



Published in final edited form as:

*Neurosci Biobehav Rev.* 2020 May ; 112: 279–299. doi:10.1016/j.neubiorev.2020.01.032.

## How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning

**Christopher M. Conway\***

Center for Childhood Deafness, Language, and Learning, Boys Town National Research Hospital, Omaha, NE, United States

### Abstract

Despite a growing body of research devoted to the study of how humans encode environmental patterns, there is still no clear consensus about the nature of the neurocognitive mechanisms underpinning statistical learning nor what factors constrain or promote its emergence across individuals, species, and learning situations. Based on a review of research examining the roles of input modality and domain, input structure and complexity, attention, neuroanatomical bases, ontogeny, and phylogeny, ten core principles are proposed. Specifically, there exist two sets of neurocognitive mechanisms underlying statistical learning. First, a “suite” of associative-based, automatic, modality-specific learning mechanisms are mediated by the general principle of cortical plasticity, which results in improved processing and perceptual facilitation of encountered stimuli. Second, an attention-dependent system, mediated by the prefrontal cortex and related attentional and working memory networks, can modulate or gate learning and is necessary in order to learn nonadjacent dependencies and to integrate global patterns across time. This theoretical framework helps clarify conflicting research findings and provides the basis for future empirical and theoretical endeavors.

### Keywords

Statistical learning; Implicit learning; Sequential learning; Artificial grammar learning

### 1. Introduction

Many events in our daily existence occur not completely randomly or haphazardly, but with a certain amount of structure, regularity, and predictability. Because of the ubiquitous presence of structured patterns in human action, perception, and cognition, the ability to

---

\*Corresponding author at: Brain, Learning, and Language Lab, Center for Childhood Deafness, Language, and Learning, Boys Town National Research Hospital, Lied Learning and Technology Center, 555N. 30<sup>th</sup> St., Omaha, NE 68131, United States. christopher.conway@boystown.org.

Declarations of Competing Interest

None.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.neubiorev.2020.01.032>.

process and represent these patterns is of paramount importance. This type of structured pattern learning – which is likely a crucial foundational ability of all higher-level organisms, and possibly of many lower-level ones as well – has been studied under the guise of different terms for what arguably tap into aspects of the same underlying construct, including “implicit learning” (A.S. Reber, 1967), “sequence learning” (Nissen and Bullemer, 1987), “sequential learning” (Conway and Christiansen, 2001), and “statistical learning” (Saffran et al., 1996).

Despite gains made in understanding how humans and other organisms learn patterned input, we are still far from an understanding of the neurocognitive mechanisms underlying learning and what factors constrain its emergence across individuals, species, and learning situations. What is needed is an integration of research findings across six key areas that have generally been treated in isolation:

1. *Input modality and domain.* How does learning proceed for inputs across different perceptual modalities (e.g., vision vs. audition) or domains (e.g., language vs. music)? Does the learning of patterns in one modality or domain involve the same neurocognitive mechanisms as learning in a different modality or domain?
2. *Input structure and complexity.* What mechanisms underpin the learning of different types of input patterns, such as associations between adjacent or co-occurring elements to more complex “global” patterns that require integration of information over longer time-scales?
3. *Role of attention.* To what extent are attention and related cognitive processes necessary for statistical learning to occur? In turn, does the outcome of learning modulate attention?
4. *Neural bases.* What is the underlying neuroanatomy of statistical learning? Is there a single, common learning and processing network? Or are there different sets of regions or networks that are used for different types of learning situations?
5. *Ontogenetic constraints.* How does statistical learning emerge and change across the lifespan? Do different aspects of learning have different developmental trajectories?
6. *Phylogenetic constraints.* Which aspects of statistical learning are shared versus unique across different animal species? What drives variation or differences across species?

Although a number of theoretical perspectives exist (e.g., Arciuli, 2017; Aslin and Newport, 2012; Daltrozzo and Conway, 2014; Forkstam and Petersson, 2005; Frost et al., 2015; Janacek and Nemeth, 2012; Keele et al., 2003; Perruchet and Pacton, 2006; Pothos, 2007; P.J. Reber, 2013; A.S. Reber, 2003; Savalia et al., 2016; Seger, 1994; Thiessen and Erickson, 2013), currently none of them sufficiently address all of these questions. In this paper, we begin by defining in more detail what is meant by “statistical learning” and how it relates to other similarly-used terms. A lack of clarity and consensus in regards to terminology has

proven to be a barrier for integrating findings across different research areas; furthermore, the use of certain terms denotes premature assumptions about what the underlying mechanisms are that characterize the construct of interest. Following this discussion, we provide a selective review and synthesis of research related to the six areas described above. Then, based on this review, we outline ten core principles that arise from an integration of the reviewed research and provide the beginnings of what could be construed as a unified theory of statistical learning. To preview, we propose there exist two primary sets of neurocognitive mechanisms – one based on the general principle of cortical plasticity and the other a specialized neural system that can provide top-down modulation of learning – with each affected and constrained by different factors in different ways. Only by taking into account the operation of these two mechanisms will we understand the neurocognitive bases of statistical learning and how they are constrained by factors such as input modality, complexity, ontogeny, and phylogeny.

## 2. Preliminary considerations

Statistical learning research began in earnest with the seminal study by Saffran et al. (1996), who showed that 8-month-old infants were sensitive to the statistical structure inherent in a short (2-minute) auditory nonword speech stream. In this study, statistical structure was operationalized as the strength of transitional probabilities between adjacent syllables (i.e., the likelihood of a given syllable occurring next based on the current syllable).

Subsequent research examined the generality of this phenomenon, demonstrating learning not only in human infants (Kirkham et al., 2002), but also adults, and not only with speech-like input, but also for non-linguistic sound sequences (Saffran et al., 1999) and visual scenes (Fiser and Aslin, 2001). Thus, statistical learning was quickly recognized as a general-purpose mechanism, robust across tasks, situations, and perhaps even species (Conway and Christiansen, 2001). It should be noted that the term “statistical learning” is limited in that it would seem to imply that learning and processing of input patterns consists of making statistical computations. Although the input in learning tasks can often be described in terms of statistical regularities (e.g., transitional probabilities between stimuli), it is as yet an open question whether in fact the brain learns and represents statistical regularities per se or whether what is learned is something different such as memory for frequently occurring clusters of items or “chunks” (Orban et al., 2008; Perruchet and Pacton, 2006; Slone and Johnson, 2018). This point will be returned to in section 3.2.

For decades prior to these initial studies, another area of research had been focused on a similar phenomenon, known as “implicit learning” (A.S. Reber, 1967, 1989). Implicit learning is generally defined as “learning without awareness” (Cleeremans et al., 1998) but it has been argued that statistical learning and implicit learning both refer to the same general learning phenomenon (Batterink et al., 2019; Christiansen, 2018; Perruchet and Pacton, 2006). Indeed, both types of learning reflect a type of incidental pattern learning (i.e., learning occurring without intention or instruction). For this reason, we regard the similarities among statistical learning and implicit learning research as indicative that there may be core processes that contribute to both, and as such, we look for insights that may be

gained by considering research findings from both areas (and other areas of research as well).

Of course, even the early research on implicit learning did not occur in a vacuum. Behaviorist approaches to associative learning and conditioning provided an important historical context for the implicit learning work (e.g., see Pearce and Bouton, 2001; Rescorla and Wagner, 1972). Gureckis and Love (2007) in fact argued that much of what the field has been studying under the guise of statistical learning is embodied by behaviorist principles of conditioning and associative learning (c.f., Goddard, 2018). Additional neurophysiological antecedents of statistical learning include Hebb's principles of learning and plasticity (i.e., the "Hebbian learning rule"; see Cooper, 2005; Hebb, 1949) and the demonstration that the development of primary visual cortex depends on environmental experience (e.g., Blakemore and Cooper, 1970). While acknowledging these important precursors, this review focuses primarily on the findings from the implicit learning and statistical learning literatures per se.

For simplicity the term "statistical learning" is used in the remainder of this paper to refer to incidental learning of structured patterns encountered in the environment. To constrain and operationalize the definition of statistical learning and provide added focus to this review, we delineate the task or situational characteristics of interest. Specifically, we propose three orthogonal dimensions that can help clarify the construct of statistical learning. These dimensions include: the level of structure present in input (i.e., random versus heavily structured sequences); the amount of exposure that is involved (i.e., a single exposure versus multiple instances); and the extent to which task situations provide explicit instruction or overt feedback (i.e., incidental versus intentional learning situations). These three dimensions are depicted graphically in Fig. 1. They create a "task space" containing a continuum of distributed points in which tasks (or situations) that are closer to the zero-point (0,0,0) can be thought of as being more characteristic of statistical learning compared to tasks at the periphery (note though, that technically speaking there is no actual zero-point as there could always be a situation with more exposures, or more structure, etc. and in that case the zero-point might be more appropriately regarded as a mathematical asymptote or singularity). Thus, the situations that we consider to be "canonical" for statistical learning have the following characteristics: structured input patterns presented over multiple exposures under incidental conditions in which there is no instruction to learn or attend to the patterns per se. Note, from this perspective, the phenomenon of statistical learning is not a categorical distinction, but a graded, continual one in which certain tasks or situations might elicit such learning more so than others.

Likewise, we can consider the types of tasks that have been used to investigate statistical learning and related learning phenomena. The primary tasks include the artificial grammar learning (AGL) task (A.S. Reber, 1967), the serial reaction time (SRT) task (Nissen and Bullemer, 1987), and the word segmentation task and its variants (Fiser and Aslin, 2001; Saffran et al., 1996). Table 1 differentiates these three tasks in terms of the measure of learning, the input structure and perceptual modality of the stimuli, and whether or not the task requires generalization to new, previously unencountered items. Despite some differences, what is common across tasks is that participants receive repeated exposure to

structured patterns, usually under incidental learning conditions and without overt feedback. The general finding is that under such conditions, participants show facilitation of or sensitivity to the underlying structure, and this often – though not always – is accompanied by an inability to verbalize one's knowledge of what has been learned.

### 3. Six key questions

#### 3.1. How is learning affected by input modality and domain?

For some time now, it has been known that statistical learning is not tied to a single perceptual modality or cognitive domain. Indeed, even a cursory review of findings from the three canonical tasks (Table 1), shows that learning can occur with auditory language-like material (Saffran et al., 1996), strings of letters (A.S. Reber, 1967), non-language auditory input such as pure tones (Saffran et al., 1999) or sequences of musical timbre (Tillman and McAdams, 2004), visual scenes and shapes (Fiser and Aslin, 2001), visual-motor patterns (Nissen and Bullemer, 1987), and tactile input (Conway and Christiansen, 2005). The demonstration of learning across such a widespread set of domains and input types immediately prompted suggestions that statistical learning should be thought of as a unitary, domain-general learning phenomenon that applies across a wide range of situations (Kirkham et al., 2002). That is, it is logically possible that statistical learning is governed by a single mechanism or neurocognitive principle that applies across a wide range of input types.

On the other hand, a series of studies showed that although learning of structured patterns can occur across various perceptual domains, the way that learning occurred in different modalities differed, suggesting the involvement of multiple modality-specific learning mechanisms (Conway and Christiansen, 2005; 2006; 2009; Emberson et al., 2011). For instance, adult participants showed higher levels of learning for auditory serial patterns compared to visual serial patterns – despite the patterns across perceptual modalities being equated in terms of low-level perceptual factors (Conway and Christiansen, 2005). In addition, the rate of presentation of serial input patterns had opposite effects on auditory and visual learning, with auditory and visual learning excelling at fast and slow presentation rates, respectively (Emberson et al., 2011). Moreover, different patterns presented in multiple streams of stimuli could be learned simultaneously and independently of each other, as long as the input streams were instantiated in different perceptual modalities (visual versus auditory) or perceptual categories (shapes versus colors; tones versus nonwords) (Conway and Christiansen, 2006). Given such findings, Conway (2005) and Conway and Christiansen (2005; 2006) proposed that aspects of statistical learning might share similarities with perceptual priming or perceptual learning (Conway et al., 2007), in which networks of neurons in modality-specific brain regions show decreased activity and improved facilitation for items that are similar to those previously experienced (P.J. Reber et al., 1998; Schacter and Badgaiyan, 2001). Furthermore, Conway (2005) suggested that although learning is implemented by a set of common computational principles or algorithms that exist across perceptual domains, there are processing differences within each perceptual modality that affect learning, such as audition and vision being differentially adept at picking up information distributed in time and space, respectively. Note, too, that

from this perspective, it is not only the sensory modality that is important (e.g., auditory), but also the type of domain or category (e.g., verbal material vs. nonlinguistic tones; Conway and Christiansen, 2006).

However, under a purely domain-specific viewpoint, learning in one perceptual modality or domain would have no bearing or relation to learning and processing in another perceptual modality or domain. This does not appear to be the case. For instance, a number of studies have demonstrated that input presented in one perceptual modality can affect pattern learning in a second concurrently presented modality (Cunillera et al., 2010; Mitchel and Weiss, 2011; Mitchel et al., 2014; Seitz et al., 2007; Thiessen, 2010). This implies an ability for learners to integrate information across different modalities or domains, a challenge for modality-specific processing accounts. However, it is important to note that for all of these demonstrations of cross-modal learning effects, the stimuli in the two different perceptual domains were presented simultaneously in time. A recent study showed that when cross-modal dependencies are created between sequentially-presented input (e.g., a visual stimulus that is followed by an auditory stimulus with a certain statistical regularity), cross-modal learning does not occur (Walk and Conway, 2016); only sequential dependencies within the same perceptual modality were shown to be learnable by adult participants. One possibility therefore is that the learning of such sequential cross-modal patterns might require additional cognitive resources such as attention or working memory in order to focus on the dependencies in question and link them together across time.

It is also important to note that the motor modality can contribute to learning. For instance, a number of studies have investigated the role of perceptual versus motor learning using the SRT task (e.g., Nemeth et al., 2009; Song et al., 2008). The general finding is that motor learning can make independent contributions to learning over and above that which occurs perceptually (Goschke, 1998). In addition, motor-response learning, but not visual perceptual learning, is unaffected by sensory manipulations of the stimuli (e.g., changes to stimulus colors; Song et al., 2008). Furthermore, motor learning and perceptual sequence learning appear to follow different time-courses of consolidation (Hallgató et al., 2013), further suggesting that the motor and perceptual modalities should be thought of as independent learning systems.

In sum, that input modality can affect statistical learning is no longer questioned (e.g., Frost et al., 2015). However, exactly how these at least partially separable and independent modality-specific learning mechanisms (e.g., visual, auditory, tactile, motor, etc.) operate in a multimodal environment is still not completely understood. It is likely that there may be a combination of modality-specific and domain-general learning processes that work together (e.g., Conway, 2005; Batterink et al., 2019; Keele et al., 2003). Fig. 2 illustrates three candidate architectures corresponding to domain-general, modality/domain-specific, and combined general/specific accounts. As an example of how a combined domain-general and domain-specific account might be instantiated in the brain, Conway and Pisoni (2008) reviewed evidence that statistical learning is associated with both modality-specific perceptual/motor brain regions – such as visual processing occipital regions for learning visual input patterns, auditory processing brain regions for learning auditory input, and motor and premotor cortex for motor learning – as well as areas such as the prefrontal cortex

(PFC) which is involved in processing input across a variety of perceptual modalities and domains. Likewise, Tecumseh and Martins (2014) proposed that for the processing of sequential patterns, the PFC and specifically Broca's area mediates domain-general predictive processing mechanisms that interact with posterior brain networks that mediate modality-specific input processing. Finally, Frost et al. (2015) proposed a similar interaction between domain-specific and domain-general processing, though their emphasis was on the hippocampus, basal ganglia, and thalamus as contributing to multimodal and domain-general processing, rather than the PFC.

More work is needed to specify to what extent different processing modes or mechanisms reflect a combination of modality-specific and domain-general learning under different situations. It is likely that certain cognitive processing resources such as selective attention and cognitive control may modulate or gate learning (e.g., Turk-Browne et al., 2005), and may be necessary for learning multimodal patterns across a temporal sequence. The role of attention as well as the neural bases of statistical pattern learning will be addressed further in subsequent sections (i.e., 3.3 and 3.4); but first, we turn to the question of input structure and complexity.

### 3.2. How is learning affected by the type of input structure?

Related to though independent of the question of input modality, is the question of input structure and complexity: what types of regularities and patterns can be learned, and what learning mechanisms are used to learn different types of structures? This question was central to much of early implicit learning research. Cleeremans et al. (1998) summarized the varying approaches emphasizing different aspects of learning, including distributional or statistical approaches (based on associative learning mechanisms as embodied for instance by neural network models; Cleeremans and McClelland, 1991), exemplar-based approaches (in which newly encountered exemplars are compared to the similarity of previously-memorized whole items; Vokey and Brooks, 1992), fragment-based or chunking approaches (in which newly encountered exemplars are evaluated to the extent to which they contain previously-encountered short chunks that were observed in previous exemplars; Perruchet and Pacteau, 1990), and abstractionist approaches (in which the structure of the relationships among stimuli is represented, independent of the stimuli surface features, perhaps taking the form of IF-THEN statements or algebraic rules; Marcus et al., 1999; A.S. Reber, 1989). Artificial grammar learning research using "balanced chunk strength designs", in which chunk/fragment information was independently varied with the rules of the artificial grammar, showed that the learning of fragment or chunk information can be at least partly dissociated from the learning of grammatical rules (Knowlton and Squire, 1996). That is, the two types of patterns can be learned independently of each other and are subserved by apparently distinct neural and cognitive mechanisms (Lieberman et al., 2004).

What exactly constitutes a grammatical "rule" has led to a certain amount of debate (Altmann and Dienes, 1999; Marcus et al., 1999). One possibility is that what is regarded as a rule is actually a type of "perceptual primitive" (Endress et al., 2009). Perceptual primitives include repetition-based structures (e.g., "ga-ti-ti" and "li-na-na" follow the same ABB repetition pattern), which are highly salient to learners, as well as edge-based

positional regularities, where items occurring at the beginning and ending of sequences tend to be learned more effectively than items in the middle. These perceptual primitives are so-named because they appear to be a type of regularity that is detected and learned on the basis of low-level perceptual mechanisms, common across both ontogeny and phylogeny (though this does not rule out the possibility that other non-perceptual memory systems might also contribute to their learning). From a slightly different perspective, following from the research on balanced chunk-strength designs (Knowlton and Squire, 1996), what is referred to as rule-based information might include positional constraints (such as which stimuli are allowable in initial, middle, or ending positions of a sequence), which is not fully captured by analysis of local chunk information alone. Thus, the umbrella term “rule” likely refers to more than one type of pattern (e.g., perceptual primitives and positional information being two likely candidates).

Apart from rule-based information, one important question raised by Perruchet and Pacton (2006; c.f., Christiansen, 2018) is whether chunk-based learning and statistical learning are the same or independent processes. Research on implicit learning has generally stressed chunk-based approaches whereas research on statistical learning has stressed statistical computations. Chunking models such as PARSER assume that attention to frequently co-occurring units results in an improved memory trace for those items, resulting in the formation of a chunk (Perruchet and Vinter, 1998). In a chunk-based view, chunking mechanisms are the primary way that learning proceeds; sensitivity to statistical relations are not learned per se but rather are a byproduct of the chunking process. On the other hand, it is possible that “chunks” are formed through the detection of transitional probabilities; in such a view, a chunk is the outgrowth of statistical learning processes, being the learned association between two items connected by high transitional probabilities. Several studies have attempted to clarify which mechanism governs pattern learning, with most of the evidence to date favoring chunk-based mechanisms (Giroux and Rey, 2009; Fiser and Aslin, 2005; Perruchet and Poulin-Charronnat, 2012; Orban et al., 2008), though at least one study appears to support a statistical learning approach (Endress and Mehler, 2009). It is possible that both chunk-based and statistical-based computations are available to learners and which process is used depends on the learning conditions, such as the availability of temporal cues which might promote chunking (Franco and Destrebecqz, 2012). It is also possible that forming a chunk among items separated in time (i.e. as part of a temporal sequence) has different cognitive requirements compared to a chunk of spatially-arranged and simultaneously-presented stimuli.

In addition to the distinctions among chunks, statistical associations, and rules, pattern structure can be quantified in other ways. For instance, patterns can differ in relation to how many preceding items are needed to predict the subsequent item in a sequence: for a 1<sup>st</sup> order dependency, only one preceding item is needed to determine the next item, whereas for a 2<sup>nd</sup> order dependency, two preceding items are required, etc. (Gomez, 1997). Likewise, in the statistical learning literature, complexity can be manipulated in terms of the strength of the transitional probabilities between items, the size of the “words” or chunks in word segmentation tasks (e.g., pairs of items or triplets), and the hierarchical arrangement of chunks in visual learning tasks. Similarly, for the serial reaction time task, complexity can be manipulated in terms of the type of sequence pattern (fixed versus probabilistic; first-order



conditional versus second-order conditional; Remillard, 2008) and the length of the sequence.

Within the artificial grammar learning literature, there have been attempts to quantify the level of complexity of input patterns (e.g., Pothos, 2010; Schiff and Katan, 2014; van den Bos and Poletiek, 2008v). For instance, Wilson et al. (2013) used a metric that quantifies the complexity of a finite-state grammar by dividing the number of different stimulus elements in the grammar by the number of unique transitions between stimulus elements. This gives a measure of the grammar's linear predictability or determinism, where a value of 1.0 denotes a perfectly deterministic grammar (i.e., a linear chain) and a lower value denotes a certain level of unpredictability (i.e., branching within the grammar). Wilson et al. (2013) used this metric to examine grammar learning of varying levels of complexity in humans and nonhumans, a point that will be returned to in section 3.6 (see also, Heimbauer et al., 2018). Pothos (2010) proposed an entropy model for quantifying complexity in AGL, borrowing concepts from information theory. Essentially, Shannon entropy is a logarithmic function of the number of different possibilities available; the greater the level of entropy the higher the level of uncertainty. Pothos (2010) found that this measure of entropy correlated with artificial grammar learning performance (greater levels of entropy were associated with lower levels of learning); entropy was also correlated with most other standard measures of complexity and regularity, such as associative chunk strength. Likewise, Schiff and Katan (2014) used a measure of topological entropy to assess 56 previously published AGL studies incorporating a total of 10 different artificial grammars. They found that their measure of entropy was significantly correlated with learning performance, despite the fact that the studies were carried out under different conditions and using different types of stimuli. In sum, it is clear that, regardless of the specific measure used, increased pattern complexity is associated with decreased learning performance on AGL tasks.

For patterns occurring in input sequences, it may also be possible to differentiate the types of structures in terms of three primary types of patterns: fixed sequences, where items in the sequence occur in an arbitrary, inflexible order (e.g., a phone number); statistical-based patterns, where the sequence consists of frequently co-occurring elements such as pairs or triplets defined by transitional probabilities; and hierarchical-based sequences, in which primitive units are combined to create more complex units, such as the case in natural language and other complex domains (Conway and Christiansen, 2001). Supporting this proposal, recent empirical work using the SRT task suggests that the learning of fixed sequences and statistical-based patterns reflect partially different characteristics, both at the behavioral and neural levels (e.g., Kóbor et al., 2018; Simor et al., 2019). For instance, statistical learning appears to occur relatively rapidly and plateaus quickly, whereas sequence learning shows a slower, gradual improvement across learning episodes (Simor et al., 2019). Furthermore, the two types of learning are reflected by different ERP components (Kóbor et al., 2018). These are interesting findings because from a certain perspective, the learning of both a fixed sequence and a statistical-based one could be construed as involving the learning of transitional probabilities inherent in the sequences, with a fixed sequence having transitional probabilities of 1.0 and statistical-based sequences containing transitional probability values less than one. However, the evidence suggests that at least partially separate mechanisms underly the learning of these two types of patterns.

One way to distinguish fixed sequences and statistical-based patterns from hierarchical patterns is by considering the difference between adjacent and nonadjacent dependencies (Gómez, 2002; Remillard, 2008). Adjacent dependencies consist of regularities between two items immediately following each other (e.g., A–B) whereas nonadjacent dependencies consist of regularities between two items in which the two have one or more intervening elements between them (e.g., A-x-B). The distinction between adjacent and nonadjacent dependencies is similar to the distinction made in formal linguistics between finite-state grammars, that generally include adjacent-item dependencies and are thought to be inadequate to describe natural language, and phrase structure grammars, which incorporate non-adjacent item dependencies, can have a recursive or hierarchical structure, and are computationally more powerful and arguably more able to characterize natural language (Fitch and Friederici, 2012; Tecumseh and Martins, 2014; Jager and Rogers, 2012).

Because nonadjacent dependency learning is thought to be a hallmark of human language and possibly other aspects of cognition (Christiansen and Chater, 2015), it is not surprising that there has been much recent interest in this type of learning (e.g., Creel et al., 2004; Deocampo et al., 2019; Frost and Monaghan, 2016; Gómez, 2002; Lany and Gómez, 2008; Pacton and Perruchet, 2008; Romberg and Saffran, 2013; Vuong et al., 2016). The learning of nonadjacent dependencies is often difficult to demonstrate in the lab and appears to generally require that the nonadjacent structure be highlighted – such as by manipulation of the transitional probabilities or through perceptual cues – or for endogenous attention to be properly oriented to the dependencies in question (de Diego-Balaguer et al., 2016d; Gómez, 2002; Newport and Aslin, 2004). More specifically, de Diego-Balaguer et al., 2016d suggested that the learning of nonadjacent dependencies likely is only possible later in development when endogenous attentional mechanisms become available to the learner. Research also suggests that the learning of nonadjacent dependencies recruits neural networks that are separate from those involved in the learning of adjacent dependencies (more on neural bases of nonadjacent dependency learning in section 3.4).

Taken together, it may be possible therefore to think about different types of input structures that vary in complexity. Table 2 presents a rough taxonomy of different types of patterns, with purportedly more “simpler” patterns (i.e., easier to learn) at the top and “more complex” patterns toward the bottom. Pattern complexity can thus be thought of as existing along a continuum from more serial, linear, and adjacent-item associations to dependencies that are more variable, nonadjacent, and/or contain recursive or hierarchical structure (c.f., Dehaene et al., 2015; Petkov and Wilson, 2012). Additional research is needed to specify the cognitive, computational, and neural prerequisites needed to learn patterns of varying structure and complexity, as there have been few studies systematically investigating these factors.

### 3.3. What is the role of attention in learning?

Although statistical learning generally occurs under “incidental” conditions (i.e., without direct instruction or feedback during the learning process), this does not necessarily imply that attention plays no role. Before examining the role of attention in statistical learning, it is

necessary to briefly define and discuss the construct of attention as well as related concepts such as automaticity, working memory, and conscious awareness.

An important distinction can be made between exogenous and endogenous attention (e.g., Chica et al., 2013). Exogenous attention is a bottom-up process in which cognitive resources are captured by salient stimuli in the environment; endogenous attention is a top-down process that provides a way to select which stimuli to process and which to ignore. Related to attention is the notion of automaticity. A cognitive process can be considered automatic if it occurs with little effort and requires few attentional resources (Hasher and Zacks, 1979). More specifically, it has been suggested that automatic behaviors or cognitive processes usually have the following four characteristics (Bargh, 1994): there is a general lack of awareness of the cognitive process that is occurring; there is no intentional initiation of the cognitive process in question; the cognitive process is difficult to stop or alter once it has been initiated; and the cognitive process has a low mental load. Thus, in regard to the role of attention in statistical learning, one question is whether statistical learning can be considered an automatic process (i.e., whether it proceeds without awareness, is initiated without intention, is unable to be controlled once it has started, and whether it has a low mental load or cost). A separate question is what roles do endogenous and exogenous attention play in learning, if any.

It is important to point out that (endogenous) attention is closely linked to the construct of working memory (Awh et al., 2006). For instance, one common definition of working memory is that it refers to processes that “hold a limited amount of information temporarily in a heightened state of availability for use in ongoing information processing” (e.g., Cowan, 1988, 2017). Thus, by this definition, working memory and (endogenous) attention are closely intertwined as the items that are in a heightened state of availability are necessarily within the focus of attention.

Finally, related to the question of attention and working memory, is to what extent statistical learning results in knowledge that is accessible to conscious awareness. Attention and awareness are related – e.g., the involvement of attention is more likely to lead to conscious awareness – but they are not synonymous (Lamme, 2003; Norman et al., 2013). Awareness can emerge when the activation strength or quality of the representations reach a sufficient level (Cleeremans, 2011), regardless of how much attention was originally deployed during the learning task. The extent to which learning proceeds intentionally versus incidentally can be manipulated by task instructions, which in turn can influence the extent that the knowledge that is learned is accessible to conscious awareness (Bertels et al., 2015). Pattern awareness can also emerge naturally during the learning process, even when no instructions are given to explicitly promote explicit strategies or conscious awareness (Singh et al., 2017). Decades of research on implicit learning has demonstrated that some aspects of learning can occur without the involvement of explicit strategies or conscious awareness (e.g., Song et al., 2007; Turk-Browne et al., 2009). To provide focus to the remaining discussion, we focus primarily on the roles of attention and working memory in relation to statistical learning.

Understanding the role of attention and working memory during statistical learning is not straightforward and in fact is a matter of some debate (e.g., Janacsek, and Nemeth, 2013; 2015; Martini et al., 2015). On the one hand, it seems plausible that having a larger working memory capacity provides a bigger “window” to encode and bind stimuli together across a temporal sequence that could subsequently improve learning of the contained regularities (Janacsek, and Nemeth, 2013). However, the empirical findings do not consistently demonstrate a functional relationship between working memory capacity and sequence learning ability as measured by the SRT task (Janacsek, and Nemeth, 2013). One possible reason for this is that if one takes a multi-component view of statistical learning (e.g., Arciuli, 2017; Daltrozzo and Conway, 2014), then each separate component in the system may depend upon attention or working memory to different degrees. For instance, the evidence appears to suggest that working memory may be more closely related to explicit forms of sequence learning compared to implicit forms of learning, as argued by Janacsek and Nemeth (2013). Likewise, the construct of working memory is multi-faceted, so different aspects of working memory may be more or less important for statistical learning. For instance, visual-spatial working memory may be closely tied to performance on statistical learning tasks that require visual-spatial encoding but less so for tasks involving the learning of verbal patterns (Janacsek, and Nemeth, 2013). In addition, for any given learning task, different participants may represent and conceptualize the task differently in terms of how they rely upon verbal, visual, or other types of representations (Martini et al., 2013). This variability in how participants represent the learning tasks could therefore explain the lack of strong correlations observed between working memory capacity and performance on statistical learning tasks.

Further illustrating the complex relationships among these constructs is a study by Hendricks et al. (2013) that attempted to examine the role of working memory in statistical learning. Hendricks et al. (2013) used a concurrent load task in conjunction with an artificial grammar learning paradigm to examine whether the learning of grammatical rules versus chunk-based information was automatic or not (i.e., required working memory resources). The concurrent load task involved participants viewing six random numbers on the screen, maintaining the numbers in memory while they subsequently viewed a trial of letters generated from an artificial grammar, and then finally typing the six numbers from memory. This concurrent load task was given to participants either during the exposure phase of the AGL task, the test phase, or both. Performance was compared to a control group that did the AGL task without having to do the concurrent load task. The results of this study showed that the learning of chunk or fragment-based information could proceed with minimal cognitive requirements (that is, the concurrent load task did not impair performance), suggesting that this form of learning can occur relatively automatically and under incidental learning conditions. On the other hand, the expression of rule-based knowledge (at test following learning), required a certain amount of cognitive resources (that is, the concurrent load task given during the test phase interfered with test performance for rule-based knowledge). These findings were interpreted by suggesting that the learning of fragment or chunk information is mediated by a form of implicit “perceptual fluency” in which perception of items is facilitated via experience (e.g., Chang and Knowlton, 2004), whereas

the learning and expression of rule-based regularities was an explicit process involving something akin to “hypothesis generation” (e.g., Dulany et al., 1984).

Note, this conjecture appears to contradict earlier work suggesting that chunk learning and rule-learning occur via declarative and procedural memory, respectively (Lieberman et al., 2004). Briefly, declarative memory refers to the recall and recognition of facts and events (Squire, 2004) whereas procedural memory is a type of nondeclarative and largely implicit form of learning (Ullman, 2004). The relationship between statistical learning and these two other forms of memory will be taken up in section 3.5. For now, it is important to point out that the findings from Hendricks et al. (2013) and Lieberman et al. (2004) are not necessarily contradictory of one another, as it is possible that chunk learning can proceed via multiple routes, using either perceptual-based or declarative memory-based forms of encoding. Likewise, rule-based learning similarly may rely on either procedural memory or hypothesis-generation depending on the particular task, learning context, or individual. As mentioned earlier, note that “rules” in the present case refer to any information in the stimulus sequences that denote grammaticality apart from bigram and trigram information, such as positional regularities (e.g., what stimuli are allowed in different positions of a sequence) or possibly even nonadjacent regularities as dictated by the grammar.

Interestingly, the concurrent load task also interfered with performance in a transfer condition in which the underlying rules were consistent but the stimulus set was changed (Hendricks et al., 2013). That is, knowledge of the underlying grammatical regularities could be transferred to a non-trained letter set but only if there were sufficient cognitive resources available during test (i.e., only if there was not a concurrent load task). It appears then that some aspects of statistical learning require attention / working memory (e.g., using hypothesis-generation strategies, expressing rule-based knowledge at test, and transferring knowledge to novel stimulus domains), whereas others appear to be automatic (e.g., perceptual fluency of chunk-based information). However, it should be noted that it is not perfectly certain that the Hendricks et al. (2013) concurrent load task completely eliminated attentional resources; some amount of attention may still have been available during learning.

Another way to manipulate attention is by capitalizing on its selective nature. Turk-Browne et al. (2005) did so by creating two interleaved streams of differently colored visual regularities and then instructing participants to detect repetitions in one stream but not the other. Across several experiments, Turk-Browne et al. (2005) determined that learning of the statistical regularities only occurred for the attended stream, not for the unattended stream. They concluded that visual learning of sequential regularities both is and is not automatic: it requires attention in the sense that the regularities are only learned if the stimuli are selectively attended; but learning is automatic in the sense that it can occur incidentally (i.e., in the face of a cover task that provided no information about the presence of regularities) and does not necessarily result in conscious awareness of what was learned. Selective attention therefore may act as a “gate” for statistical learning, at least for certain learning situations and task paradigms (e.g., Baker et al., 2004; Emberson et al., 2011; Toro et al., 2005; Turk-Browne et al., 2005).

Interestingly, there appears to be a reciprocal relationship, in which learning itself can modulate attention (Alamia and Zénon, 2016; Hard et al., 2018; Zhao et al., 2013). That is, attention affects learning by facilitating encoding of particular aspects of the input; and yet, learning itself can affect attention, for instance by creating a “pop-out” effect, drawing (exogenous) attention to input that violates the expectations that have been generated based on previous experience (Kristjansson et al., 2007). In support of this idea, Sengupta et al. (2018) recently found that functional connectivity between brain networks supporting attention and working memory processes changed following exposure to the statistical regularities presented in an artificial language.

Attention also plays a crucial role in the framework of de Diego-Balaguer et al., 2016d, in which top-down control of attention (i.e., endogenous attention) is a prerequisite for learning nonadjacent but not adjacent sequential dependencies. Consistent with such a dissociation are findings from Romberg and Saffran (2013). They constructed artificial languages in which the first and third items of 3-word phrases had nonadjacent deterministic relationships while the intervening elements had adjacent probabilistic relationships with the surrounding two items. Although adults were able to demonstrate learning of both the adjacent and nonadjacent dependencies, higher confidence ratings on nonadjacent trials were associated with higher accuracy, while greater confidence on adjacent trials was not associated with greater accuracy. This may suggest that learning of the nonadjacent dependencies occurred explicitly while learning of adjacent dependencies was accomplished by more implicit means. Interestingly, Turk-Browne et al. (2005) pointed out that, because their design involved two interleaved streams of stimuli, learning was occurring in many cases over intervening items, thus requiring learning of nonadjacent dependencies. Together these findings are consistent with the idea that selective attention perhaps is most needed for learning nonadjacent dependencies across a temporal stream (Diego-Balaguer et al., 2016).

A different perspective, however, stresses not only the *necessity* of attention for pattern learning, but also its *sufficiency* (Pacton and Perruchet, 2008; Perruchet and Vinter, 1998). Under this view, the learning of patterns in input is a natural consequence of attentional processing due to the laws of memory and associative learning. From this perspective, attention may be necessary not only for learning nonadjacent but also adjacent dependencies (Pacton and Perruchet, 2008). However, as reviewed previously, it appears that adjacent-item chunks can also be learned without the availability of attentional resources (Hendricks et al., 2013).

Other studies are consistent with the notion that while some aspects of learning require attention or intention to learn, other aspects of learning can indeed proceed automatically and under incidental conditions. For instance, Bekinschtein et al. (2009) (see also Wacongne et al., 2011) used a local-global paradigm in which auditory sequences of pure tones that contained either local (within-sequence) or global (across-sequence) violations were presented. For instance, in “XXXXXY” the “Y” is a local violation in this sequence because it violates the expected pattern of “X”s. However, after repeated exposures of “XXXXXY”, if the sequence “xref” is then encountered, the final “X” in the sequence becomes a global violation because the “Y” is expected after exposure to the “XXXXXY” sequences. Bekinschtein et al. (2009) found that the local deviants were processed automatically and

non-consciously; these types of violations were immune to attentional manipulations and were even elicited in coma patients as measured by event-related potentials (ERPs). On the other hand, the global deviants required explicit or controlled processing: they were accompanied by conscious awareness in healthy participants; in coma patients the ERPs related to these types of deviants were not detected.

Thus, it appears likely that both “implicit” (i.e., attention-independent / automatic) and “explicit” (i.e., attention-dependent) learning processes operate alongside each other. Such “dual-theory” approaches are common in the literature. For instance, Dale et al (2012) proposed that during a learning episode, implicit associative or reactive learning occurs initially, which leads to the formulation of predictive “wagers” that steadily become more correct and that in turn lead to explicit awareness of the learned patterns. This perspective is also consistent with research using a predictor-target paradigm in which visual “target” stimuli are predicted to varying degrees by “high” or “low” predictor stimuli; although learning is incidental, over the course of the experiment, adults and children display the emergence of a P300-like ERP component elicited by the high predictor stimulus (Jost et al., 2015). This ERP component is strongly related to participants’ conscious awareness of the predictor-target contingency (Singh et al., 2017). This attention-based ERP component is distinct from participants’ learning as assessed through reaction times, which appeared to be indexing learning of the contingencies occurring outside attention and awareness (Singh et al., 2018).

Similarly, Batterink et al. (2015) proposed that implicit and explicit learning systems operate in parallel, with the implicit system more or less always engaged but the explicit system optional. They suggested that in the standard familiarity task often used in statistical learning research, the familiarity judgement reflects explicit knowledge but that implicit learning can also be displayed and measured indirectly using reaction times or possibly ERPs. Another dual-system approach is that of Keele et al. (2003), who proposed a theoretical perspective based on a review of findings from the SRT task. In their view, a dorsal neural system mediates implicit learning of unimodal or unidimensional stimuli, whereas a ventral system mediates the learning of cross-modal or cross-dimensional input, which can involve both implicit and explicit learning mechanisms. This last tenet is consistent with Walk and Conway (2016) who proposed that implicit learning is sufficient for learning unimodal sequential regularities (i.e., sequential dependencies between items in the same perceptual modality) but that additional cognitive resources such as selective attention or working memory may be required to learn cross-modal sequential patterns. Similarly, Daltrozzo and Conway (2014) also proposed a two-system view of pattern learning: a bottom-up implicit-perceptual learning system that develops early in life and encodes the surface structure of input; and a second system that is dependent on attention, develops later in life, and relies to a greater extent on top-down information to encode and represent more complex patterns.

Finally, insight can also be gained from a related though somewhat distinct research literature on category learning. Smith and colleagues (Smith and Grossman, 2008; Smith et al., 1998) proposed that there are multiple types of category-learning systems: rule-based and similarity-based. They proposed that rule-based category learning involves selective attention and working memory processes to enable a decision to be made about whether an

item belongs to a particular category. On the other hand, similarity-based categorization processes can be mediated by the involvement of both explicit and implicit learning processes. The implicit learning system involves processes such as perceptual fluency and perceptual priming; that is, one decides whether an item belongs to the category in question in terms of the ease with which the perceptual features of the item can be processed (Smith and Grossman, 2008).

Taken together, we believe the evidence supports the idea that statistical learning reflects both implicit / automatic and attention-dependent / explicit aspects of processing. The attention-dependent learning system shares similarities with Baars (1988; 2005) global workspace theory of consciousness, in which consciousness is construed as a limited capacity attentional spotlight that “enables access between brain functions that are otherwise separate” (Baars, 2005, p.46). What determines the mode of learning (explicit vs. implicit) likely depends at least in part on the type of input to be learned; some types of structures appear to require attention to adequately process and encode the patterns, such as nonadjacent dependencies (de Diego-Balaguer et al., 2016d), global patterns (Bekinschtein et al., 2009), cross-modal dependencies (Keele et al., 2003), and rule-based processing (Hendricks et al., 2013; Smith et al., 1998). Other factors that may affect the involvement of automatic versus attention-dependent mechanisms include whether learning is assessed through the use of direct/explicit judgments versus indirect measures such as reaction times (Batterink et al., 2015) and whether learning requires generalization or transfer to new stimulus sets (Hendricks et al., 2013). We furthermore propose that the automatic learning system is “obligatory” in the sense that it is always active, whereas the attention-dependent system is optional and is only engaged when selective attention and working memory are brought to bear on the learning task (Batterink et al., 2015) via the involvement of endogenous or exogenous attentional mechanisms. It is also possible that the involvement of one or both systems is not an “either-or” phenomenon but may be graded; as learning proceeds, exogenous attention can be increasingly drawn to the regularities in question (Alamia and Zénon, 2016), which would necessarily gradually activate the attention-dependent learning system. Thus, the involvement of automatic versus attention-dependent learning mechanisms could likely change across a learning episode or across multiple episodes.

### 3.4. An interim summary

Based on the preceding three sections, it is clear that statistical learning: 1) consists of both modality-specific and domain-general learning mechanisms; 2) can be used to learn patterns along a continuum of complexity from relatively simple to more complex structures; 3) and involves both implicit / automatic as well as explicit / attention-dependent modes of learning. We furthermore propose that these three factors “line up” so to speak, suggesting the involvement of two primary modes of learning (see Fig. 3). Specifically, statistical learning is mediated through the functioning of at least two (or more) distinct processing mechanisms. The first is the classic “implicit” learning system, that proceeds automatically and with minimal attentional requirements, is likely a perceptual-based process, and can mediate the learning of local, unimodal, and associative-based patterns. The second is an explicit, attention-dependent system that is necessary for learning nonadjacent, global, and



crossmodal dependencies; it is also needed for when transferring learning to new stimulus sets or contexts. These two systems likely operate in parallel with each other (Batterink et al., 2015), and each can be more or less engaged depending on the learning requirements and situation (Daltrozzo and Conway, 2014).

This dual-system approach has similarities to the distinction between “model-based” and “model-free” reinforcement learning (e.g., Savalia et al. 2016; Kurdi et al., 2019) which describes learning that is goal-directed, flexible, and reliant upon long-term knowledge (i.e., model-based learning) versus learning that is data-driven, automatic, and relatively inflexible (i.e., model-free learning). Together, these two types of learning provide complementary ways to best learn about and interact with the environment.

However, it should be pointed out that the two systems may possibly operate competitively, rather than independently of or in cooperation with one another, as has been suggested up to this point. For instance, some studies have shown that executive control processes may have an antagonistic relationship with implicit pattern learning (Ambrus et al., 2019; Filoteo et al., 2010; Nemeth et al., 2013; Tóth et al., 2017; Virag et al., 2015). Virag et al. (2015) observed a negative correlation between executive functions and implicit learning as measured by a variant of the SRT task. And Nemeth et al. (2013) used hypnosis to reduce explicit attentional processes in their subjects, which resulted in improved learning on the SRT task. At present it is not clear under what conditions these two systems act synergistically versus antagonistically but one possibility is that it is due to the task requirements. Most of the studies cited above that observed a competitive relationship used the SRT task, which differs in a number of important respects with other statistical learning tasks such as the segmentation task or AGL task. One important difference is that the SRT task requires a motor response on each trial, whereas the other two tasks have a passive exposure phase that involves perceptual learning or memory-based encoding without a motor response. It is possible that top-down attentional control interferes with the type of trial-by-trial stimulus-response learning that the SRT task elicits; it is currently an open question whether this holds true for learning during other types of statistical learning tasks that do not involve the same type of stimulus-response learning.

In the next section (3.5), we review the neuroanatomical bases of statistical learning through the lens of the operation of these two proposed learning systems. Then, we examine how each type of learning system might change across human development and may differ across phylogeny (sections 3.6 and 3.7) before concluding with a summary of ten core principles that flesh out the neurocognitive mechanisms underlying statistical learning (section 4).

### 3.5. What are the neuroanatomical bases of statistical learning?

Brain areas that have shown significant activation during different types of statistical learning and implicit learning tasks include practically the entire brain, including: perceptual regions (e.g., Turk-Browne et al., 2010), parietal cortex (e.g., Forkstam et al., 2006), prefrontal cortex and Broca’s area specifically (e.g., Ablá and Okanoya, 2008), as well as subcortical regions such as the hippocampus (and medial temporal lobe, MTL) (e.g., Schapiro et al., 2014), and basal ganglia (e.g., Karuza et al., 2013). Rather than reviewing all of the available evidence in detail, we instead focus on theoretical perspectives that can help

explain why certain brain regions may or may not be active depending on the task or situation. For added focus, we mainly discuss the neocortical bases of statistical learning, while still acknowledging the important role played by subcortical structures such as the hippocampus, cerebellum and basal ganglia (see Batterink et al., 2019). At the end of this section we then discuss how the neocortical systems proposed here interact with “classic” learning and memory systems (i.e., declarative and procedural memory), thought to be mediated largely by subcortical structures.

One perspective that is consistent with the neural findings showing multiple brain regions involved with statistical learning is P.J. Reber’s (2013) proposal that implicit learning reflects a general principle of plasticity of neural networks that results in improved processing. That is, learning is an emergent property of neural plasticity that is pervasive and universal, not localized to a particular brain region, nor confined to any specific task, but contributes to cognition and behavior very broadly. Under this view, implicit learning cannot be defined exclusively by whether or not it involves for instance the MTL or conscious awareness; instead, it reflects the gradual tuning of neural networks and synapses to adapt to statistical structure encountered in the environment. Such neural plasticity and tuning generally is associated with reduction of neural activity that reflects increased processing efficiency (P.J. Reber, 2013).

Under this “plasticity of processing” perspective of statistical learning, the areas of the brain that will reflect learning are those same areas involved in processing the input in question. Thus, it is perhaps not surprising that perceptual regions of the brain are implicated in statistical learning (Turk-Browne et al., 2010), as perceptual processing is necessary in order to encode the stimuli in the first place. Note, perceptual areas have shown activity reflecting not just perception of the individual stimuli, but *learning* of the patterns themselves. But what about studies showing activity in frontal, parietal, subcortical, and other areas? It is likely that general constraints on processing in different neural networks determines which brain areas will reflect learning. There appear to be two primary sets of cortical regions involved in implicit learning of sequential structure (Conway and Pisoni, 2008): sensory/perceptual regions as already discussed, but also frontal regions such as the prefrontal cortex (PFC) that have connective loops with subcortical networks including the basal ganglia and cerebellum. For tasks involving processing of sequential (i.e., temporal) structure in particular, working memory and selective attention are likely necessary, which in turn relies on PFC and associated brain networks. Thus, these two systems – perceptual and frontal – together constitute a dynamic and adaptive cortical network that is used to perceive, encode, and adapt to most types of input patterns encountered in the world.

The distinction between frontal “executive” cortical regions and posterior “perceptual” regions is nicely summarized by Fuster and Bressler (2012), who argued this dichotomy reflects a general characteristic of neural functioning, with frontal areas needed for actions (e.g., behavior, language) as well as higher-level planning, and posterior regions involved in sensory, perceptual, and memory operations. Under this view, all aspects of cognition involve the operation of large-scale cortical networks, not modular regions, and in particular involves the interaction of the posterior and frontal systems. Lateral PFC is argued specifically to be crucial for the temporal organization of behavior (Fuster, 2001).

Furthermore, working memory neuroimaging studies generally show lateral PFC involvement in conjunction with posterior areas that vary on the sensory modality of the particular input type that is encountered: “If the memorandum is visual, that posterior region includes inferotemporal and parastriate cortex ..... if it is auditory, superior temporal cortex; if it is spatial, posterior parietal cortex” (Fuster and Bressler, 2012, p.215). Integrating Fuster and Bressler’s (2012) view with P.J. Reber’s (2013) leads to the following conclusion: items encountered in a temporal sequence necessarily recruit PFC as well as sensory/posterior regions (the exact sensory region active depending on the input modality) in order to process the sequence; if that particular type of sequence is encountered repeatedly, containing structural regularities, then the networks involved in processing these sequences (i.e., PFC and perceptual regions) will show plasticity and tuning, resulting in learning of the underlying structure.

In a similar perspective, Hasson et al. (2015) noted that virtually all cortical circuits can accumulate information (i.e. learn) over time, but that timescales vary hierarchically in the brain: lower sensory areas can only process information on the order of 10 s to 100 s of milliseconds, whereas higher-order areas can process information over much longer timescales (many seconds or minutes) (see also Farbood et al., 2015; Kiebel et al., 2008). This appears to be due to the hierarchical arrangement of neural systems. That is, lower-order sensory areas respond to relatively simple features (such as single tones or lines of particular orientations), whereas higher-order areas integrate across this information to represent increasingly complex stimuli (such as speech or faces). This same general “rostro-caudal” framework appears to apply to temporal dynamics as well, in which timescales of representation generally increase as one moves from lower sensory areas to higher-level frontal areas (Kiebel et al., 2008).

Hasson et al. (2015) furthermore argued against a memory versus processing distinction; instead, in their view, prior information continuously shapes processing in the present moment, very similar to P.J. Reber’s (2013) view of implicit learning consisting of cortical tuning of processing networks. In addition, Hasson et al. (2015) argued for the existence of modulatory circuits: “attentional control processes supported by fronto-parietal circuits” (related to traditional working memory operations), “and binding and consolidation processes supported by [medial temporal lobe] circuits (related to episodic memory)”. Thus, the general principle of cortical plasticity is constrained by differences in processing characteristics of different areas of the brain (e.g., short vs. longer timescales) but is also modulated by attentional control, working memory, and consolidation processes. We thus suggest there are at least two primary neurocognitive (primarily, cortical-based) mechanisms that embody statistical learning: 1) gradual tuning of cortical networks based on experience (i.e., cortical plasticity); and 2) top-down modulatory control mechanisms that guide selective attention and working memory, which is especially needed for learning patterns that require integration of information across time (i.e., statistical patterns in temporal sequences).

As reviewed in section 3.1, statistical learning appears to be partly, and perhaps largely, based on perceptual processing mechanisms (Conway and Christiansen, 2005; Frost et al., 2015). However, based on the review in this section so far, we now know there are at least

two reasons why sensory/perceptual regions cannot mediate all aspects of statistical learning. One is that due to the sizes of their temporal receptive windows, sensory areas cannot process information that spans longer than the order of milliseconds. Thus, learning sequential patterns over a temporal sequence, especially for long-distance or nonadjacent dependencies, cannot occur in these perceptual processing brain regions, but must rely on downstream networks including frontoparietal networks and perhaps PFC specifically. Second, the PFC and related frontoparietal networks appear to modulate learning in any given situation, even if these frontal regions don't reflect cortical tuning and plasticity themselves. For example, through frontoparietal network involvement, attention to particular stimuli may occur, directing perceptual processing regions to then engage with those inputs, which over the course of repeated experience, results in those perceptual regions exhibiting neural plasticity and learning. It is also likely that the frontoparietal networks themselves may show neural tuning and plasticity with exposure, allowing learning itself to modulate attention, as reviewed in section 3.3.

This framework where learning (perhaps of sequential input specifically) might involve a combination of higher-level frontal areas as well as lower perceptual regions, is consistent with other relevant theoretical perspectives. For instance, Uhrig et al. (2014) provided evidence consistent with the idea that learning sequences occurs at hierarchical levels in the brain: lower/modality-specific areas are independent of attention and can mediate "local" processing operations whereas higher levels are attention-dependent and are needed for more "global" processing. Thothathiri and Rattinger (2015) further argued that frontal areas and controlled processing are necessary for sequence processing specifically. Their argument, based on a review of both neuroimaging and neuropsychological studies and focusing on sequence production, is that sequencing involves cognitive control (the ability to order items, reject incorrect items, resolve interference, and choose the correct item to produce). Therefore, the frontal lobe (specifically, left ventrolateral PFC) is needed for sequencing because cognitive control functions are necessary for selecting the correct stimulus among various alternatives in a sequence.

Cognitive control and selective attention are likely important not just for sequence *production* but also for sequence learning, especially for long-distance or nonadjacent dependencies. As mentioned earlier, de Diego-Balaguer et al., 2016d proposed that in order to learn a sequential nonadjacent dependency, cognitive control is needed to inhibit processing of intervening items occurring between the nonadjacent dependencies and to focus on the long-distance dependency itself. Indeed, areas of the PFC such as left inferior frontal gyrus (LIFG, or Broca's region) have often been implicated in sequence learning that specifically involves structures that contain long-distance regularities (e.g., Bahlmann et al., 2008; Friederici et al., 2006). In fact, LIFG has been proposed to be a "supramodal hierarchical processor" (Tettamanti and Weniger, 2006). It is possible that the reason that LIFG may appear to be necessary for hierarchical operations is that hierarchical dependencies necessarily require the processing (and learning) of long-distance dependencies, which relies on PFC involvement.

Finally, a crucial aspect of statistical learning appears to be prediction and expectation (Dale et al., 2012), which is mediated by both sensory and downstream areas such as the PFC

(though it is possible that the PFC generates the predictions and subsequently modulates sensory areas, Bubic, 2010). In particular, temporal sequencing appears to be an area where prediction is most important due to the prominent role of time and uncertainty (Bubic, 2010). An important part of making predictions of upcoming events is the necessity of inhibiting the representation of events or stimuli that are not predicted, which likely involves PFC (Bar, 2009). However, predictive processing appears to be inherent to all levels of the hierarchically organized nervous system (Friston, 2005) and thus appears to go hand-in-hand with a “plasticity of processing” approach. Furthermore, Huettig (2015) suggested a dual-system account of predictive processing similar to our proposal of statistical learning consisting of an implicit, automatic processing system and an explicit, attention-dependent one.

To summarize, there are a number of considerations that can illuminate why certain brain regions show consistent activation in studies of statistical learning. Under a “plasticity of processing” approach (P.J. Reber, 2013), whatever neural substrate is involved in processing the input in question, through repeated exposure and experience, becomes tuned through general principles of neural plasticity to become more efficient at processing that type of stimulus, resulting in lower levels of neural activation. This explains why it is common to observe a variety of distributed neural regions active for different kinds of tasks and input types, such as auditory and visual processing regions during auditory and visual statistical learning tasks, respectively. In addition, the brain is organized hierarchically, with upstream brain regions showing relatively short temporal receptive windows and downstream areas (such as PFC) showing the largest temporal receptive windows. This acts as a further constraint on processing: for sequences and especially long-distance or global dependencies, only brain regions with temporal receptive windows that are large enough to process the stimuli across longer periods of time will reflect learning of such dependencies. Finally, in addition to the general mechanism of cortical plasticity, the involvement of PFC and frontoparietal networks can act as a modulatory mechanism on learning, providing top-down control of attention and cognitive control, which can affect and direct learning, especially for more complex patterns such as hierarchical or long-distance dependencies. This distributed versus specialized dichotomy appears to map loosely onto the “implicit / automatic” and “explicit / attention-dependent” distinction outlined in section 3.4 and Fig. 3, with cortical plasticity instantiated in lower perceptual regions reflecting attention-independent, automatic implicit learning mechanisms, and downstream brain regions reflecting attention-dependent specialized functions needed for processing and learning certain aspects of structural regularities, especially those that require integration over longer periods of time.

However, as reviewed earlier in section 3.4, some research suggests that attention-independent and attention-dependent systems might operate antagonistically, rather than synergistically. That is, frontal-based executive and cognitive control functions might operate competitively with more implicit forms of learning (e.g., Nemeth et al., 2013). For instance, Tóth et al. (2017) used EEG to measure functional connectivity during implicit sequence learning with the SRT task. They found that learning performance was negatively correlated with functional connectivity in anterior sites, which they proposed suggests that top-down attentional control interferes with automatic, implicit learning of the visual-motor sequences. More recently, Ambrus et al. (2019) investigated the relationship between these

two systems using inhibitory transcranial magnetic stimulation (TMS) on the dorsolateral prefrontal cortex (DLPFC) while participants engaged in the SRT task. The results revealed that disrupting this area of the frontal lobe resulted in better learning of nonadjacent dependencies. Thus, these findings appear to show that at least for the SRT task, frontal-mediated executive and attentional functions act antagonistically with implicit learning, with the latter improving when the former are weak or disrupted.

The framework proposed here has been focused mainly on neocortical processing mechanisms. However, subcortical structures such as the hippocampus, cerebellum, and basal ganglia clearly play a central role in learning and memory more generally, and perhaps statistical learning specifically. For instance, the cerebellum is known to play an important role in associative motor learning (Steinmetz, 2000) but also possibly non-motor learning and other cognitive functions (e.g., Desmond and Fiez, 1998; Ivry and Baldo, 1992; Timmann et al., 2010). The classic memory systems view holds that declarative memory – which refers to the recall and recognition of facts and events – depends on the hippocampus and MTL (e.g., Squire, 2004). Procedural memory on the other hand – a type of nondeclarative and largely implicit form of learning – relies specifically on the basal ganglia, though the cerebellum also appears to play a role (Ullman, 2004; Ullman et al., 2020). These two forms of memory likely are both involved during statistical learning (Batterink et al., 2019; Sawi and Rueckl, 2019), possibly in a competitive manner. The MTL and basal ganglia often show competitive interactions, which may be modulated by the PFC (Poldrack and Rodriguez, 2004). One way to unite the cortical perspective presented above with the workings of subcortical structures is to take a complementary learning systems approach (e.g., O'Reilly and Norman, 2002). Under this view, there is a trade-off between different types of learning and memory, necessary in order to achieve different goals and to meet certain demands, which is best handled by functional specialization of brain regions. For instance, the hippocampus is well-suited for rapidly encoding arbitrary associations and memories of specific events, while the neocortex can handle slowly developing representations of the general statistical structure of the environment (O'Reilly and Norman, 2002). Atallah et al. (2004) proposed a tripartite model consisting of the hippocampus (for rapid learning of specific events and details), posterior neocortex (for learning general statistical information about the environment), and the PFC (with connections to the basal ganglia, for maintaining information in an active state). Together, these three brain systems can support different types of behavioral functions, with each brain area satisfying different kinds of demands. The cortical model proposed above encompasses two of the three components of this tripartite model (the posterior neocortex and PFC), but we acknowledge the recent work suggesting that the hippocampus also plays a role in statistical learning and needs to be integrated into such a model (Schapiro et al., 2014). More work is needed to outline the exact interactions between the cortical systems outlined here and the other (subcortical) brain systems underlying learning and memory more generally.

### 3.6. How does ontogeny constrain learning?

Most aspects of sensorimotor, cognitive, and social functioning increase from childhood to adulthood (Plebanek and Sloutsky, 2017). Is the same true for statistical learning? At least some aspects of statistical learning, such as the learning of adjacent transitional

probabilities, are present from very early in development. For instance, auditory learning of adjacent transitional probabilities is clearly available by 8-months (Saffran et al., 1996) and possibly even at birth (Teinonen et al., 2009). Similarly, visual adjacency learning has been demonstrated at 2 months (Kirkham et al., 2002) and in newborns (Bulf et al., 2011). Similarly, the detection of adjacent-item repetition structure also appears to be available very early in life (Endress et al., 2009; Gervain et al., 2008). This evidence of some aspects of statistical learning developing very early is consistent with A.S. Reber's (2003) classic view of implicit learning as being an invariant ability, present across all (typically developing) individuals, with very little individual variation or change across development (Amso and Davidow, 2012; Jost et al., 2015).

On the other hand, other evidence points to a more complex developmental picture. When visual input sequences were created that allowed separate investigation of co-occurrence frequency information and transitional probabilities, a developmental progression was found, with 2.5-month-olds showing sensitivity to co-occurrence frequency only, but 4.5-month-olds and older infants showing sensitivity to transitional probabilities as well (Marcovitch and Lewkowicz, 2009). Similarly, 5-month-olds were able to segment visual sequences that contained redundant co-occurrence frequency and transitional probability cues, but 2-month-olds were unable to do so (Slone and Johnson, 2015). Thus, potentially important changes occur in infants' capacity to track statistical patterns in visual sequential input between 2 and 5 months of age, likely due to developmental changes to attention and memory (Slone and Johnson, 2015).

Complicating the developmental findings is that, as reviewed above, statistical learning itself appears to be a heterogeneous construct (e.g., Arciuli, 2017; Daltrozzo and Conway, 2014; Thiessen and Erickson, 2013). If there are multiple neurocognitive processes underlying statistical learning, then each one may be governed by different developmental constraints, leading to different patterns of development depending on what aspect of statistical learning is being measured in a given study. Thus, taking a multiple-systems approach to understanding statistical learning may provide some clarity. As an example, Janacek et al. (2012) examined age-related changes in statistical learning in over 400 individuals between 4 and 85 years of age using the SRT task. They found that 4–12-year-olds had the greatest learning effects as measured by RTs, with a dramatic decrease in learning ability around 12 years that continued to decline across the lifespan. However, accuracy scores were lowest in the children and elderly participants with highest scores at the middle ages. Janacek et al. (2012) suggested that these findings may be the result of their measures tapping into two separate learning systems, with accuracy related to voluntary attentional control (an underdeveloped executive function mechanism in early childhood) and RT related to involuntary mechanisms. Other evidence for different developmental trajectories for different aspects of learning includes the distinction between nonadjacent and adjacent dependencies, with the former being not as easily learned early in development as the latter (e.g., Gervain et al., 2008). In de Diego-Balaguer et al. (2016d) view, the learning of nonadjacent dependencies develops later in childhood, only when endogenous attentional control is mature, which they propose is needed to learn these types of dependencies. Thus, the type of input pattern appears to interact with age, with some types of structures learnable early in development

but others requiring the development of attention and memory mechanisms to support such learning.

Interestingly, input modality might also interact with age to determine learning success. Visual statistical learning (as measured by a variation of the classic triplet segmentation task) was found to have a gradual developmental progression between 5 and 12 years of age (Arciuli and Simpson, 2011; Raviv and Arnon, 2017). On the other hand, auditory learning showed no such age differences, at least within this age range (Raviv and Arnon, 2017). A subsequent follow-up study showed that the apparent modality differences might be driven by the specific nature of the stimuli in terms of being composed of linguistic stimuli (i.e., syllables) or not (Shufaniya and Arnon, 2018). That is, the most recent evidence suggests that whereas statistical learning of nonlinguistic auditory and visual input may show a steady increase with age, auditory learning of linguistic materials may be developmentally invariant (Raviv and Arnon, 2017; Shufaniya and Arnon, 2018). It is important to realize, however, that there may be other important developmental changes occurring earlier in life, before age 5, that are not captured by these two studies.

Paradoxically, some aspects of statistical learning might actually be more efficient earlier in development, when cognitive abilities such as top-down attentional control and working memory have not yet reached mature levels of ability (Thompson-Schill et al., 2009). For instance, Plebanek and Sloutsky (2017) showed that 4- and 5-year-old children outperformed adults on a change-detection and a visual search task. They suggested that it was because children at that age tend to distribute attention across multiple aspects of stimuli, even when it is not relevant to the goal. This more distributed attention resulted in better processing of task-irrelevant information, which allowed the children to perform better on the change-detection and visual search tasks. Similarly, Juhasz et al. (2019) found that young children showed superior learning on the SRT task relative to adolescents and adults (note that this study also took into account the average response speed differences between age groups, an important methodological point for any developmental study that uses response times as the measure of learning). The idea that cognitive limitations early in development may confer a computational advantage for learning is not new (e.g., Elman, 1993; Newport, 1990). In general, it could be evolutionarily adaptive for organisms to have more efficient and flexible learning mechanisms early in development (e.g., Johnson and Wilbrecht, 2011). This “less is more” proposal fits nicely with the theoretical framework offered by Ambrus et al. (2019) and related studies suggesting that top-down executive control (instantiated in frontal-based neural circuits) may impede (implicit) statistical learning. Under this framework, the reason that young children perform better than adults on statistical learning is that their PFC is under-developed, which allows for unhindered bottom-up, data-driven learning of environmental patterns (Ambrus et al., 2019).

As argued above, one general mechanism that underlies statistical learning is cortical plasticity (P.J. Reber, 2013). The ability for cortical networks to adapt and modify themselves based on environmental experience appears to be an intrinsic property of neural networks, present across the lifespan (Pascual-Leone et al., 2011). However, it is also clear that in general, neural plasticity declines with age (Kleim and Jones, 2008; Pascual-Leone et al., 2011). Furthermore, different neural systems may have different degrees of plasticity at



different points in development. For instance, different brain areas have different timescales of synaptic proliferation and pruning, with motor and sensory processing areas maturing first, followed by spatial and language processing regions (parietal lobe), and executive functions (frontal lobe) developing last (Gogtay et al., 2004). Furthermore, although there is a general trend for reduced plasticity with age, there appears to be a substantial amount of individual variability, with different individual brains having different “starting points” of plasticity as well as having different “slopes of change”, due to variations in genetic and environmental factors (Pascual-Leone et al., 2011). The environment plays a key role in dictating changes in plasticity due to the principle of neural commitment: as learning and experience with the environment progresses, neural networks become entrenched and tuned to the particular patterns of information experienced, making further plasticity-related changes more difficult (Kuhl, 2004). This mechanism of entrenchment has been argued to be a major factor giving rise to the existence of sensitive periods in language and other cognitive and perceptual domains (Kuhl et al., 2005; Meltzoff et al., 2009).

Thus, cortical perceptual plasticity changes over development (White et al., 2013). Early in development, plasticity is driven primarily by bottom-up (implicit) learning. Later in development (after the sensitive period ends), plasticity becomes increasingly reliant on top-down factors, such as knowledge of higher order representations and categories gained through experience, which directs attention to particular features or types of input. In most cases, learning and plasticity is achieved through the interaction of these two processes. However, because of neural commitment and cortical maturation, there is a gradual developmental decline in the extent that bottom-up processes impact plasticity; at the same time, top-down influences such as selective attention increasingly modulate the capacity for cortical plasticity (Kral and Eggermont, 2007; White et al., 2013). A key developmental shift may occur around the age of 4 years (Mueller et al., 2018), from automatic associative learning that dominates infancy, to attention-guided frontal cortex-based mechanisms that guide learning (see also Deocampo and Conway, 2016).

Thus, the developmental trajectory of statistical learning appears to be due to the functioning of at least two interacting mechanisms. Early in development, statistical learning is driven almost entirely by bottom-up, automatic, associative learning mechanisms that reflect principles of cortical plasticity, in which neural networks slowly become attuned to environmental regularities with experience. Later in development, as selective attention and other executive functions mature, learning is increasingly modulated by such abilities, allowing for the learning of more complex input patterns such as nonadjacent dependencies and other types of structural patterns that can be considered to be more global. However, increased reliance on top-down control in learning, and a concurrent reduction in cortical plasticity, may not always be beneficial, leading to situations that paradoxically result in poorer learning by adults relative to children (e.g., Ambrus et al., 2019; Plebanek and Sloutsky, 2017). Likewise, in old age, as executive functions begin to show decline, it is expected that statistical learning will worsen, as both bottom-up and top-down mechanisms will be less effective.

In sum, it is argued that statistical learning is influenced by two different types of mechanisms that change over developmental time (see Daltrozzo and Conway, 2014). It

should be pointed out that the bottom-up learning system actually consists of multiple sub-systems (i.e., visual, auditory, motor, etc.), with each neural sub-system having different developmental trajectories in terms of cortical plasticity and maturation. More research is needed to investigate the changes that occur in statistical learning across development, with careful attention paid to the factors reviewed so far (e.g., input modality, input complexity, and the role of attention) and how they impact different components of learning. Missing, too, from this view is the role of consolidation in learning and how that might change with age. For example, Adams et al. (2018) suggested that the ability to “off-load” information from the focus of attention into long-term memory might improve with age; such developmental changes would likely impact the efficiency of statistical learning as well. Thus, taking into account the multiple processes that influence statistical learning will help illuminate the complex developmental picture. Importantly, measures of learning need to be developed that tap into each of the different purported aspects of statistical learning (Arciuli and Conway, 2018) in order to track how each relevant process changes across the lifespan.

### 3.7. How does phylogeny constrain learning?

As with the developmental research, findings related to species comparisons is complicated by the variety of methods and approaches used to assess learning. Based on the review provided to this point, it is proposed that statistical learning in nonhuman species is governed by the operation of at least two partially dissociable learning systems, one based on the principle of cortical plasticity that mediates basic associative learning and perceptual processes, and the other a top-down modulatory “executive” system that directs attention and allows for the learning of more complex patterns. If true, then there are likely areas of overlap across species for evolutionarily-conserved mechanisms, primarily the learning of simple statistical associations mediated by mechanisms of cortical plasticity. In the same respect, it would not be surprising to observe species differences for the learning of more complex patterns such as nonadjacent dependencies or global patterns that require integration over larger timescales, which are proposed to be mediated by top-down cognitive control and attention processes instantiated in the frontal lobe.

Such a distinction was proposed by Conway and Christiansen (2001), who reviewed the extant findings on sequential learning in nonhuman primates, and concluded that all species of primates demonstrate the ability to learn relatively simple patterns (such as repeating sequences and adjacent dependencies) but that species differences are observed in the learning of more complex hierarchical sequential structures that are characterized by nonadjacent dependencies. This framework was based on earlier work, for instance, by Johnson-Pynn et al. (1999), who showed that whereas children 2–3 years of age display hierarchically-based behavioral strategies for organizing nesting cups, three species of nonhuman primates (chimpanzees, bonobos, and capuchin monkeys) do not spontaneously display such complex strategies, relying solely on simpler combinatorial actions. Furthermore, it was argued that these limitations in sequencing abilities could be a key reason for why nonhuman primates do not display human-like language (Conway and Christiansen, 2001). Such a perspective was echoed by Hauser et al. (2002) who argued that what nonhuman animals lack is a narrow faculty of language, specifically the ability to compute complex syntactic hierarchical structures. Thus, in both cases, the argument is that

certain simpler kinds of pattern learning are common across species, but that more complex processing of hierarchical structure containing nonadjacent dependencies is found only in some animal species (including humans).

Since Conway and Christiansen's (2001) and Hauser et al.'s (2002) initial proposals, there have been demonstrations of learning of nonadjacent dependencies in chimpanzees (e.g., Sonnweber et al., 2015) and learning of center-embedded recursive structures in baboons (e.g., Rey et al., 2012). However, Rey et al. (2012) postulated that the baboons were not learning the recursive structure using specialized computational mechanisms as proposed by Hauser et al. (2002), but rather were doing so based on more elementary learning mechanisms such as associative learning and working memory processes. In an attempt to understand the neural basis of auditory sequence learning in rhesus monkeys, Uhrig et al. (2014) incorporated a version of the auditory "local-global" paradigm used in previous human work as described in sections 3.3 and 3.5. They used fMRI to determine that local transitions were mediated by bilateral auditory areas, whereas the learning of global rules showed more distributed activity in downstream prefrontal and parietal areas, similar to the findings found with humans (Bekinschtein et al., 2009). Findings such as these suggest that although sequence processing may be mediated by distinct systems for learning different types of regularities, these systems appear to be present at least in some nonhuman primate species and may be common across human and nonhuman primates.

On the other hand, in a review of the structure of animal communication and learning in artificial grammar studies, Cate and Okanoya (2012) concluded that nonhuman animals' natural productions are only as syntactically complex as finite-state grammars, not more complex structures. Similarly, in a recent comparative investigation of statistical-sequential learning, Rey et al. (2018) showed that both humans and guinea baboons could learn local regularities (i.e., adjacent transitions between sequentially-presented items). However, humans but not baboons were also able to extract the global structure of hierarchically arranged sets of sequences. That is, learning the sequence  $A_1B_1C_1$  involves local, adjacent transitions only. However, learning the arrangement of this sequence  $A_1B_1C_1$  as it occurs in conjunction with other sequences such as  $A_2B_2C_2$  and  $A_3B_3C_3$  involves a more global understanding of the patterns that is not based on local transitions alone.

Clearly, more research is needed to investigate the role of input complexity in statistical learning across species. Wilson et al. (2013) and Stobbe et al. (1598) have both argued for the importance of cross-species research to better understand the origins and emergence of statistical learning and its role in human functions. Petkov and Wilson (2012) suggested that we need better ways to "bridge" findings between humans and nonhumans by using similar methods across species, such as eye-tracking and neuroimaging of nonhuman animals. They furthermore argued that it is important to focus on the learning of finite-state grammars (containing primarily adjacent dependencies) of varying complexity in order to better understand the evolutionary precursors of language and other human skills. As a recent example of such an approach, Heimbauer et al. (2018) used a SRT task with rhesus macaques using a more complex finite-state grammar than typically used with nonhuman primates, in order to probe the limits and extent of learning. Though the monkeys were able to show learning and generalization of this relatively complex grammar, for sequences up to

8 items in length, it took them hundreds of trials over multiple days to do so, whereas humans demonstrated learning of a similarly complex grammar within a single session (Jamieson and Mewhort, 2005).

In addition to input complexity, an area with even less research is the role of input modality on nonhuman learning, as pointed out by Heimbauer et al. (2018) and Milne et al., 2018b; Milne et al., 2018a. As reviewed in section 3.1, there is evidence to suggest that the learning of statistical patterns differs for different perceptual modalities, with generally superior learning observed for auditory compared to visual sequential patterns, at least in humans (e.g., Conway and Christiansen, 2005; Frost et al., 2015). It is not clear as of yet whether nonhuman animals show similar constraints on sensory modality in regard to statistical learning. In one of the few studies to directly compare auditory and visual statistical learning in both humans and nonhumans, Milne et al., 2018b; Milne et al., 2018a showed comparable learning in humans and macaque monkeys. However, other research suggests in fact that nonhuman primates may be better at visual temporal processing compared to auditory processing (Merchant and Honing, 2014). These modality differences were argued to be due to the nonhuman primate brain having impoverished auditory-motor connections in comparison to humans (Merchant and Honing, 2014). Such an inversion of the typical modality effect observed in humans has interesting implications for the evolution of language learning mechanisms, and perhaps could be a contributing factor for why humans but not nonhuman primates show complex (auditory-vocal) linguistic abilities.

Finally, more work is needed to understand the neural underpinnings of statistical learning across species. There have been some recent advances in this regard (e.g., Attaheri et al., 2015; Meyer and Olson, 2011; Meyer et al., 2014; Milne et al., 2016; Petkov and Wilson, 2012; Wilson et al., 2015, 2017; Uhrig et al., 2014). The current neuroscience evidence supports a combination of modality-specific neural networks (Meyer and Olson, 2011; Meyer et al., 2014; Uhrig et al., 2014) and anterior regions of the brain including frontal cortex (Wilson et al., 2017; Uhrig et al., 2014) that together support sequence processing, prediction, and statistical learning in nonhuman primates (Kikuchi et al., 2018). As reviewed above, a similar interplay between downstream frontal areas and low-level sensory regions appears to mediate statistical learning in humans (Conway and Pisoni, 2008). Frontal brain regions, such as the PFC and perhaps Broca's area especially, may differ across species and may be an important crucial factor in the evolution of complex hierarchical functions such as language (Tecumseh and Martins (2014)).

It is important to examine not just brain regions that are active or elicited during tasks in nonhumans, but also the patterns of interconnectivity among different brain regions and the way that different networks may have evolved to take on specialized functions in different species. For example, humans, relative to other primate species, are known to have a larger PFC, which may be due to an increase in white matter and neural connections to the rest of the brain (Tecumseh and Martins (2014)). Furthermore, there may be differences across species in terms of how different neural pathways (ventral and dorsal) connecting frontal cortex to the rest of the brain are used to learn sequential patterns of varying complexity in language and other domains (e.g., Wilson et al., 2017). It is also important to consider not just learning abilities themselves, but also how learning mechanisms coevolved with

attentional and motivational biases to direct learning and support complex functions (Lotem and Halpern, 2012). Likewise, it is important to consider how statistical learning is used by different species in different ecological niches (Santolin and Saffran, 2018).

In sum, there is much we still do not know about how phylogeny constrains statistical learning. There has been progress in three areas of comparative research (input complexity, modality effects, and neural bases) but there are more questions than answers at this time. We suggest that considering statistical learning as being made up of two primary types of mechanisms (cortical plasticity interacting with higher-level modulatory control processes) might be helpful for constraining the types of research inquiries and hypotheses that are explored. The tentative proposal offered here is that associative-based perceptual learning mediated through general mechanisms of cortical plasticity will be conserved across species (Rey et al., 2018). For instance, despite vast differences in brain size across mammalian species, the temporal dynamics that govern neural communication and information integration is remarkably similar (Buzsáki et al., 2013). Similarly, the use of neurotrophic factors, which impact neuronal survival and differentiation in modulating synaptic plasticity, is relatively conserved across animal species (Casey et al., 2015). Even so, despite the apparent conservation of neural plasticity across species, there are differences, even among mammals: for example, in rodents, neurogenesis is a lifelong process, whereas in humans, adult neurogenesis is much more reduced (La Rosa and Bonfanti, 2018). However, at a behavioral level, it seems relatively clear that basic associative learning and sequence learning abilities are relatively conserved (Wilson et al., 2017).

We suggest that where species differences are observed, they are likely to be due less to variations in cortical plasticity-based mechanisms and more to differences in higher-level cognitive processing, such as top-down cognitive control, attention, and working memory processes, supported by frontal cortex and related networks, and in the different patterns of connectivity among PFC and sensory brain regions. Thus, examining the way that frontal cortex, which mediates top-down control of information-processing, interacts with bottom-up sensory-motor processes, may offer insights into both commonalities and variations across different species (Mishra and Gazzaley, 2016).

#### 4. Ten core principles

Based on the review of findings related to these six areas of research, we outline ten core principles that we believe provide a scaffolding for the construct of statistical learning and lead to testable predictions to help focus future research. Together, the principles argue for the existence of two primary sets of neurocognitive mechanisms or modes of learning that interact to support statistical learning across a variety of contexts. Each principle is described fully below and then presented succinctly in Table 3.

1. Statistical learning is a *multifaceted construct* (e.g., Arciuli, 2017; Daltrozzo and Conway, 2014; Thiessen and Erickson, 2013). We specifically propose two primary cortical mechanisms that underlie statistical learning. The first is based on the general principle of cortical plasticity, which is not localized to any particular area but is prevalent throughout the brain, that mediates basic

associative learning and perceptual learning processes. The second is a top-down modulatory “executive” system (primarily centered in the prefrontal cortex) that directs attention and allows for the learning of more complex patterns. The operation of these two systems or modes of learning likely occurs independently and in parallel (Batterink et al., 2015), though the functioning of the executive system can also emerge as learning occurs through the associative system. These two cortical mechanisms also interact with hippocampal, cerebellar, and basal-ganglia based learning, with each subcortical system contributing to learning depending on the demands of the task and situation (Atallah et al., 2004).

2. The general property of *cortical plasticity*, which mediates bottom-up, associative learning, is instantiated over multiple, hierarchically-embedded networks (Hasson et al., 2015; P.J. Reber, 2013). When we perceive, encode, or act upon a given stimulus, the particular neurocognitive processes that were active tune adaptively with experience, thereby facilitating further processing with the same or similar stimuli. This “plasticity of processing” approach explains the wide and distributed pattern of activity observed in statistical learning neuroimaging studies (e.g., as reviewed in Conway and Pisoni, 2008; Frost et al., 2015; Keele et al., 2003). Furthermore, this type of learning based on cortical plasticity is ever-present and obligatory, always being active.
3. Different areas of neocortex have different *cortical processing capabilities* in terms of perceptual modality and temporal timescales (Fuster and Bressler, 2012), and these processing differences provide constraints on which brain areas will reflect plasticity and learning in any given situation (Frost et al., 2015). For instance, posterior modality-specific perceptual brain regions generally encompass shorter timescales of processing (Hasson et al., 2015), and thus can mediate the learning of local or adjacent dependencies in a modality-specific manner. More anterior downstream regions such as the PFC have longer timescales for integrating information (Hasson et al., 2015), thus allowing for the processing and learning of nonadjacent, long-distance, and global patterns.
4. Existing alongside the general principle of cortical plasticity are endogenous attention and working memory processes that provide *top-down modulation of learning* (Fuster and Bressler, 2012; Hasson et al., 2015). Top-down modulation, a specialized process centered primarily in PFC and related frontoparietal networks, is used to direct attention and processing toward certain aspects of stimuli encountered in the environment, and it may be necessary for learning nonadjacent dependencies in temporal sequences (de Diego-Balaguer et al. (2016d)). Top-down control also allows for inhibition of stimuli or features contained in stimuli, in order for selective attention to be deployed in strategic ways and thus is crucial for learning patterns arrayed across temporal sequences, such as crossmodal sequential dependencies (Walk and Conway, 2016) and global patterns that require integration over longer timescales (Wacongne et al., 2011). Top-down control and endogenous attention may also be needed for the expression of knowledge; it may be particularly necessary when attempting to generalize or apply knowledge to new settings such as when the perceptual

features are changed and a mapping between the old patterns and new ones must be ascertained (Hendricks et al., 2013). On the other hand, at least for some types of learning situations such as perceptual-motor sequence learning embodied by the SRT task, top-down control may actually interfere with or impede implicit learning (Ambrus et al., 2019).

5. Central to statistical learning is *prediction and expectation* (Dale et al., 2012), mediated across both sensory and downstream areas such as the PFC (though it is possible that the PFC generates the predictions and subsequently modulates sensory areas, Bubic, 2010). Prediction is particularly important for serial learning and sequencing because of the prominent role of time and uncertainty in temporal sequences (Bubic, 2010). Predictive processing appears to be inherent to all levels of the hierarchically organized nervous system (Friston, 2005) and thus goes hand-in-hand with a “plasticity of processing” approach. Predictive processing consists both of implicit/automatic as well as explicit, attention-dependent processing (Huettig, 2015). An important part of making predictions of upcoming events is the necessity of inhibiting the internal representations of events or stimuli that are not predicted, which likely involves PFC (Bar, 2009).
6. *Modality and domain effects* arise due to plasticity of processing in perceptual cortical networks, similar to the mechanism of perceptual priming (Conway and Christiansen, 2006; Frost et al., 2015; P.J. Reber, 2013). Cross-modal learning may be possible though only under particular conditions, such as when the cross-modal dependencies are presented simultaneously in time, rather than across a temporal sequence (Walk and Conway, 2016), or when selective attention is deployed. Thus, domain-general cognitive mechanisms such as selective attention are needed to gate or modulate learning (e.g., Turk-Browne et al., 2005). Likewise, certain aspects of statistical learning appear to be domain-general in the sense that some amount of transfer or correspondence across different stimulus sets can occur, such as the recognition of repetition structures and other perceptual primitives (Endress et al., 2009; Gomez et al., 2000). These cross-modal or domain-general functions likely involve PFC and related attention and working memory processes.
7. Learning can occur for a variety of *input structures* that vary in complexity, ranging from simple associations between two stimuli and perceptual “chunks” to more complex and highly variable patterns that span across a temporal sequence and that form recursive or hierarchical structure (Dehaene et al., 2015; Petkov and Wilson, 2012). The limited research suggests that complexity directly influences learning performance, with more complex input patterns leading to lower levels of learning (Schiff and Katan, 2014). It appears that learning different types of structures entails different processing requirements, with some types of perceptual-based and simple patterns being learned relatively automatically and effortlessly (Hendricks et al., 2013), and more complex, nonadjacent patterns requiring the involvement of selective attention or working memory (de Diego-Balaguer et al. (2016d)). Thus, nonadjacent or long-distance dependencies, as well as “global” learning that requires integrating information

across exemplars over time necessarily requires processing by brain networks such as the PFC that can handle information over these larger timescales (Fuster and Bressler, 2012).

8. Statistical learning and attention have a *bidirectional* relationship: attention can modulate or gate learning (Turk-Browne et al., 2005) and learning itself can lead to heightened levels of attention for the structure that has been learned (Zhao et al., 2013). However, learning can also proceed in a relatively implicit or automatic fashion. The involvement of these two primary “modes” or systems, one explicit (attention-dependent) and the other implicit (attention-independent) appears to occur in parallel (Batterink et al., 2015), with the implicit system always “on” but the explicit system optional. As reviewed, the learning of some types of input structures appear to require the explicit system to greater extents, including nonadjacent dependencies, global patterns, abstract rule-based processing, and possibly cross-modal temporal patterns (Bekinschtein et al., 2009; de Diego-Balaguer et al. (2016d); Walk and Conway, 2016). Other types of patterns, such as chunks, can likely be learned through implicit learning via a form of perceptual learning (Chang and Knowlton, 2004; Hendricks et al., 2013) although explicit, attention-based learning can also mediate the formation of chunks in memory (Pacton and Perruchet, 2008). Conscious awareness of what is learned occurs when the strength or quality of the representations reaches a threshold level (Cleeremans, 2011).
9. *Ontogeny* differentially constrains the development of different aspects of statistical learning (cortical plasticity and top-down modulatory control). Because cortical plasticity is a general property of nervous systems, present in varying degrees across all brain networks and in all individuals across the lifespan, from a certain point of view, plasticity-based statistical learning is likely relatively age-invariant. However, this is an oversimplification; plasticity generally is heightened early in development before neural entrenchment results in the ending of sensitive periods (Kuhl et al., 2005; White et al., 2013). Thus, plasticity-mediated associative learning mechanisms dominate learning in infancy and early in development. A second and independent ontogenetic constraint is due to the relatively late maturation of the frontal lobe (Thompson-Schill et al., 2009), which mediates top-down control of learning. Thus, certain aspects of learning that rely on selective attention and cognitive control, such as the learning of nonadjacent and global regularities as well as crossmodal temporal associations, is more effective later in development when the frontal system matures (Daltrozzo and Conway, 2014; de Diego-Balaguer et al. (2016d)). Later in life, in old age, plasticity is still present but likely not at pre-sensitive period peak levels; the frontal system also shows a certain amount of decline in healthy aging (e.g., Van Petten et al., 2004). Thus, a full account of developmental changes across the lifespan must take into account the independent trajectories of these two mechanisms that impact learning.
10. *Phylogeny* also constrains different aspects of statistical learning in different ways. It is likely that neural plasticity is evolutionarily conserved, and thus basic



associative learning principles are likely to be present across the phyla, at least in species that have nervous systems (Rey et al., 2018). On the other hand, it is likely that top-down modulatory control differs across species. This is expected to result in a relatively attenuated ability to learn nonadjacent dependencies, global patterns, and hierarchical structures (Conway and Christiansen, 2001; Rey et al., 2018), which depend crucially on the PFC and is believed to be less developed in most nonhuman species (Tecumseh and Martins, 2014). There may be species differences, too, in terms of neural connectivity and the integrity of specific neural pathways that connect PFC to other brain areas (Tecumseh and Martins, 2014). It is also expected that there may be species-specific differences that vary based on ecological niche and the unique selection pressures faced by the organism (Santolin and Saffran, 2018), making it possible for instance that some species excel at statistical learning but only in certain types of functions or contexts.

## 5. Conclusion

In sum, it is proposed that the construct of statistical learning can be decomposed into multiple components; primary among them are two mostly dissociable, cortically-based, cognitive mechanisms (Fig. 4). The first mechanism is based on the principle of neural plasticity, and therefore encompasses the entire neocortex. Through experience with particular types of patterned input, the brain networks involved in processing that input will show improved processing due to cortical tuning. This system is likely to be largely automatic and attention-independent and is constrained by the processing limitations of the cortical network(s) in question. One such limitation is the timescale of processing, with more posterior networks processing information over shorter timescales, and more anterior networks able to process information over longer timescales. An additional constraint on processing is sensory modality, with modality-specific perceptual regions able to show plasticity only for the types of input available to those networks. It is expected that the general principle of plasticity is present across most individuals and species, though there may be variations within development (increased plasticity prior to the end of the sensitive period) as well as across species (with variations specific to the ecological demands and requirements). In a sense, this first mechanism can be thought of as “obligatory”; it is always online and active in encoding regularities in the environment.

A second mechanism that acts in concert– or sometimes in competition – with the first is one that mediates top-down modulatory control to help filter and selectively attend to particular inputs or features. This “executive” system involves frontoparietal networks (and perhaps specifically the PFC) that mediate endogenous attention and working memory to modulate learning. This system is specifically needed to learn nonadjacent and global regularities as well as crossmodal contingencies across temporal sequences. This mechanism is less likely to be available early in development; it is also likely to be found to varying degrees across species but mainly in cognitively more advanced species such as chimpanzees and humans. At least for the learning of perceptual-motor sequences, the executive system may actually impede learning rather than contribute to it (Ambrus et al., 2019). One could consider this

second system to be “optional” in the sense that it does not seem to be universally active in all situations or contexts, unlike the first mechanism.

Note that this multi-component proposal is similar to and draws upon a number of other theoretical frameworks (e.g., Arciuli, 2017; Batterink et al., 2015; Daltrozzo and Conway, 2014; Frost et al., 2015; Keele et al., 2003; P.J. Reber, 2013; Thiessen and Erickson, 2013). However, the current proposal is the only one that addresses all six factors reviewed above (i.e., how learning proceeds across different *input modalities* and domains and for different types of *input structures*; the role of *attention* in statistical learning; the underlying *neuroanatomy* of statistical learning; and *ontogenetic* and *phylogenetic* constraints on learning). This framework also makes specific and unique predictions that future research can usefully examine. For instance, this framework predicts that the learning of nonadjacent dependencies as well as global patterns that span across time will necessarily require the involvement of PFC and frontoparietal networks and related attentional and working memory processes. Likewise, the learning of cross-modal sequential dependencies will also require these same frontal-based neural processes. It is also predicted that where statistical learning will differ across age and across species will be for the learning of nonadjacent, global, and cross-modal sequential patterns; on the other hand the learning of perceptual chunks or local transitions in a sequence will be relatively conserved across species and across age.

Finally, there are a number of outstanding questions that this review did not address and thus remain as ripe areas for future research. Five specific areas are highlighted here (see additional discussion by Arciuli and Conway, 2018):

1. *What is the relationship between statistical learning and other forms of learning and memory?* Although the stance taken here is that statistical learning, implicit learning, and sequence learning are essentially referring to the same underlying construct, it is also likely that different tasks used to probe learning (e.g., the AGL task versus the SRT task) may reflect partially dissociable aspects of learning. Likewise, further work is needed to specify to what extent statistical learning overlaps with for instance procedural memory (Ullman, 2004), category learning (Smith and Grossman, 2008), or other forms of nondeclarative memory (Squire, 2004). The evidence so far supports the notion that statistical learning relies upon both procedural and declarative forms of memory (Batterink et al. (2019), but more work is needed to work out the exact interactions.
2. *What is the relationship between statistical learning and language processing and development?* Although much work has highlighted the role of statistical learning as a language learning mechanism (e.g., Nemeth et al., 2011; Romberg and Saffran, 2010), there are still unanswered questions about which aspects of statistical learning map onto which aspects of language (e.g., phonology, word learning, syntax, etc.) at different points in development.
3. *What is the relationship between statistical learning and development in non-language domains?* Because statistical learning is a general and pervasive learning mechanism, it should impact a wide range of domains in addition to

language, such as music, perceptual and motor skill development, and educational outcomes, but these connections are still underspecified and thus represent potentially rich areas in need of further investigation.

4. *Which aspects of atypical development are associated with atypical statistical learning abilities?* There is a growing body of evidence suggesting a link between atypical statistical learning and language learning disabilities such as developmental language disorder (Obeid et al., 2016), developmental dyslexia (Gabay et al., 2015), and autism spectrum disorder (Jeste et al., 2015). There is also evidence that variation in statistical learning abilities explains variability in language outcomes in children who are deaf or hard of hearing (Conway et al., 2011; Deocampo et al., 2018; Greppe et al., 2019). More work is needed to understand which aspects of learning causally impact which aspects of atypical development across a variety of clinical populations (c.f., Arciuli and Conway, 2018; Krishnan et al., 2016; Zwart et al., 2019).
5. *Finally, to what extent is statistical learning itself affected by experience?* There are two related questions here. The first is whether the mechanisms underlying statistical learning can be improved to increase the effectiveness of learning. The second concerns the issue of “rewiring”: to what extent can knowledge of statistical regularities be modified or even unlearned in order to assimilate new regularities? In regards to the first question, there is some initial work to suggest that statistical learning may be modifiable to some degree (Onnis et al., 2015; Smith et al., 2015). Likewise, in regards to the second question, it appears possible that knowledge of statistical regularities can be rewired, allowing for the learning of new regularities (Szegedi-Hallgató et al., 2017). If it is possible to improve learning and/or rewire one’s knowledge of what has been learned, then this represents an unprecedented opportunity to use targeted intervention or controlled manipulation of environmental factors to help promote statistical learning across a range of language and learning disorders (Plante and Gómez, 2018). Similarly, it is yet unknown whether methods of improving statistical learning could have an impact across developmental, language, and educational outcomes even in typical developing individuals.

Statistical learning is a robust learning mechanism that provides adaptability, flexibility, and improved behavioral functioning for organisms that can capitalize on the structure inherent in the world. The quest for a unified theory of statistical learning requires continued research to help understand how learning proceeds for different types of input, what cognitive and neural systems undergird learning, and how it emerges across species and within individuals across developmental time. Future research that integrates findings across a number of key areas, as embodied by the ten core principles outlined here, will help us better understand how the brain learns environmental structure.

## Acknowledgments

This work was supported by the National Institute on Deafness and other Communication Disorders (R01DC012037). The sponsor had no role in the writing of this article or in the decision to submit it for publication. Earlier versions of this work were presented by Conway (2016a; 2016b) and Conway, Deocampo, Smith, &

Eghbalzad (2016). We thank Margo Appenzeller, Joanne Deocampo, Samantha Emerson, Karla McGregor, and Michael Ullman for their helpful comments on an earlier draft of this manuscript.

## References

- Abla D, Okanoya K, 2008 Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia* 46 (11), 2787–2795. 10.1016/j.neuropsychologia.2008.05.012. [PubMed: 18579166]
- Adams EJ, Nguyen AT, Cowan N, 2018 Theories of working memory: differences in definition, degree of modularity, role of attention, and purpose. *Lang. Speech Hear. Serv. Sch* 49 (3), 340–349. 10.1044/2018\_LSHSS-17-0114. [PubMed: 29978205]
- Alamia A, Zénon A, 2016 Statistical regularities attract attention when task-relevant. *Front. Hum. Neurosci* 10 10.3389/fnhum.2016.00042.
- Altmann GTM, Dienes Z, 1999 Rule learning by seven-month-old infants and neural networks. *Science* 284 875a.
- Ambrus GG, Vékony T, Janacsek K, Trimborn ABC, Kovács G, Nemeth D, 2019 When less is more: enhanced statistical learning of non-adjacent dependencies after disruption of bilateral DLPFC. *BioRxiv* 10.1101/198515.
- Amso D, Davidow J, 2012 The development of implicit learning from infancy to adulthood: item frequencies, relations, and cognitive flexibility. *Dev. Psychobiol* 54 (6), 664–673. 10.1002/dev.20587. [PubMed: 22714674]
- Arciuli J, 2017 The multi-component nature of statistical learning. *Philos. Trans. Biol. Sci* 372 (1711), 20160058 10.1098/rstb.2016.0058.
- Arciuli J, Conway CM, 2018 The promise—and challenge—of statistical learning for elucidating atypical language development. *Curr. Dir. Psychol. Sci* 9.
- Arciuli J, Simpson IC, 2011 Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Dev. Sci* 14 (3), 464–473. 10.1111/j.1467-7687.2009.00937.x. [PubMed: 21477186]
- Aslin RN, Newport EL, 2012 Statistical learning: from acquiring specific items to forming general rules. *Curr. Dir. Psychol. Sci* 21 (3), 170–176. 10.1177/0963721412436806. [PubMed: 24000273]
- Atallah HE, Frank MJ, O'Reilly RC, 2004 Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems. *Neurobiol. Learn. Mem* 82 (3), 253–267. 10.1016/j.nlm.2004.06.004. [PubMed: 15464408]
- Attaheri A, Kikuchi Y, Milne AE, Wilson B, Alter K, Petkov CI, 2015 EEG potentials associated with artificial grammar learning in the primate brain. *Brain Lang* 148, 74–80. 10.1016/j.bandl.2014.11.006. [PubMed: 25529405]
- Awh E, Vogel EK, Oh S-H, 2006 Interactions between attention and working memory. *Neuroscience* 139 (1), 201–208. 10.1016/j.neuroscience.2005.08.023. [PubMed: 16324792]
- Baars BJ, 1988 *A Cognitive Theory of Consciousness*. Cambridge University Press, New York.
- Baars BJ, 2005 Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog. Brain Res* 150, 45–53. 10.1016/S0079-6123(05)50004-9. [PubMed: 16186014]
- Bahlmann J, Schubotz RI, Friederici AD, 2008 Hierarchical artificial grammar processing engages Broca's area. *NeuroImage* 42 (2), 525–534. 10.1016/j.neuroimage.2008.04.249. [PubMed: 18554927]
- Baker CI, Olson CR, Behrmann M, 2004 Role of attention and perceptual grouping in visual statistical learning. *Psychol. Sci* 15 (7), 460–466. 10.1111/j.0956-7976.2004.00702.x. [PubMed: 15200630]
- Bar M, 2009 The proactive brain: memory for predictions. *Philos. Trans. Biol. Sci* 364 (1521), 1235–1243. 10.1098/rstb.2008.0310.
- Bargh JA, 1994 The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition In: Wyer RS Jr., Srull TK (Eds.), *Handbook of Social Cognition, Vol.1: Basic Processes*, 2nd ed. Psychology Press, New York, NY, pp. 1–40.
- Batterink LJ, Reber PJ, Neville HJ, Paller KA, 2015 Implicit and explicit contributions to statistical learning. *J. Mem. Lang* 83, 62–78. 10.1016/j.jml.2015.04.004. [PubMed: 26034344]

- Batterink LJ, Paller KA, Reber PJ, 2019 Understanding the neural bases of implicit and statistical learning. *Top. Cogn. Sci* 1–22. 10.1111/tops.12420.
- Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L, 2009 Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci* 106 (5), 1672–1677. 10.1073/pnas.0809667106. [PubMed: 19164526]
- Bertels J, Destrebecqz A, Franco A, 2015 Interacting effects of instructions and presentation rate on visual statistical learning. *Front. Psychol* 6 10.3389/fpsyg.2015.01806.
- Blakemore C, Cooper GF, 1970 Development of the brain depends on the visual environment. *Nature* 228, 477–478. [PubMed: 5482506]
- Bubic, 2010 Prediction, cognition and the brain. *Frontiers in Human Neuroscience* 10.3389/fnhum.2010.00025.
- Bulf H, Johnson SP, Valenza E, 2011 Visual statistical learning in the newborn infant. *Cognition* 121 (1), 127–132. 10.1016/j.cognition.2011.06.010. [PubMed: 21745660]
- Buzsáki G, Logothetis N, Singer W, 2013 Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80 (3), 751–764. 10.1016/j.neuron.2013.10.002. [PubMed: 24183025]
- Casey BJ, Glatt CE, Lee FS, 2015 Treating the developing versus developed brain: translating preclinical mouse and human studies. *Neuron* 86 (6), 1358–1368. 10.1016/j.neuron.2015.05.020. [PubMed: 26087163]
- Chang GY, Knowlton BJ, 2004 Visual feature learning in artificial grammar classification. *J. Exp. Psychol. Learn. Mem. Cogn* 30 (3), 714–722. 10.1037/0278-7393.30.3.714. [PubMed: 15099138]
- Chica AB, Bartolomeo P, Lupiáñez J, 2013 Two cognitive and neural systems for endogenous and exogenous spatial attention. *Behavioral Brain Research* 237, 107–123. 10.1016/j.bbr.2012.09.027.
- Christiansen MH, 2018 Implicit statistical learning: a tale of two literatures. *Top. Cogn. Sci* 1–14. 10.1111/tops.12332.
- Christiansen MH, Chater N, 2015 The language faculty that wasn't: a usage-based account of natural language recursion. *Front. Psychol* 6 10.3389/fpsyg.2015.01182.
- Cleeremans A, 2011 The radical plasticity thesis: how the brain learns to be conscious. *Front. Psychol* 2 10.3389/fpsyg.2011.00086.
- Cleeremans A, McClelland JL, 1991 Learning the structure of event sequences. *J. Exp. Psychol. Gen* 120 (3), 235–253. [PubMed: 1836490]
- Cleeremans A, Destrebecqz A, Boyer M, 1998 Implicit learning: news from the front. *Trends Cogn. Sci. (Regul. Ed.)* 2 (10), 406–416. 10.1016/S1364-6613(98)01232-7. [PubMed: 21227256]
- Conway CM, 2005 An Odyssey Through Sight, Sound, and Touch: Toward a Perceptual Theory of Implicit Statistical Learning Unpublished doctoral dissertation. Cornell University, Ithaca, NY.
- Conway CM, Christiansen MH, 2001 Sequential learning in non-human primates. *Trends Cogn. Sci. (Regul. Ed.)* 5 (12), 539–546. 10.1016/S1364-6613(00)01800-3. [PubMed: 11728912]
- Conway CM, Christiansen MH, 2005 Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn* 31 (1), 24–39. 10.1037/0278-7393.31.1.24. [PubMed: 15641902]
- Conway CM, Christiansen MH, 2006 Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychol. Sci* 17 (10), 905–912. 10.1111/j.1467-9280.2006.01801.x. [PubMed: 17100792]
- Conway CM, Christiansen MH, 2009 Seeing and hearing in space and time: effects of modality and presentation rate on implicit statistical learning. *Eur. J. Cogn. Psychol* 21 (4), 561–580. 10.1080/09541440802097951.
- Conway CM, Pisoni DB, 2008 Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Ann. N. Y. Acad. Sci* 1145 (1), 113–131. 10.1196/annals.1416.009. [PubMed: 19076393]
- Conway CM, Goldstone RL, Christiansen MH, 2007 Spatial constraints on visual statistical learning of multi-element scenes. In: *In Proceedings of the 29th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society pp. 185–190.

- Conway CM, Pisoni DB, Anaya EM, Karpicke J, Henning SC, 2011 Implicit sequence learning in deaf children with cochlear implants. *Dev. Sci* 14 (1), 69–82. 10.1111/j.1467-7687.2010.00960.x. [PubMed: 21159089]
- Cooper SJ, 2005 Donald O. Hebb's synapse and learning rule: a history and commentary. *Neurosci. Biobehav. Rev* 28 (8), 851–874. [PubMed: 15642626]
- Cowan N, 1988 Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull* 104 (2), 163–191. [PubMed: 3054993]
- Cowan N, 2017 The many faces of working memory and short-term storage. *Psychon. Bull. Rev* 24 (4), 1158–1170. 10.3758/s13423-016-1191-6. [PubMed: 27896630]
- Creel SC, Newport EL, Aslin RN, 2004 Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol. Learn. Mem. Cogn* 30 (5), 1119–1130. 10.1037/0278-7393.30.5.1119. [PubMed: 15355140]
- Cunillera T, Càmarà E, Laine M, Rodríguez-Fornells A, 2010 Speech segmentation is facilitated by visual cues. *Q. J. Exp. Psychol* 63 (2), 260–274. <https://doi.org/10.1080/17470210902888809>.
- Dale R, Duran N, Morehead R, 2012 Prediction during statistical learning, and implications for the implicit/explicit divide. *Adv. Cogn. Psychol* 8 (2), 196–209. 10.5709/acp-0115-z. [PubMed: 22723817]
- Daltrozzo J, Conway CM, 2014 Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Front. Hum. Neurosci* 8 10.3389/fnhum.2014.00437.
- de Diego-Balaguer R, Martínez-Alvarez A, Pons F, 2016d Temporal attention as a scaffold for language development. *Front. Psychol* 7 10.3389/fpsyg.2016.00044.
- Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C, 2015 The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88 (1), 2–19. 10.1016/j.neuron.2015.09.019. [PubMed: 26447569]
- Deocampo JA, Conway CM, 2016 A developmental shift in the relationship between sequential learning, executive function, and language ability as revealed by event-related potentials. In: *In Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Philadelphia, PA: Cognitive Science Society pp. 1074–1079.
- Deocampo JA, Smith GNL, Kronenberger WG, Pisoni DB, Conway CM, 2018 The role of statistical learning in understanding and treating spoken language outcomes in deaf children with cochlear implants. *Lang. Speech Hear. Serv. Sch* 49 (3S), 723 10.1044/2018\_LSHSS-STLT1-17-0138. [PubMed: 30120449]
- Deocampo JA, King TZ, Conway CM, 2019 Concurrent learning of adjacent and non-adjacent dependencies in visuo-spatial and visuo-verbal sequences. *Front. Psychol* 10, 1107 10.3389/fpsyg.2019.01107. [PubMed: 31164850]
- Desmond JE, Fiez JA, 1998 Neuroimaging studies of the cerebellum: language, learning, and memory. *Trends Cogn. Sci. (Regul. Ed.)* 2 (9), 355–362. [PubMed: 21227232]
- Dulany DE, Carlson RA, Dewey GI, 1984 A case of syntactical learning and judgment: How conscious and how abstract? *J. Exp. Psychol. Gen* 113 (4), 541–555.
- Elman JL, 1993 Learning and development in neural networks: the importance of starting small. *Cognition* 48 (1), 71–99. 10.1016/0010-0277(93)90058-4. [PubMed: 8403835]
- Emberson LL, Conway CM, Christiansen MH, 2011 Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Q. J. Exp. Psychol* 64 (5), 1021–1040. 10.1080/17470218.2010.538972.
- Endress AD, Mehler J, 2009 The surprising power of statistical learning: when fragment knowledge leads to false memories of unheard words. *J. Mem. Lang* 60 (3), 351–367. 10.1016/j.jml.2008.10.003.
- Endress AD, Nespor M, Mehler J, 2009 Perceptual and memory constraints on language acquisition. *Trends Cogn. Sci. (Regul. Ed.)* 13 (8), 348–353. 10.1016/j.tics.2009.05.005. [PubMed: 19647474]
- Farbood MM, Heeger DJ, Marcus G, Hasson U, Lerner Y, 2015 The neural processing of hierarchical structure in music and speech at different timescales. *Front. Neurosci* 9 10.3389/fnins.2015.00157.

- Filoteo JV, Lauritzen S, Maddox WT, 2010 Removing the frontal Lobes: the effects of engaging executive functions on perceptual category learning. *Psychol. Sci* 21 (3), 415–423. 10.1177/0956797610362646. [PubMed: 20424079]
- Fiser J, Aslin RN, 2001 Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci* 12 (6), 499–504. 10.1111/1467-9280.00392. [PubMed: 11760138]
- Fiser J, Aslin RN, 2005 Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen* 134 (4), 521–537. 10.1037/0096-3445.134.4.521. [PubMed: 16316289]
- Fitch WT, Friederici AD, 2012 Artificial grammar learning meets formal language theory: an overview. *Philos. Trans. Biol. Sci* 367 (1598), 1933–1955. 10.1098/rstb.2012.0103.
- Forkstam C, Petersson KM, 2005 Towards an explicit account of implicit learning. *Curr. Opin. Neurol* 18 (4), 435–441. 10.1097/01.wco.0000171951.82995.c4. [PubMed: 16003121]
- Forkstam C, Hagoort P, Fernandez G, Ingvar M, Petersson KM, 2006 Neural correlates of artificial syntactic structure classification. *NeuroImage* 32 (2), 956–967. 10.1016/j.neuroimage.2006.03.057. [PubMed: 16757182]
- Franco A, Destrebecqz A, 2012 Chunking or not chunking? How do we find words in artificial language learning? *Adv. Cogn. Psychol* 8 (2), 144–154. 10.5709/acp-0111-3. [PubMed: 22723813]
- Friederici AD, Bahlmann J, Heim S, Schubotz RI, Anwander A, 2006 The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci* 103 (7), 2458–2463. 10.1073/pnas.0509389103. [PubMed: 16461904]
- Friston K, 2005 A theory of cortical responses. *Philos. Trans. Biol. Sci* 360 (1456), 815–836. 10.1098/rstb.2005.1622.
- Frost RLA, Monaghan P, 2016 Simultaneous segmentation and generalisation of nonadjacent dependencies from continuous speech. *Cognition* 147, 70–74. 10.1016/j.cognition.2015.11.010. [PubMed: 26638049]
- Frost R, Armstrong BC, Siegelman N, Christiansen MH, 2015 Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci. (Regul. Ed.)* 19 (3), 117–125. 10.1016/j.tics.2014.12.010. [PubMed: 25631249]
- Fuster JM, 2001 The prefrontal cortex - an update: time is of the essence. *Neuron* 30, 319–333. [PubMed: 11394996]
- Fuster JM, Bressler SL, 2012 Cognit activation: a mechanism enabling temporal integration in working memory. *Trends Cogn. Sci. (Regul. Ed.)* 16 (4), 207–218. 10.1016/j.tics.2012.03.005. [PubMed: 22440831]
- Gabay Y, Thiessen ED, Holt LL, 2015 Impaired statistical learning in developmental dyslexia. *J. Speech Lang. Hear. Res* 58 (3), 934–945. 10.1044/2015\_JSLHR-L-14-0324. [PubMed: 25860795]
- Gervain J, Macagno F, Cogoi S, Pena M, Mehler J, 2008 The neonate brain detects speech structure. *Proc. Natl. Acad. Sci* 105 (37), 14222–14227. 10.1073/pnas.0806530105. [PubMed: 18768785]
- Giroux I, Rey A, 2009 Lexical and sublexical units in speech perception. *Cogn. Sci* 33 (2), 260–272. 10.1111/j.1551-6709.2009.01012.x. [PubMed: 21585470]
- Goddard MJ, 2018 Extending B. F. Skinner's selection by consequences to personality change, implicit theories of intelligence, skill learning, and language. *Rev. Gen. Psychol* 22 (4), 421–426. 10.1037/gpr0000168.
- Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, Thompson PM, 2004 Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci* 101 (21), 8174–8179. 10.1073/pnas.0402680101. [PubMed: 15148381]
- Gomez RL, 1997 Transfer and complexity in artificial grammar learning. *Cogn. Psychol* 33 (2), 154–207. 10.1006/cogp.1997.0654. [PubMed: 9245469]
- Gómez RL, 2002 Variability and detection of invariant structure. *Psychol. Sci* 13 (5), 6.
- Gomez RL, Gerken L, Schvaneveldt RW, 2000 The basis of transfer in artificial grammar learning. *Mem. Cognit* 28 (2), 253–263.
- Goschke T, 1998 Implicit learning of perceptual and motor sequences: evidence for independent learning systems In: Stadler MA, Frensch PA (Eds.), *Handbook of Implicit Learning* SAGE Publications, London, pp. 401–444.

- Grep MA, Deocampo JA, Walk AM, Conway CM, 2019 Visual sequential processing and language ability in children who are deaf or hard of hearing. *J. Child Lang* 1–15. 10.1017/S0305000918000569.
- Gureckis TM, Love BC, 2007 Behaviorism reborn? Statistical learning as simple conditioning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* 29 (29), 335–340.
- Hallgató E, Gyri-Dani D, Pékár J, Janacsek K, Nemeth D, 2013 The differential consolidation of perceptual and motor learning in skill acquisition. *Cortex* 49 (4), 1073–1081. 10.1016/j.cortex.2012.01.002. [PubMed: 22325422]
- Hard BM, Meyer M, Baldwin D, 2018 Attention reorganizes as structure is detected in dynamic action. *Mem. Cognit* 10.3758/s13421-018-0847-z.
- Hasher L, Zacks RT, 1979 Automatic and effortful processes in memory. *J. Exp. Psychol. Gen* 108, 356–388. 10.1037/0096-3445.108.3.356.
- Hasson U, Chen J, Honey CJ, 2015 Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn. Sci. (Regul. Ed.)* 19 (6), 304–313. 10.1016/j.tics.2015.04.006. [PubMed: 25980649]
- Hauser MD, Chomsky N, Fitch WT, 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 11.
- Hebb DO, 1949 *The Organization of Behavior: a Neuropsychological Theory*. John Wiley & Sons, Inc, New York, NY.
- Heimbauer LA, Conway CM, Christiansen MH, Beran MJ, Owren MJ, 2018 Visual artificial grammar learning by rhesus macaques (*Macaca mulatta*): exploring the role of grammar complexity and sequence length. *Anim. Cogn* 21 (2), 267–284. 10.1007/s10071-018-1164-4. [PubMed: 29435770]
- Hendricks MA, Conway CM, Kellogg RT, 2013 Using dual-task methodology to dissociate automatic from nonautomatic processes involved in artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn* 39 (5), 1491–1500. 10.1037/a0032974. [PubMed: 23627281]
- Huetig F, 2015 Four central questions about prediction in language processing. *Brain Res* 1626, 118–135. 10.1016/j.brainres.2015.02.014. [PubMed: 25708148]
- Ivry RB, Baldo JV, 1992 Is the cerebellum involved in learning and cognition? *Curr. Opin. Neurobiol* 2 (2), 212–216. [PubMed: 1638157]
- Jager G, Rogers J, 2012 Formal language theory: refining the Chomsky hierarchy. *Philos. Trans. Biol. Sci* 367 (1598), 1956–1970. 10.1098/rstb.2012.0077.
- Jamieson RK, Mewhort DJK, 2005 The influence of grammatical, local, and organizational redundancy on implicit learning: an analysis using information theory. *J. Exp. Psychol. Learn. Mem. Cogn* 31 (1), 9–23. 10.1037/0278-7393.31.1.9. [PubMed: 15641901]
- Janacsek K, Nemeth D, 2012 Predicting the future: from implicit learning to consolidation. *Int. J. Psychophysiol* 83 (2), 213–221. 10.1016/j.ijpsycho.2011.11.012. [PubMed: 22154521]
- Janacsek K, Nemeth D, 2013 Implicit sequence learning and working memory: Correlated or complicated? *Cortex* 49 (8), 2001–2006. 10.1016/j.cortex.2013.02.012. [PubMed: 23541152]
- Janacsek K, Nemeth D, 2015 The puzzle is complicated: when should working memory be related to implicit sequence learning, and when should it not? Response to Martini et al.. *Cortex* 64, 411–412. 10.1016/j.cortex.2014.07.020. [PubMed: 25239854]
- Janacsek K, Fiser J, Nemeth D, 2012 The best time to acquire new skills: age-related differences in implicit sequence learning across the human lifespan. *Dev. Sci* 15 (4), 496–505. 10.1111/j.1467-7687.2012.01150.x. [PubMed: 22709399]
- Jeste SS, Kirkham N, Senturk D, Hasenstab K, Sugar C, Kupelian C, Johnson SP, 2015 Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Dev. Sci* 18 (1), 90–105. 10.1111/desc.12188. [PubMed: 24824992]
- Johnson C, Wilbrecht L, 2011 Juvenile mice show greater flexibility in multiple choice reversal learning than adults. *Dev. Cogn. Neurosci* 1 (4), 540–551. 10.1016/j.dcn.2011.05.008. [PubMed: 21949556]
- Johnson-Pynn J, Frigaszy DM, Brakke KE, Hirsh EM, Greenfield PM, 1999 Strategies used to combine seriated cups by chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), and capuchins (*Cebus apella*). *J. Comp. Psychol* 113 (2), 137–148. [PubMed: 10384722]



- Jost E, Conway CM, Purdy JD, Walk AM, Hendricks MA, 2015 Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Res* 1597, 95–107. 10.1016/j.brainres.2014.10.017. [PubMed: 25475992]
- Juhász D, Nemeth D, Janáček K, 2019 Is there more room to improve? The lifespan trajectory of procedural learning and its relationship to the between- and within-group differences in average response times. *PLoS One* 14 (7), e0215116. [PubMed: 31314804]
- Karuzá EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D, 2013 The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang* 127 (1), 46–54. 10.1016/j.bandl.2012.11.007. [PubMed: 23312790]
- Keele SW, Ivry R, Mayr U, Hazeltine E, Heuer H, 2003 The cognitive and neural architecture of sequence representation. *Psychol. Rev* 110 (2), 316–339. 10.1037/0033-295X.110.2.316. [PubMed: 12747526]
- Kiebel SJ, Daunizeau J, Friston KJ, 2008 A hierarchy of time-scales and the brain. *PLoS Comput. Biol* 4 (11), e1000209 10.1371/journal.pcbi.1000209. [PubMed: 19008936]
- Kikuchi Y, Sedley W, Griffiths TD, Petkov CI, 2018 Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. *Curr. Opin. Behav. Sci* 21, 145–153. 10.1016/j.cobeha.2018.05.002. [PubMed: 30057937]
- Kirkham NZ, Slemmer JA, Johnson SP, 2002 Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83 (2), B35–B42. 10.1016/S0010-0277(02)00004-5. [PubMed: 11869728]
- Kleim JA, Jones TA, 2008 Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. *J. Speech Lang. Hear. Res* 51 (1), S225 10.1044/1092-4388(2008/018). [PubMed: 18230848]
- Knowlton BJ, Squire LR, 1996 Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J. Exp. Psychol. Learn. Mem. Cogn* 22 (1), 169–181. [PubMed: 8648284]
- Kóbor A, Takács Á, Kardos Z, Janáček K, Horváth K, Csépe V, Nemeth D, 2018 ERPs differentiate the sensitivity to statistical probabilities and the learning of sequential structures during procedural learning. *Biol. Psychol* 135, 180–193. 10.1016/j.biopsycho.2018.04.001. [PubMed: 29634990]
- Kral A, Eggermont JJ, 2007 What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Res. Rev* 56 (1), 259–269. 10.1016/j.brainresrev.2007.07.021. [PubMed: 17950463]
- Krishnan S, Watkins KE, Bishop DVM, 2016 Neurobiological basis of language learning difficulties. *Trends Cogn. Sci. (Regul. Ed.)* 20 (9), 701–714. 10.1016/j.tics.2016.06.012. [PubMed: 27422443]
- Kristjansson A, Vuilleumier P, Schwartz S, Macaluso E, Driver J, 2007 Neural basis for priming of pop-out during visual search revealed with fMRI. *Cereb. Cortex* 17 (7), 1612–1624. 10.1093/cercor/bhl072. [PubMed: 16959868]
- Kuhl PK, 2004 Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci* 5 (11), 831–843. 10.1038/nrn1533. [PubMed: 15496861]
- Kuhl PK, Conboy BT, Padden D, Nelson T, Pruitt J, 2005 Early speech perception and later language development: Implications for the “critical period”. *Lang. Learn. Dev* 1 (3&4), 237–264.
- Kurdi B, Gershman SJ, Banaji MR, 2019 Model-free and model-based learning processes in the updating of explicit and implicit evaluations. *Proc. Natl. Acad. Sci* 116 (13), 6035–6044. 10.1073/pnas.1820238116. [PubMed: 30862738]
- La Rosa C, Bonfanti L, 2018 Brain plasticity in mammals: an example for the role of comparative medicine in the neurosciences. *Front. Vet. Sci* 5 10.3389/fvets.2018.00274.
- Lamme VAF, 2003 Why visual attention and awareness are different. *Trends Cogn. Sci. (Regul. Ed.)* 7 (1), 12–18. 10.1016/S1364-6613(02)00013-X. [PubMed: 12517353]
- Lany J, Gómez RL, 2008 Twelve-month-old infants benefit from prior experience in statistical learning. *Psychol. Sci* 19 (12), 1247–1252. 10.1111/j.1467-9280.2008.02233.x. [PubMed: 19121132]

- Lieberman MD, Chang GY, Chiao J, Bookheimer SY, Knowlton BJ, 2004 An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *J. Cogn. Neurosci* 16 (3), 427–438. 10.1162/089892904322926764. [PubMed: 15072678]
- Lotem A, Halpern JY, 2012 Coevolution of learning and data-acquisition mechanisms: a model for cognitive evolution. *Philos. Trans. Biol. Sci* 367 (1603), 2686–2694. 10.1098/rstb.2012.0213.
- Marcovitch S, Lewkowicz DJ, 2009 Sequence learning in infancy: the independent contributions of conditional probability and pair frequency information. *Dev. Sci* 12 (6), 1020–1025. 10.1111/j.1467-7687.2009.00838.x. [PubMed: 19840056]
- Marcus GF, Vijayan S, Bandi Rao S, Vishton PM, 1999 Rule learning by seven-month-old infants. *Science* 283 (5398), 77–80. 10.1126/science.283.5398.77. [PubMed: 9872745]
- Martini M, Sachse P, Furtner MR, Gaschler R, 2015 Why should working memory be related to incidentally learned sequence structures? *Cortex* 64, 407–410. 10.1016/j.cortex.2014.05.016. [PubMed: 25113155]
- Meltzoff AN, Kuhl PK, Movellan J, Sejnowski TJ, 2009 Foundations for a new science of learning. *Science* 325 (5938), 284–288. 10.1126/science.1175626. [PubMed: 19608908]
- Merchant H, Honing H, 2014 Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience* 7 10.3389/fnins.2013.00274.
- Meyer T, Olson CR, 2011 Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci* 108 (48), 19401–19406. 10.1073/pnas.1112895108. [PubMed: 22084090]
- Meyer T, Ramachandran S, Olson CR, 2014 Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *J. Neurosci* 34 (28), 9332–9337. 10.1523/JNEUROSCI.1215-14.2014. [PubMed: 25009266]
- Milne AE, Mueller JL, Männel C, Attaheri A, Friederici AD, Petkov CI, 2016 Evolutionary origins of non-adjacent sequence processing in primate brain potentials. *Sci. Rep* 6 (1). 10.1038/srep36259.
- Milne A, Wilson B, Christiansen M, 2018a Structured sequence learning across sensory modalities in humans and nonhuman primates. *Curr. Opin. Behav. Sci* 21, 39–48. 10.1016/j.cobeha.2017.11.016.
- Milne AE, Petkov CI, Wilson B, 2018b Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience* 389, 104–117. 10.1016/j.neuroscience.2017.06.059. [PubMed: 28687306]
- Mishra J, Gazzaley A, 2016 Cross-species approaches to cognitive neuroplasticity research. *NeuroImage* 131, 4–12. 10.1016/j.neuroimage.2015.09.002. [PubMed: 26348561]
- Mitchel AD, Weiss DJ, 2011 Learning across senses: cross-modal effects in multisensory statistical learning. *J. Exp. Psychol. Learn. Mem. Cogn* 37 (5), 1081–1091. 10.1037/a0023700. [PubMed: 21574745]
- Mitchel AD, Christiansen MH, Weiss DJ, 2014 Multimodal integration in statistical learning: evidence from the McGurk illusion. *Front. Psychol* 5 10.3389/fpsyg.2014.00407.
- Mueller JL, Milne A, Männel C, 2018 Non-adjacent auditory sequence learning across development and primate species. *Curr. Opin. Behav. Sci* 21, 112–119. 10.1016/j.cobeha.2018.04.002.
- Nemeth D, Hallgató E, Janacsek K, Sándor T, Londe Z, 2009 Perceptual and motor factors of implicit skill learning. *NeuroReport* 20 (18), 1654–1658. 10.1097/WNR.0b013e328333ba08. [PubMed: 19901856]
- Nemeth D, Janacsek K, Csifcsak G, Szvoboda G, Howard JH, Howard DV, 2011 Interference between sentence processing and probabilistic implicit sequence learning. *PLoS One* 6 (3), e17577 10.1371/journal.pone.0017577. [PubMed: 21408117]
- Nemeth D, Janacsek K, Polner B, Kovacs ZA, 2013 Boosting human learning by hypnosis. *Cereb. Cortex* 23 (4), 801–805. 10.1093/cercor/bhs068. [PubMed: 22459017]
- Newport EL, 1990 Maturation constraints on language learning. *Cogn. Sci* 14 (1), 11–28. 10.1207/s15516709cog1401\_2.
- Newport EL, Aslin RN, 2004 Learning at a distance I. Statistical learning of nonadjacent dependencies. *Cogn. Psychol* 48 (2), 127–162. 10.1016/S0010-0285(03)00128-2. [PubMed: 14732409]

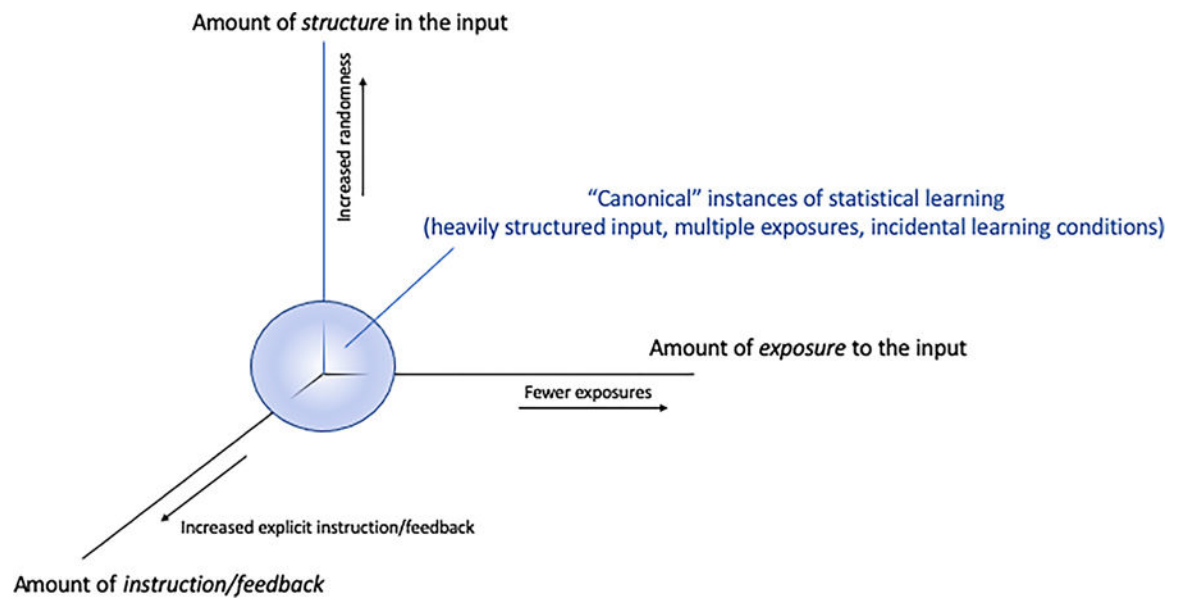
- Nissen MJ, Bullemer P, 1987 Attentional requirements of learning: evidence from performance measures. *Cogn. Psychol* 19 (1), 1–32.
- Norman LJ, Heywood CA, Kentridge RW, 2013 Object-based attention without awareness. *Psychol. Sci* 24 (6), 836–843. 10.1177/0956797612461449. [PubMed: 23572282]
- O'Reilly RC, Norman KA, 2002 Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends Cogn. Sci. (Regul. Ed.)* 6 (12), 505–510. 10.1016/S1364-6613(02)02005-3. [PubMed: 12475710]
- Obeid R, Brooks PJ, Powers KL, Gillespie-Lynch K, Lum JAG, 2016 Statistical learning in specific language impairment and autism spectrum disorder: a meta-analysis. *Front. Psychol* 7 10.3389/fpsyg.2016.01245.
- Onnis L, Lou-Magnuson M, Yun H, Thiessen ED, 2015 Is statistical learning trainable? In: Noelle DC, Dale R, Warlaumont AS, Yoshimi J, Matlock T, Jennings CD, Maglio PP (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* Cognitive Science Society, Austin, TX, pp. 1781–1786.
- Orban G, Fiser J, Aslin RN, Lengyel M, 2008 Bayesian learning of visual chunks by human observers. *Proc. Natl. Acad. Sci* 105 (7), 2745–2750. 10.1073/pnas.0708424105. [PubMed: 18268353]
- Pacton S, Perruchet P, 2008 An attention-based associative account of adjacent and nonadjacent dependency learning. *J. Exp. Psychol. Learn. Mem. Cogn* 34 (1), 80–96. 10.1037/0278-7393.34.1.80. [PubMed: 18194056]
- Pascual-Leone A, Freitas C, Oberman L, Horvath JC, Halko M, Eldaief M, Rotenberg A, 2011 Characterizing brain cortical plasticity and network dynamics across the age-span in health and disease with TMS-EEG and TMS-fMRI. *Brain Topogr* 24 (3–4), 302–315. 10.1007/s10548-011-0196-8. [PubMed: 21842407]
- Pearce JM, Bouton ME, 2001 Theories of associative learning in animals. *Annu. Rev. Psychol* 52 (1), 111–139. 10.1146/annurev.psych.52.1.111. [PubMed: 11148301]
- Perruchet P, Pacteau C, 1990 Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *J. Exp. Psychol. Gen* 119 (3), 264–275.
- Perruchet P, Pacton S, 2006 Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci. (Regul. Ed.)* 10 (5), 233–238. 10.1016/j.tics.2006.03.006. [PubMed: 16616590]
- Perruchet P, Poulin-Charronnat B, 2012 Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *J. Mem. Lang* 66 (4), 807–818. 10.1016/j.jml.2012.02.010.
- Perruchet P, Vinter A, 1998 PARSER: a model for word segmentation. *J. Mem. Lang* 39 (2), 246–263. 10.1006/jmla.1998.2576.
- Petkov CI, Wilson B, 2012 On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philos. Trans. Biol. Sci* 367 (1598), 2077–2088. 10.1098/rstb.2012.0073.
- Plante E, Gómez RL, 2018 Learning without trying: the clinical relevance of statistical learning. *Lang. Speech Hear. Serv. Sch* 49 (3S), 710 10.1044/2018\_LSHSS-STLT1-17-0131. [PubMed: 30120448]
- Plebanek DJ, Sloutsky VM, 2017 Costs of selective attention: when children notice what adults miss. *Psychol. Sci* 28 (6), 723–732. 10.1177/0956797617693005. [PubMed: 28388275]
- Poldrack RA, Rodriguez P, 2004 How do memory systems interact? Evidence from human classification learning. *Neurobiol. Learn. Mem* 82 (3), 324–332. 10.1016/j.nlm.2004.05.003. [PubMed: 15464413]
- Pothos E, 2007 Theories of artificial grammar learning. *Psychol. Bull* 133 (2), 227–244. [PubMed: 17338598]
- Pothos E, 2010 An entropy model for artificial grammar learning. *Front. Psychol* 10.3389/fpsyg.2010.00016.
- Raviv L, Arnon I, 2017 The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Dev. Sci* 21 (4), e12593 10.1111/desc.12593. [PubMed: 28901038]
- Reber AS, 1967 Implicit learning of artificial grammars. *J. Verbal Learning Verbal Behav* 6, 855–863.

- Reber AS, 1989 Implicit learning and tacit knowledge. *J. Exp. Psychol. Gen* 118 (3), 219–235.
- Reber AS, 2003 *Implicit Learning and Tacit Knowledge: an Essay on the Cognitive Unconscious*. Oxford University Press, Oxford.
- Reber PJ, 2013 The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia* 51 (10), 2026–2042. 10.1016/j.neuropsychologia.2013.06.019. [PubMed: 23806840]
- Reber PJ, Stark CEL, Squire LR, 1998 Cortical areas supporting category learning identified using functional MRI. *Proc. Natl. Acad. Sci* 95 (2), 747–750. 10.1073/pnas.95.2.747. [PubMed: 9435264]
- Remillard G, 2008 Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *Q. J. Exp. Psychol* 61 (3), 400–424. 10.1080/17470210701210999.
- Rescorla R, Wagner A, 1972 A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement In: Black A, Prokasy W (Eds.), *Classical Conditioning ii: Current Research and Theory*. Appleton-Century-Crofts, New York, pp. 64–99.
- Rey A, Perruchet P, Fagot J, 2012 Centre-embedded structures are a by-product of associative learning and working memory constraints: evidence from baboons (*Papio Papio*). *Cognition* 123 (1), 180–184. 10.1016/j.cognition.2011.12.005. [PubMed: 22225966]
- Rey A, Minier L, Malassis R, Bogaerts L, Fagot J, 2018 Regularity extraction across species: Associative learning mechanisms shared by human and non-human primates. *Top. Cogn. Sci* 10.1111/tops.12343.
- Romberg AR, Saffran JR, 2010 Statistical learning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci* 1 (6), 906–914. 10.1002/wcs.78. [PubMed: 21666883]
- Romberg AR, Saffran JR, 2013 All together now: concurrent learning of multiple structures in an artificial language. *Cogn. Sci* 37 (7), 1290–1320. 10.1111/cogs.12050. [PubMed: 23772795]
- Saffran JR, Aslin RN, Newport EL, 1996 Statistical learning by 8-month-old infants. *Science* 274 (5294), 1926–1928. [PubMed: 8943209]
- Saffran JR, Johnson EK, Aslin RN, Newport EL, 1999 Statistical learning of tone sequences by human infants and adults. *Cognition* 70 (1), 27–52. 10.1016/S0010-0277(98)00075-4. [PubMed: 10193055]
- Santolin C, Saffran JR, 2018 Constraints on statistical learning across species. *Trends Cogn. Sci. (Regul. Ed.)* 22 (1), 52–63. 10.1016/j.tics.2017.10.003. [PubMed: 29150414]
- Savalia T, Shukla A, Bapi RS, 2016 A unified theoretical framework for cognitive sequencing. *Front. Psychol* 7 10.3389/fpsyg.2016.01821.
- Sawi OM, Rueckl J, 2019 Reading and the neurocognitive bases of statistical learning. *Sci. Stud. Read* 23 (1), 8–23. 10.1080/10888438.2018.1457681. [PubMed: 31105421]
- Schacter DL, Badgaiyan RD, 2001 Neuroimaging of priming: new perspectives on implicit and explicit memory. *Curr. Dir. Psychol. Sci* 10 (1), 1–4. 10.1111/1467-8721.00101.
- Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB, 2014 The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci* 26 (8), 1736–1747. 10.1162/jocn\_a\_00578. [PubMed: 24456393]
- Schiff R, Katan P, 2014 Does complexity matter? Meta-analysis of learner performance in artificial grammar tasks. *Front. Psychol* 5 10.3389/fpsyg.2014.01084.
- Seger CA, 1994 Implicit learning. *Psychol. Bull* 115 (2), 163–196. [PubMed: 8165269]
- Seitz AR, Kim R, van Wassenhove V, Shams L, 2007 Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception* 36 (10), 1445–1453. 10.1068/p5843. [PubMed: 18265827]
- Sengupta P, Burgaleta M, Zamora-López G, Basora A, Sanjuán A, Deco G, Sebastian-Galles N, 2018 Traces of statistical learning in the brain's functional connectivity after artificial language exposure. *Neuropsychologia* 10.1016/j.neuropsychologia.2018.12.001.
- Shufaniya A, Arnon I, 2018 Statistical learning is not age-invariant during childhood: performance improves with age across modality. *Cogn. Sci* 42 (8), 3100–3115. 10.1111/cogs.12692. [PubMed: 30276848]

- Simor P, Zavecz Z, Horváth K, Éltet N, Török C, Pesthy O, Gombos F, Janacsek K, Nemeth D, 2019 Deconstructing procedural memory: different learning trajectories and consolidation of sequence and statistical learning. *Front. Psychol* 9 10.3389/fpsyg.2018.02708.
- Singh S, Daltrozzo J, Conway CM, 2017 Effect of pattern awareness on the behavioral and neurophysiological correlates of visual statistical learning. *Neurosci. Conscious* 2017 (1). 10.1093/nc/nix020.
- Singh S, Walk AM, Conway CM, 2018 Atypical predictive processing during visual statistical learning in children with developmental dyslexia: an event-related potential study. *Ann. Dyslexia* 68 (2), 165–179. 10.1007/s11881-018-0161-2. [PubMed: 29907920]
- Slone LK, Johnson SP, 2015 Infants' statistical learning: 2- and 5-month-olds' segmentation of continuous visual sequences. *J. Exp. Child Psychol* 133, 47–56. 10.1016/j.jecp.2015.01.007. [PubMed: 25757016]
- Slone LK, Johnson SP, 2018 When learning goes beyond statistics: infants represent visual sequences in terms of chunks. *Cognition* 178, 92–102. 10.1016/j.cognition.2018.05.016. [PubMed: 29842989]
- Smith EE, Grossman M, 2008 Multiple systems of category learning. *Neurosci. Biobehav. Rev* 32 (2), 249–264. 10.1016/j.neubiorev.2007.07.009. [PubMed: 17904637]
- Smith EE, Patalano AL, Jonides J, 1998 Alternative strategies of categorization. *Cognition* 65 (2–3), 167–196. 10.1016/S0010-0277(97)00043-7. [PubMed: 9557382]
- Smith GNL, Conway CM, Bauernschmidt A, Pisoni DB, 2015 Can we improve structured sequence processing? Exploring the direct and indirect effects of computerized training using a mediational model. *PLoS One* 10 (5), e0127148 10.1371/journal.pone.0127148. [PubMed: 25946222]
- Song S, Howard JH, Howard DV, 2007 Implicit probabilistic sequence learning is independent of explicit awareness. *Learn. Mem* 14 (3), 167–176. 10.1101/lm.437407. [PubMed: 17351140]
- Song S, Howard JH, Howard DV, 2008 Perceptual sequence learning in a serial reaction time task. *Exp. Brain Res* 189 (2), 145–158. 10.1007/s00221-008-1411-z. [PubMed: 18478209]
- Sonnweber R, Ravnani A, Fitch WT, 2015 Non-adjacent visual dependency learning in chimpanzees. *Anim. Cogn* 18 (3), 733–745. 10.1007/s10071-015-0840-x. [PubMed: 25604423]
- Squire LR, 2004 Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem* 82 (3), 171–177. 10.1016/j.nlm.2004.06.005. [PubMed: 15464402]
- Steinmetz JE, 2000 Brain substrates of classical eyeblink conditioning: a highly localized but also distributed system. *Behav. Brain Res* 110 (1–2), 13–24. [PubMed: 10802300]
- Stobbe N, Westphal-Fitch G, Aust U, Fitch WT, 2012 Visual artificial grammar learning: comparative research on humans, kea (*Nestor notabilis*) and pigeons (*Columba livia*). *Philos. Trans. Biol. Sci* 367 (1598), 1995–2006. 10.1098/rstb.2012.0096.
- Szegedi-Hallgató E, Janacsek K, Vékony T, Tasi LA, Kerepes L, Hompoth EA, Bálint A, Németh D, 2017 Explicit instructions and consolidation promote rewiring of automatic behaviors in the human mind. *Sci. Rep* 7 (1), 4365 10.1038/s41598-017-04500-3. [PubMed: 28663547]
- Tecumseh FW, Martins MD, 2014 Hierarchical processing in music, language, and action: lashley revisited: music, language, and action hierarchical processing. *Ann. N. Y. Acad. Sci* 1316 (1), 87–104. 10.1111/nyas.12406. [PubMed: 24697242]
- Teinonen T, Fellman V, Näätänen R, Alku P, Huotilainen M, 2009 Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci* 10 (1), 21 10.1186/1471-2202-10-21. [PubMed: 19284661]
- ten Cate C, Okanoya K, 2012 Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philos. Trans. Biol. Sci* 367 (1598), 1984–1994. 10.1098/rstb.2012.0055.
- Tettamanti M, Weniger D, 2006 Broca's area: A supramodal hierarchical processor? *Cortex* 42 (4), 491–494. 10.1016/S0010-9452(08)70384-8. [PubMed: 16881256]
- Thiessen ED, 2010 Effects of visual information on adults' and infants' auditory statistical learning. *Cogn. Sci* 34 (6), 1093–1106. 10.1111/j.1551-6709.2010.01118.x. [PubMed: 21564244]
- Thiessen ED, Erickson LC, 2013 Beyond word segmentation: a two- process account of statistical learning. *Curr. Dir. Psychol. Sci* 22 (3), 239–243. 10.1177/0963721413476035.

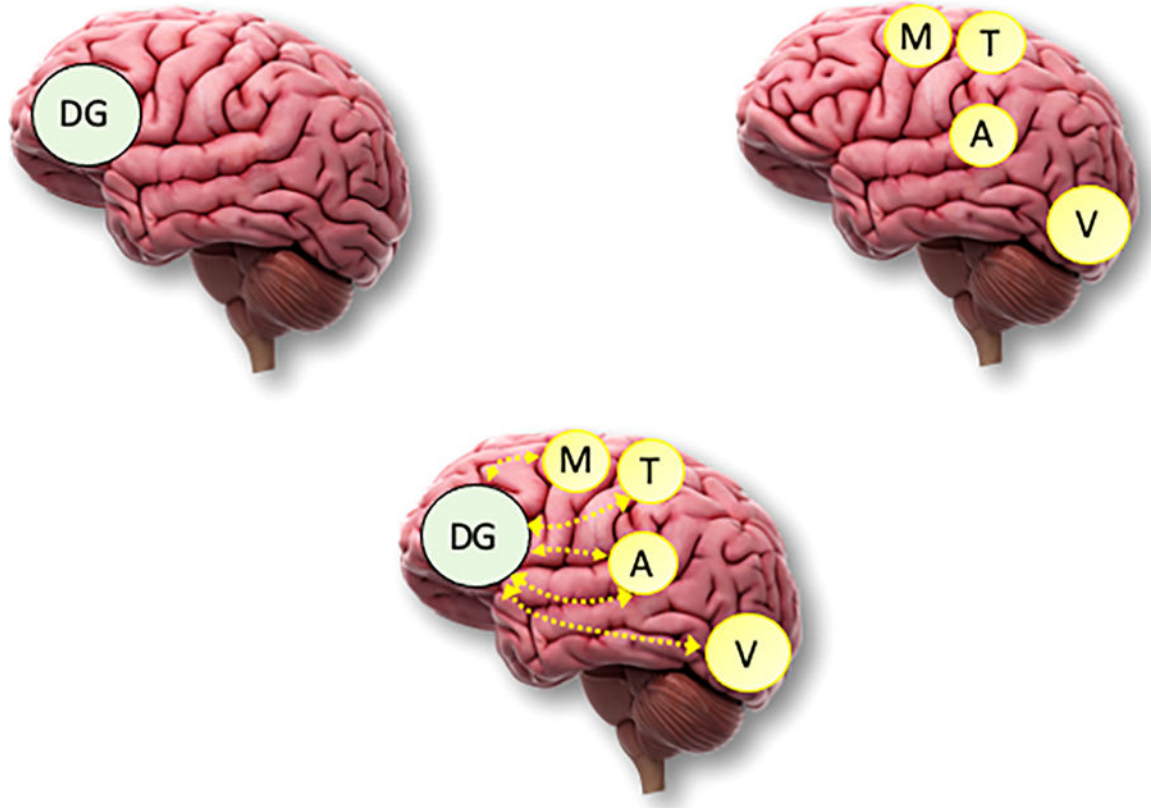
- Thompson-Schill SL, Ramscar M, Chrysikou EG, 2009 Cognition without control: when a little frontal lobe goes a long way. *Curr. Dir. Psychol. Sci* 18 (5), 259–263. 10.1111/j.1467-8721.2009.01648.x. [PubMed: 20401341]
- Tothathiri M, Rattinger M, 2015 Controlled processing during sequencing. *Front. Hum. Neurosci* 9 10.3389/fnhum.2015.00599.
- Tillman B, McAdams S, 2004 Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. *J. Exp. Psychol. Learn. Mem. Cogn* 30 (5), 1131–1142. [PubMed: 15355141]
- Timmann D, Drepper J, Frings M, Maschke M, Richter S, Gerwig M, Kolb FP, 2010 The human cerebellum contributes to motor, emotional and cognitive associative learning. A review. *Cortex* 46 (7), 845–857. [PubMed: 19665115]
- Toro JM, Sinnott S, Soto-Faraco S, 2005 Speech segmentation by statistical learning depends on attention. *Cognition* 97 (2), B25–B34. 10.1016/j.cognition.2005.01.006. [PubMed: 16226557]
- Tóth B, Janacek K, Takács Á, Kóbor A, Zavecz Z, Nemeth D, 2017 Dynamics of EEG functional connectivity during statistical learning. *Neurobiol. Learn. Mem* 144, 216–229. 10.1016/j.nlm.2017.07.015. [PubMed: 28803867]
- Turk-Browne NB, Jungé JA, Scholl BJ, 2005 The automaticity of visual statistical learning. *J. Exp. Psychol. Gen* 134 (4), 552–564. 10.1037/0096-3445.134.4.552. [PubMed: 16316291]
- Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK, 2009 Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *J. Cogn. Neurosci* 21 (10), 1934–1945. 10.1162/jocn.2009.21131. [PubMed: 18823241]
- Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM, 2010 Implicit perceptual anticipation triggered by statistical learning. *J. Neurosci* 30 (33), 11177–11187. 10.1523/JNEUROSCI.0858-10.2010. [PubMed: 20720125]
- Uhrig L, Dehaene S, Jarraya B, 2014 A hierarchy of responses to auditory regularities in the macaque brain. *J. Neurosci* 34 (4), 1127–1132. 10.1523/JNEUROSCI.3165-13.2014. [PubMed: 24453305]
- Ullman MT, 2004 Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92 (1–2), 231–270. 10.1016/j.cognition.2003.10.008. [PubMed: 15037131]
- Ullman MT, Earle FS, Walenski M, Janacek K, 2020 The neurocognition of developmental disorders of language. *Annu. Rev. Psychol* 71 (1), 389–417. 10.1146/annurev-psych-122216-011555. [PubMed: 31337273]
- van den Bos E, Poletiek FH, 2008v Effects of grammar complexity on artificial grammar learning. *Mem. Cognit* 36 (6), 1122–1131. 10.3758/MC.36.6.1122.
- Van Petten C, Plante E, Davidson PS, Kuo TY, Bajuscak L, Glisky EL, 2004 Memory and executive function in older adults: relationships with temporal and prefrontal gray matter volumes and white matter hyperintensities. *Neuropsychologia* 42 (10), 1313–1335. 10.1016/j.neuropsychologia.2004.02.009. [PubMed: 15193940]
- Virag M, Janacek K, Horvath A, Bujdosó Z, Fabo D, Nemeth D, 2015 Competition between frontal lobe functions and implicit sequence learning: evidence from the long-term effects of alcohol. *Exp. Brain Res* 233 (7), 2081–2089. [PubMed: 25910995]
- Vokey JR, Brooks LR, 1992 Salience of item knowledge in learning artificial grammars. *J. Exp. Psychol. Learn. Mem. Cogn* 18 (2), 328–344.
- Vuong LC, Meyer AS, Christiansen MH, 2016 Concurrent statistical learning of adjacent and nonadjacent dependencies. *Lang. Learn* 66 (1), 8–30. 10.1111/lang.12137.
- Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S, 2011 Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci* 108 (51), 20754–20759. 10.1073/pnas.1117807108. [PubMed: 22147913]
- Walk AM, Conway CM, 2016 Cross-domain statistical–sequential dependencies are difficult to learn. *Front. Psychol* 7 10.3389/fpsyg.2016.00250.
- White EJ, Hutka SA, Williams LJ, Moreno S, 2013 Learning, neural plasticity and sensitive periods: implications for language acquisition, music training and transfer across the lifespan. *Front. Syst. Neurosci* 7 10.3389/fnsys.2013.00090.

- Wilson B, Slater H, Kikuchi Y, Milne AE, Marslen-Wilson WD, Smith K, Petkov CI, 2013 Auditory artificial grammar learning in macaque and marmoset monkeys. *J. Neurosci* 33 (48), 18825–18835. 10.1523/JNEUROSCI.2414-13.2013. [PubMed: 24285889]
- Wilson B, Kikuchi Y, Sun L, Hunter D, Dick F, Smith K, Petkov CI, 2015 Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat. Commun* 6 (1). 10.1038/ncomms9901.
- Wilson B, Marslen-Wilson WD, Petkov CI, 2017 Conserved sequence processing in primate frontal cortex. *Trends Neurosci* 40 (2), 72–82. 10.1016/j.tins.2016.11.004. [PubMed: 28063612]
- Zhao J, Al-Aidroos N, Turk-Browne NB, 2013 Attention is spontaneously biased toward regularities. *Psychol. Sci* 24 (5), 667–677. 10.1177/0956797612460407. [PubMed: 23558552]
- Zwart FS, Vissers CTM, Kessels RPC, Maes JHR, 2019 Procedural learning across the lifespan: a systematic review with implications for atypical development. *J. Neuropsychol* 13 (2), 149–182. 10.1111/jnp.12139. [PubMed: 28990357]

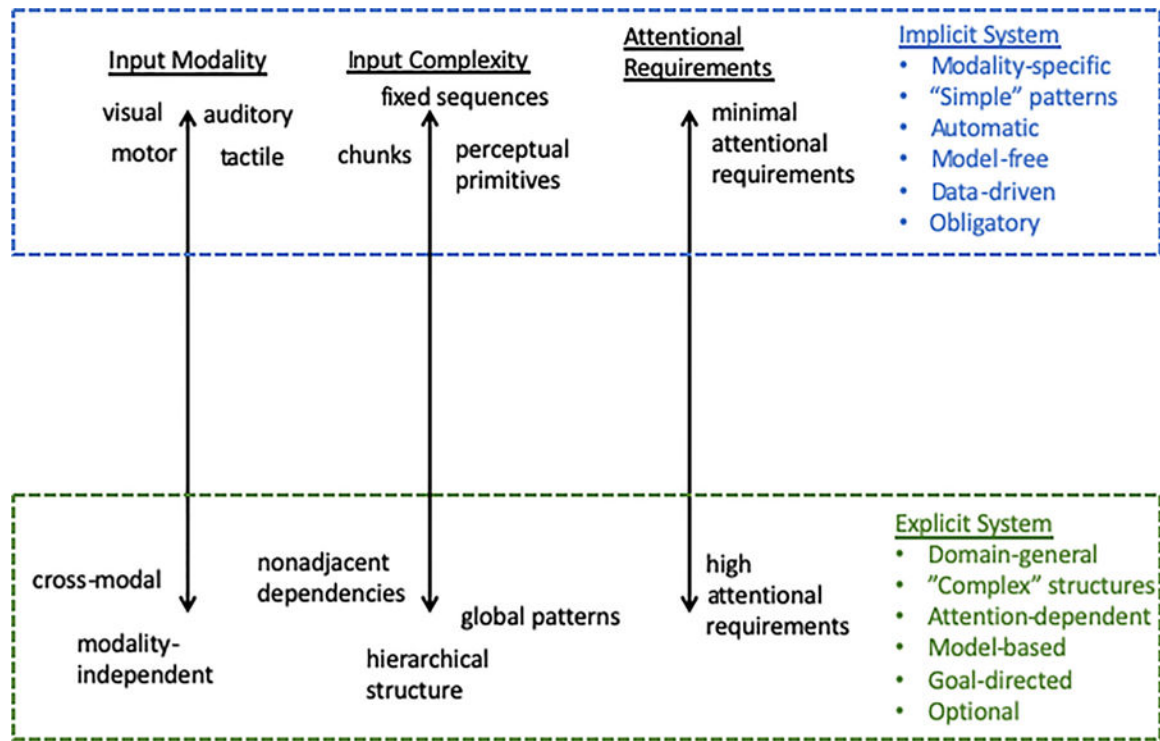


**Fig. 1.** Three orthogonal dimensions outlining a proposed task space for statistical learning.

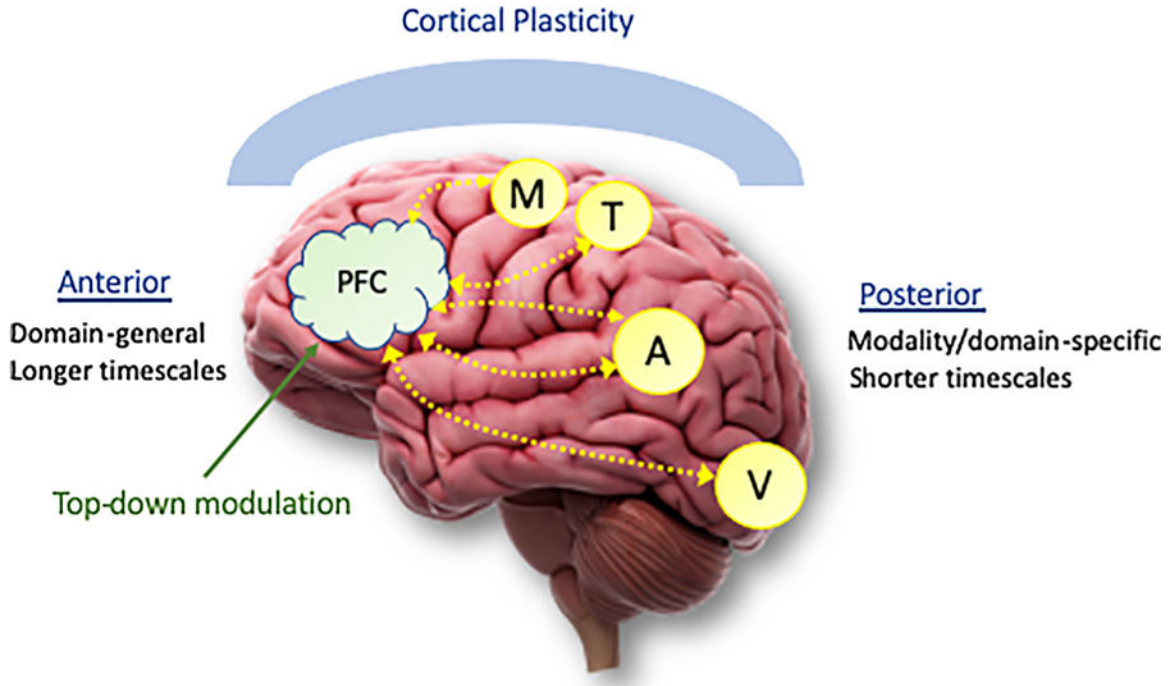




**Fig. 2.** Three candidate architectures for how input modality and domain interacts with statistical learning. (Top left): A domain-general (DG) account posits a single, unitary mechanism that implements statistical learning for all input modalities and domains. Here, the frontal lobe of the brain (prefrontal cortex) is offered as one potential domain-general brain region, though others areas are also likely candidates. (Top right): A modality/domain-specific account posits multiple, relatively independent mechanisms, each handling a specific type of input such as auditory (A), visual (V), tactile (T), and motor (M) patterns. For simplicity, only these four modality/domain-specific regions are illustrated though others would be posited to exist as well. (Bottom): Finally, perhaps the most viable account is one which combines domain-general and domain-specific architectures, here shown with connections between modality-specific brain regions and the prefrontal cortex.



**Fig. 3.** Depiction of two distinct learning mechanisms, “implicit” and “explicit”, in relation to the factors of input modality, input complexity, and attention.



**Fig. 4.** A multicomponent model of statistical learning. The first set of (“implicit”) mechanisms is based on the principle of cortical plasticity, which allows networks to adaptively change through experience. Posterior modality-specific networks allow for improved perceptual processing of input patterns spanning short timescales (A: auditory; V: visual; T: tactile; M: motor). More anterior networks can handle patterns spanning across perceptual modalities and over larger timescales. The second “executive” (or “explicit”) mechanism is rooted in frontal lobe (prefrontal cortex, PFC) and frontoparietal networks that mediate top-down control of attention and working memory to modulate learning and allow for the learning of more complex patterns such as nonadjacent dependencies. Not only does the executive system modulate plasticity-based learning; but through the principle of plasticity, as learning occurs, attention itself can be affected, drawing resources toward certain environmental events or stimuli, thus affecting the operation of the executive system.

**Table 1**

Summary of three common paradigms for studying statistical learning.

	AGL	SRT	Segmentation
Measure of learning	Explicit judgement of grammaticality (usually)	Reaction times (though response accuracy can also be used)	Explicit judgement of familiarity (usually)
Input structure	Defined by artificial grammar	Repeating sequences (usually)	Defined by transitional probabilities
Modality	Perceptual (any)	Visual-motor (mostly, though auditory stimuli can also be used)	Perceptual (any)
Generalization at test?	Yes	No	No

**Table 2**

A rough taxonomy of input structures learnable through statistical learning.

---

Perceptual primitives (repetitions, etc.)
Serial transitions
Chunks
Finite state grammars (of varying complexity)
Nonadjacent dependencies
Recursive / hierarchical / phrase structure

---

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

## Ten core principles for statistical learning.

**Table 3:**

<b>Principle</b>	<b>Description</b>
Multi-faceted	Statistical learning consists of two primary cortical mechanisms: 1) associative and perceptual learning based on principles of cortical plasticity; 2) top-down modulatory control of attention for learning more complex patterns
Cortical plasticity	The primary mechanism of statistical learning is mediated by cortical plasticity, which results in reduced neural activity and heightened behavioral/perceptual facilitation
Cortical processing constraints	Cortical plasticity is influenced by differences in processing capabilities (i.e., perceptual modality, timescales)
Top-down modulation of learning	Attention and working memory can modulate and gate statistical learning, by directing attention and processing to specific stimuli or features, allowing for the learning of patterns arrayed across temporal sequences and/or crossmodal dependencies; however, in some situations, top-down modulatory control may instead impede implicit pattern learning
Prediction and expectation	Central to statistical learning of temporal sequences is prediction and expectation of upcoming stimuli/events
Modality effects	Modality effects arise from plasticity of processing of perceptual brain networks
Input structures	Statistical learning can occur for a variety of input structures; simpler patterns such as serial transitions and adjacent dependencies can be learned automatically in a bottom-up fashion whereas learning more complex, global patterns require selective attention, working memory, and cognitive control
Bidirectional relationship with attention	Statistical learning and attention have a bidirectional relationship: attention can modulate learning and learning can affect levels of attention
Ontogeny	Plasticity-mediated associative learning dominates learning early in development; over developmental time, selective attention and cognitive control mechanisms progressively become available to influence learning
Phylogeny	Plasticity-mediated associative learning is relatively conserved across species; the learning of nonadjacent or hierarchical dependencies, which requires specialized cognitive mechanisms, varies across species