning have not focused on the MTL, it may nevertheless play a domain-general role. In fact, this may also be true about some of the regions implicated in auditory statistical learning. For example, the IFG also supports visual statistical learning: Its response to regularities correlates with later behavioral familiarity (Turk-Browne et al., 2009), and it shows analogous representational changes to what was found in the MTL (Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013).

More Complex Statistical Structure

The previous section focused on the extraction of reliable groupings of stimuli in the service of segmenting continuous streams. Typically, these groupings are deterministic, with syllables or shapes within a regularity co-occurring with high or perfect probability. But the contingencies in naturalistic input can be much more variable, resulting in weaker probabilistic relationships between stimuli. Many studies have examined how such contingencies are extracted by exposing participants to sequences constructed from a complex generative process.

This work tends not to focus on segmentation per se, but rather on sensitivity to the underlying structure.

Because the relationships in these studies are probabilistic, any given stimulus will be somewhat associated with multiple stimuli. This is sometimes conceptualized in terms of a transition matrix, where each stimulus is indexed by a row and column, and each cell reflects the probability of transitioning from the stimulus indexed by the row to the stimulus indexed by the column. These matrices have been used to quantify the degree of predictability in a sequence – or the opposite, its entropy – which can then be used to identify brain systems sensitive to varying levels of structure. The temporoparietal junction (TPJ) is engaged at times of low predictability in both visual (Bischoff-Grethe, Proper, Mao, Daniels, & Berns, 2000) and auditory (Furl et al., 2011; see also Nastase, Iacovella, & Hasson, 2013) modalities; this may reflect reorienting to unexpected stimuli, especially in the right hemisphere (Corbetta, Patel, & Shulman, 2008). Consistent with the studies in the previous section, the MTL is engaged by sequences with high predictability (Durrant, Cairney, & Lewis, 2012); interestingly, the MTL is less involved, and the striatum more involved, after a period of consolidation.

A related paradigm known as artificial grammar learning uses finite-state automata to generate structured sequences (Reber, 1967). These grammars have start and end states, and each sequence can be presented simultaneously as a string (Figure 3). After exposure to several of these strings, participants are able to discriminate between new test strings that are consistent versus inconsistent with the grammar. These grammaticality judgments can be based either on similarity to
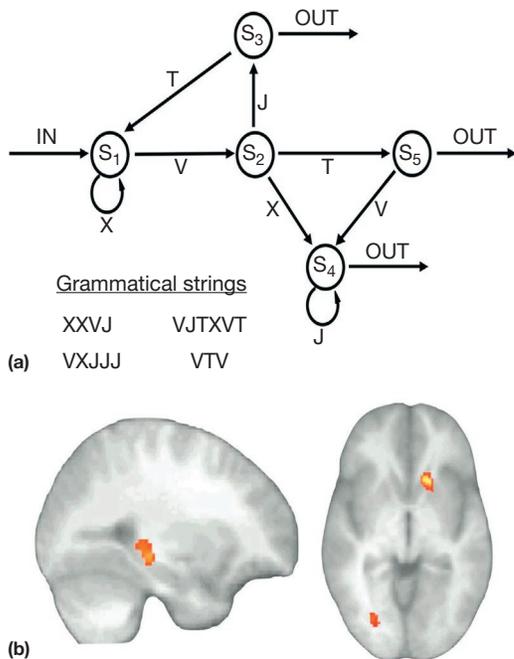
Figure 3   Artificial grammar learning. (a) Grammar used to generate sequences of strings in Lieberman, Chang, Chiao, Bookheimer, and Knowlton (2004). Generation of a string begins at IN, and at every transition from one state to the next, the letter corresponding to the arrow between states is added to the string. Examples of grammatical strings are shown. (b) Hippocampal activation (left) is associated with grammaticality judgments based on the similarity of test strings to training strings, whereas caudate activation (right) is associated with judgments that require sensitivity to the rules of the grammar. Adapted with permission from Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*, 16, 427–438.

substrings that had been experienced during training, which seems to be supported by the hippocampus, or on adherence to the set of rules embodied by the grammar, which seems to be supported by the striatum and frontal cortex (Channon et al., 2002; Dolan & Fletcher 1999; Lieberman et al., 2004; Strange, Duggins, Penny, Dolan, & Friston, 2005; see also Poldrack et al., 2001).

In addition to revealing sensitivity to stimulus predictability, a similar transition matrix approach has been used to study contingencies in the motor domain. In this paradigm, known as the serial reaction time task (SRTT), participants make a predetermined response whenever a stimulus appears in a particular location (Nissen & Bullemer, 1987). Both the hippocampus and striatum are sensitive to different measures of structure during this task, including entropy, conditional entropy, and mutual information (Bornstein & Daw, 2012; Harrison, Duggins, & Friston, 2006; Strange et al., 2005; Thomas et al., 2004; see also Huettel, Mack, & McCarthy, 2002). The profile of hippocampal responses observed across these studies suggests that the hippocampus may be making predictions about possible outcomes, with more widespread predictions arising at times of higher uncertainty about the future (Bornstein & Daw, 2012, 2013).

In these SRTT studies, only the current stimulus contributes information about what will occur next in the sequence. However, in contexts in which probabilities are weak, it can sometimes be beneficial to consider a longer history of events in order to better predict the future. This has been tested in modified versions of the SRTT where both the frequency of items and frequency of item pairs are equated, forcing participants to rely on at least second-order relationships (Reber & Squire, 1994). Although all forms of the SRTT are thought to rely on the striatum (Grafton, Hazeltine, & Ivry, 1995; Rauch et al., 1997; Seidler et al., 2005), the MTL seems to be recruited for this more complex form in particular (Curran, 1997; Schendan, Searl, Melrose, & Stern, 2003; Shanks, Channon, Wilkinson, & Curran, 2006; cf. Reber & Squire, 1994).

Spatial Regularities

The focus of the article so far has been on sequential regularities, but the spatial arrangement of objects is also highly structured over time (Biederman, Mezzanotte, & Rabinowitz, 1982). For example, every time you visit a building, the layout of rooms and the configuration of furniture are roughly the same. The learning of spatial regularities has been studied in a paradigm analogous to the Saffran et al. (1996) study discussed earlier in the text, where participants extract recurring configurations of objects from complex scenes rather than subsequences of objects from a continuous stream (Fiser & Aslin, 2001). Another paradigm, contextual cueing, was designed specifically to examine the learning of spatial regularities (Chun & Jiang, 1998). In contextual cueing, participants search for a target letter in an array of distractors. Unbeknownst to them, distractor arrays are paired with a consistent target location, such that the target can be found more quickly over time. Learning in this task seems to depend on the MTL, as evidenced by patient work (Chun & Phelps, 1999; cf. Manns & Squire, 2001) and fMRI studies (Greene, Gross, Elsinger, & Rao, 2007; Manelis & Reder, 2012; Westerberg, Miller, Reber, Cohen, & Paller, 2011). This suggests a parallel between statistical learning in temporal and spatial domains.

Conclusions

Across the diverse statistical learning tasks considered here, there are two broad classes of brain systems involved. One class consists of cortical regions, which seem to participate only in certain forms of statistical learning. The STG is engaged by sequential regularities in the auditory domain and may reflect sensory and/or linguistic processes. The IFG is involved in segmenting sequential regularities in both the auditory and visual domains and seems to be correlated with behavioral expressions of learning. Finally, the TPJ responds when sequential regularities are violated, consistent with its role in reorienting.

The other class of brain regions consists of specialized memory systems, which seem to be involved across all of the paradigms discussed here. The striatum, which is known to be involved in procedural learning, also contributes to many forms of statistical learning – even ones that do not involve

motor responses. Perhaps more surprising, the MTL and hippocampus in particular are involved in virtually all types of statistical learning that we considered. This is surprising for two reasons: First, the hippocampus is specialized for encoding individual episodes (McClelland, McNaughton, & O'Reilly, 1995), and yet statistical learning requires aggregating across multiple episodes. Second, the hippocampus is known to support declarative memory (Squire, Stark, & Clark, 2004), and yet many of the forms of learning we reviewed occur incidentally and often implicitly (e.g., Chun & Phelps, 1999; Destrebecqz & Cleeremans, 2001; Kim et al., 2009). These findings are consistent with the view that computations in the hippocampus can support multiple types of learning without always resulting in awareness (Hannula & Greene, 2012; Henke, 2010). Specifically, associations rapidly formed within the hippocampus via relational binding may provide the building blocks of statistical learning.

The interactions between these cortical regions and memory systems during statistical learning remain to be explored. For example, given the rapid learning ability of the hippocampus, it may play a role in mediating cortical responses to structure. Regardless, one conclusion that can be drawn at this stage – based on the repeated observation of the same brain regions across tasks – is that statistical learning is at least partly a domain-general process. This explains why statistical learning in one domain influences performance in another, such as visual segmentation predicting language abilities (Arciuli & Simpson, 2012; Frost, Siegelman, Narkiss, & Afek, 2013), and suggests that statistical learning is a pervasive element of cognition.

See also: **INTRODUCTION TO ANATOMY AND PHYSIOLOGY**: Auditory Cortex; Basal Ganglia; Topographic Layout of Monkey Extrastriate Visual Cortex; **INTRODUCTION TO CLINICAL BRAIN MAPPING**: Language; **INTRODUCTION TO COGNITIVE NEUROSCIENCE**: Category Learning; Familiarity; Neural Correlates of Motor Skill Acquisition and Consolidation; Neuroimaging Studies of Reinforcement-Learning; Prediction and Expectation; The Medial Temporal Lobe and Episodic Memory; Uncertainty; **INTRODUCTION TO METHODS AND MODELING**: Multi-voxel Pattern Analysis; **INTRODUCTION TO SYSTEMS**: Memory.

References

Abla, D., & Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia*, *46*, 2787–2795.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*, 286–304.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*, 170–176.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*, 1382–1407.

Berkes, P., Orban, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, *331*, 83–87.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.

Bischoff-Grethe, A., Proper, S. M., Mao, H., Daniels, K. A., & Berns, G. S. (2000). Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *Journal of Neuroscience*, *20*, 1975–1981.

Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, *35*, 1011–1023.

Bornstein, A. M., & Daw, N. D. (2013). Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLoS Computational Biology*, *9*(12), e1003387.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*, 678–685.

Buckley, M. J., & Gaffan, D. (1998). Perirhinal cortex ablation impairs configural learning and paired-associate learning equally. *Neuropsychologia*, *36*, 535–546.

Campbell, K. L., Zimerman, S., Healey, M. K., Lee, M. M., & Hasher, L. (2012). Age differences in visual statistical learning. *Psychology and Aging*, *27*, 650–656.

Channon, S., Shanks, D., Johnstone, T., Vakili, K., Chin, J., & Sinclair, E. (2002). Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia*, *40*, 2185–2197.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, *2*, 844–847.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24–39.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*, 306–324.

Cunillera, T., Camara, E., Toro, J. M., Marco-Pallares, J., Sebastian-Galles, N., Ortiz, H., et al. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*, *48*, 541–553.

Curran, T. (1997). Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience*, *9*, 522–533.

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin and Review*, *8*, 343–350.

Dolan, R. J., & Fletcher, P. F. (1999). Encoding and retrieval in human medial temporal lobes: an empirical investigation using functional magnetic resonance imaging (fMRI). *Hippocampus*, *9*, 25–34.

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2012). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cerebral Cortex*, *23*, 2467–2478.

Erickson, C. A., & Desimone, R. (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. *Journal of Neuroscience*, *19*, 10404–10416.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 458–467.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*, 1243–1252.

Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., & Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *NeuroImage*, *54*, 2267–2277.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin and Review*, *16*, 486–490.

Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, *7*, 497–510.

Greene, A. J., Gross, W. L., Elsinger, C. L., & Rao, S. M. (2007). Hippocampal differentiation without recognition: An fMRI analysis of the contextual cueing task. *Learning and Memory*, *14*, 548–553.

Hannula, D. E., & Greene, A. J. (2012). The hippocampus reevaluated in unconscious learning and memory: At a tipping point? *Frontiers in Human Neuroscience*, *6*, 80.

Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks*, *19*, 535–546.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*, 523–532.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.

Higuchi, S., & Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 739–743.

Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience*, *5*, 485–490.

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, *127*, 46–54.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, *7*, e36399.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, *461*, 145–149.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843.

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, *321*, 1502–1507.

Li, W., Piech, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience*, *7*, 651–657.

Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*, *16*, 427–438.

Manelis, A., & Reder, L. M. (2012). Procedural learning and associative memory mechanisms contribute to contextual cueing: Evidence from fMRI and eye-tracking. *Learning and Memory*, *19*, 527–534.

Manns, J. R., & Squire, L. R. (2001). Perceptual learning, awareness, and the hippocampus. *Hippocampus*, *11*, 776–782.

Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 350–364.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *Journal of Neuroscience*, *26*, 7629–7639.

Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 12239–12244.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, *335*, 817–820.

Nastase, S., Iacovella, V., & Hasson, U. (2013). Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Human Brain Mapping*, *35*, 1111–1128.

Naya, Y., Yoshida, M., & Miyashita, Y. (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science*, *291*, 661–664.

Naya, Y., Yoshida, M., & Miyashita, Y. (2003). Forward processing of long-term associative memory in monkey inferotemporal cortex. *Journal of Neuroscience*, *23*, 2861–2871.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning – Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.

Osada, T., Adachi, Y., Kimura, H. M., & Miyashita, Y. (2008). Towards understanding of the cortical network underlying associative memory. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *363*, 2187–2199.

Overath, T., Cusack, R., Kumar, S., Von Kriegstein, K., Warren, J. D., Grube, M., et al. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biology*, *5*, e288.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546–550.

Purves, D., Wojtach, W. T., & Lotto, R. B. (2011). Understanding vision in wholly empirical terms. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(Suppl. 3), 15588–15595.

Rauch, S. L., Whalen, P. J., Savage, C. R., Curran, T., Kendrick, A., Brown, H. D., et al. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping*, *5*, 124–132.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855.

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning and Memory*, *1*, 217–229.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

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

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747.

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*, 1622–1627.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*, 486–492.

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*, 1013–1025.

Seidler, R. D., Purushotham, A., Kim, S. G., Ugurbil, K., Willingham, D., & Ashe, J. (2005). Neural correlates of encoding and expression in implicit sequence learning. *Experimental Brain Research*, *165*, 114–124.

Shanks, D. R., Channon, S., Wilkinson, L., & Curran, H. V. (2006). Disruption of sequential priming in organic and pharmacological amnesia: A role for the medial temporal lobes in implicit contextual learning. *Neuropsychopharmacology*, *31*, 1768–1776.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, *27*, 279–306.

Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Networks*, *18*, 225–230.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*, 792–814.

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of developmental differences in implicit sequence learning: an fMRI study of children and adults. *Journal of Cognitive Neuroscience*, *16*, 1339–1351.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 399–407.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*, 1934–1945.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*, 11177–11187.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.

Westerberg, C. E., Miller, B. B., Reber, P. J., Cohen, N. J., & Paller, K. A. (2011). Neural correlates of contextual cueing are modulated by explicit learning. *Neuropsychologia*, *49*, 3439–3447.

Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science*, *300*, 1578–1581.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*, 667–677.