Abstract

Statistical learning (SL) is involved in a wide range of basic and higher-order cognitive functions and is taken to be an important building block of virtually all current theories of information processing. In the last two decades, a large and continuously growing research community has therefore focused on the ability to extract embedded patterns of regularity in time and space. This work has mostly focused on transitional probabilities, in vision, audition, by newborns, children, adults, in normal developing and clinical populations. Here we appraise this research approach, we critically assess what it has achieved, what it has not, and why it is so. We then center on present SL research to examine whether it has adopted novel perspectives. These discussions lead us to outline possible blueprints for a novel research agenda.

Keywords: *Statistical learning*, *regularities, distributional properties, patterning, information processing, cognition, language, memory.*

Public Significance Statement:

This review targets a fundamental theoretical construct in cognitive science, the learning of regularities in the environment. A critical analysis of past and present achievements of this field of research reveals possible novel experimental directions and theoretical perspectives.

1. Introduction

Statistical learning (SL)—learning from the distributional properties of sensory input across time and space—has become a major theoretical construct in cognitive science. Providing the primary means by which organisms learn about the regularities in the environment, SL is involved in a wide range of basic and higher-order cognitive functions such as vision, audition, motor planning, event processing, reading, speech perception, language acquisition, semantic memory, and social cognition, to name a few. SL, therefore, is taken to be a necessary building block of virtually all current theories of information processing, and its importance in advancing theories throughout the cognitive and brain sciences cannot be overestimated (see Saffran & Kirkham, 2018, for review).

Although the roots of SL can be traced back nearly a century (see Christiansen, 2019, for review), the recent impetus for SL research can be found in the published finding of Saffran and her colleagues (Saffran, Aslin, & Newport, 1996), showing that infants are sensitive to transitional probabilities (TPs) of syllables in a continuous speech stream. The paper made two critical points: first, that information regarding word boundaries could be detected in the input from differences in TPs within and between word boundaries. Second, that children can rapidly perceive and use this information to parse the continuous speech input. This paper sparked intense theoretical debates in the domain of language acquisition (e.g., Christiansen & Curtin, 1999; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Peña, Bonatti, Nespor & Mehler, 2002; Seidenberg, 1997; Yang, 2004). It was seen as providing evidence that experience-based learning mechanisms can potentially account for language learning—hence, there is no need to revert to nativist accounts of language acquisition (Chomsky, 1965).

Saffran and her colleagues were careful in their original paper to qualify the scope of their claims: "It remains unclear whether the statistical learning we observed is indicative of a mechanism specific to language acquisition or of a general learning mechanism applicable to a broad range of distributional analyses of environmental input (p. 1928).” However, given the intriguing possibility that Saffran et al. (1996) raised, SL research has expanded broadly, and related debates spilled over to other domains of learning and cognition. To date, the *Science* paper by Saffran and colleagues has reached nearly 4900 citations, with about a stable rate of more than 300 citations per year[1].

Research on learning regularities was pervasive decades before the paper by Saffran et al. (1996), mainly through implicit learning using artificial grammar learning (AGL; e.g., Reber, 1967) and serial-reaction time (SRT; e.g., Nissen & Bullemer, 1987) paradigms (see Christiansen, 2019; Hunt & Aslin, 2001; Perruchet & Pacton, 2006, for discussions). However, the groundbreaking finding by Saffran and her colleagues inspired a large research community to focus on the ability to extract embedded patterns of regularity in time and space, mostly TPs, across vision, audition, and tactile modality, in newborns, children and adults. Figure 1 shows how this field has exploded in particular over the last decade (i.e., since 2006) relative to the overall expansion rate of research in other major domains of cognitive science[2]. Our search shows that the first two decades of research on SL (1996-2016) have produced over 760 papers[3], we hereafter refer to this body of work as “past” research. In the most recent two years alone (2016-2018), over 150 papers on SL have been published. We consider this set of articles to represent the “present” state of the art in SL research. Given that the field is now expanding at an almost exponential rate, it seems like a

[1] Impact according to Google Scholar, June 2019.

[2] The data for the other major domains of research was extracted by entering the labels presented in Figure 1 (e.g., “attention”, “memory”, etc.) into the same Scopus search procedure used to identify the papers on statistical learning. The choice of normalizing publication rates relative to 2006 was taken as it is the mid-point point of our data. The overall trends presented in this figure hold, however, across a range of different normalization schemes.

[3] The search included all papers with SL in their title, abstract, and/or their keywords, excluding machine learning, see our discussion in section 2.1 Methodological considerations.

good time to take stock of what has been accomplished so far, what is missing from the current research focus, and why this might be so. This is the first aim of the present paper. We do so by examining the empirical work of "past" SL research in Part 1 versus "present" work in Part 2, considering several important criteria. These include, the scope of empirical research in terms of range of methodologies, the validity of theoretical presuppositions, the extent of integration with adjacent fields of cognitive science, and the extent of ecological validity. In the third part, these discussions are harnessed to point to several avenues regarding how future research can address some of the missing pieces.

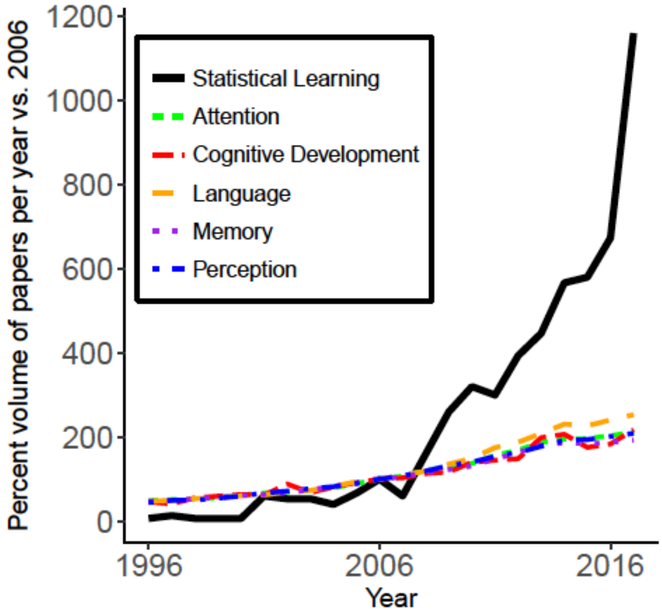

Figure 1. Percent volume of papers per year relative to 2006. The number of papers published in 2006 is taken as the baseline from which percent volume is measured.

We should clarify from the outset that the first two parts of the paper are not aimed to provide a comprehensive review of all empirical work that has been done in the field, but to critically discuss some of the directions (and also misdirections) that this field has taken since the original paper by Saffran and colleagues in 1996. Here, we do not take issue with a specific finding, an individual study, its experimental design, inferences, or conclusions. Problems at this level are not the target of the present discussion. Instead, our paper aims to focus on broader conceptual and methodological issues. We outline the fundamental characteristics of the initial SL research program when taken as a whole, distilling out what it has and has not accomplished. To foreshadow what follows, our take is that SL research has provided considerable important evidence, insights, and theoretical contributions. However, research paradigms often get entrenched in methodologies, basic axioms, prototypical metaphors, and homogeneous ways of thinking about particular issues. Pointing these out has the potential of moving the field forward, opening novel research avenues. This is the focus of Parts 1 and 2 of our discussion. In Part 3, we offer suggestions for ways in which the field may move forward by building on past work and dealing with current limitations.

1.1. Tracing the boundaries of SL phenomena

Before we begin the review of SL research, we must first ask and answer a fundamental question: What should be considered SL? Typically, a research community can at least agree on the scope of the issues that they are studying, yet there is no broadly agreed upon formal definition[4]. An imperative first step is, therefore, a precise description of our inclusion criteria, which allows the drawing of a clear line regarding what phenomena belong to our present investigation and what do not. We should emphasize that our claims in this section are not ontological in nature. Rather, they are aimed at providing a common ground for discussions by clarifying from the outset which phenomena will undergo scrutiny and which will not. While we do recognize that other potential demarcation lines can be drawn, we

[4] Anecdotally, at the conference on *Interdisciplinary Advances on Statistical Learning* (Bilbao, 2017), the question of how to define SL was at the center of a panel discussion that concluded without reaching any general agreement. Opinions ranged from a narrow definition of SL, to "all learning is SL".

naturally assume that our inclusion criteria are constructive in the sense that they focus on the core aspects and phenomena related to SL. Here, we do not voice a principled disagreement with the claim that *all* (or almost all) learning is, in fact, statistical learning. We simply argue that even if convincing arguments can be put forward in its defense, adopting it will not be constructive in providing nuanced distinctions, precise predictions, and a tractable scope for future SL research.

The present paper targets, therefore, all phenomena related to *perceiving and learning any forms of patterning in the environment that are either spatial or temporal in nature*. Patterning requires, by definition, that there would be more than one stimulus (an independent stimulus is not a pattern), and that there would be more than a single occurrence of events in the stream (one appearance of something is not a pattern). This inclusion criterion is wide enough to incorporate all learning of ordered auditory, visual, or tactile stimuli, but precludes instances of one-shot learning (e.g., Laska & Metzker, 1998). It also precludes simple frequency effects when a single stimulus is repeated again and again leading to changes in its representational state in the visual, auditory, or somatosensory cortex (e.g., Grill-Spector, Henson, & Martin, 2006). To clarify, we will not consider a rhythmic repetition of a single stimulus (e.g., a metronome's tick, a flickering light at a given frequency), to be SL. Hence, entrainment of neural populations to this form of "regularity" is not within the present scope. Indeed, current evidence suggests that entrainment to rhythm per se (timing expectation) is very different than predictions regarding upcoming structure (e.g., Ding, et al., 2016). In a similar vein, a sudden change or cessation of rhythmic repetition, such as revealed in typical oddball paradigms, are also excluded (e.g., the repetition of /pa/ occasionally replaced by /ba/, e.g., Getzmann & Näätänen, 2015; Näätänen, Gaillard, & Mäntysalo, 1978).

In this sense, we focus on how organisms encode and use the regularities related to *relationships between recurrent events* (frequencies, associations, distributions, positions) to enable and enhance learning, and how neural changes occur due to such patterning. Hence, the boundaries of SL phenomena that are of interest for this paper do not include typical reinforcement learning that investigates how probabilistic reinforcement shapes behavior, or how supervised, semi-supervised, or unsupervised learning can be used to simply summarize the environment. Rather, our discussion targets phenomena where the organism not only mirrors the statistical properties of the environment (for example, mirroring the TPs structure within an input stream), but uses the statistical information to derive representational content that go beyond mirroring (for example, deriving representations of "words" given the differences in TPs within the input). This is what made SL potentially influential in the cognitive sciences. We should emphasize that within this scope, we do not focus just on learning TPs, but on a range of potential regularities. One may learn, for instance, that A occurs more frequently than B, that B is always in the middle of a sequence of three stimuli, that C co-occurs with D, or that ABCD is not a grammatical event. These are but a few examples of SL, hence our definition is anything but narrow. Thus, in addition to the work directly inspired by the Saffran et al. (1996) study, we also include AGL, SRT learning, and cross-situational learning[5] under the umbrella of "statistical learning." Importantly, though, our definition avoids the presupposition that "everything is SL", because if everything is SL, practically, nothing substantial can be said about it.

2. Part 1: Past accomplishments in SL

[5] Cross-situational learning involves learning the referent for individual words across multiple exposures, in which each exposure is ambiguous with respect to the words' identity (e.g., Yu & Smith, 2007). From an SL perspective, this requires computing distributional statistics over possible word-referent mappings given their patterns of co-occurrence.

In this part we aim to review and summarize SL past research, first by evaluating its scope in terms of research questions and methodologies. We then examine various theoretical perspectives on SL mechanism(s), mainly whether one or more mechanisms underlie the learning of regularities. Next, we assess how SL has been integrated within other research areas in cognitive science given its initial promise to inform most theories of information processing. Finally, we discuss what we see as potential weaknesses or pitfalls of this research enterprise, focusing on issues such as extent of theoretical specification, and ecological validity.

2.1 Methodological considerations

We start our discussion by outlining our methodology for reviewing SL research. Our guidelines in structuring our review of past research followed the flow chart of PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses, see Figure 2). PRISMA offers state-of-the-art protocols for appraising research efforts (see, http://www.prisma-statement.org). Our first decision point in this flow chart concerned the inclusion criteria for constructing the database of experimental papers on SL. Our search thus targeted all journal articles that contained the term “statistical learning” in their abstract, title, or keyword list, published from 1996 to 2016. In terms of screening, we excluded a few specific journals where “statistical learning” is used frequently in a machine-learning or analytical interpretation that is not related to cognition (e.g., IEEE journals on information theory, image processing, etc.).

Admittedly, given our discussion of what SL is, there is no doubt a broader community doing research related to SL per our definition, without self-identifying their research as such. We discuss in length further on the reasons for such demarcation line between research paradigms (see our section on “domain integration”). However, our aim in this review was to specifically target the community that identifies itself as engaging in SL research, and we assumed that our search criteria would encompass this community in an optimal way. Our exploration procedure undoubtedly excluded a number of papers that, for one reason or another, omitted a reference to SL in their title, abstract or keywords (we note, for example, that the influential study of Aslin, Saffran & Newport, 1998, on the computation of transitional probabilities statistics by infants, falls into this category). However, an exhaustive search to locate all potential papers that examine the learning of regularities throughout the full scope of the cognitive sciences is not a tractable enterprise, as it requires a manual inspection of thousands and thousands of papers. Importantly, expanding the search by devising a list of other potential keywords or perhaps a list of potential authors known to work on SL, would be a thorny issue. Indeed, it is unlikely that the SL research community would agree on exactly what those keywords or authors should be. Critically, any choice of keywords (e.g., “word segmentation”, “conditional probabilities”, etc.) would inevitably create a sampling bias towards inclusion of specific topics. Because there are many ways to assemble a database of papers for reviews and meta-analyses, each one with its own pros and cons, we have sought here to make explicit the rational for our decisions regarding inclusion or exclusion criteria.

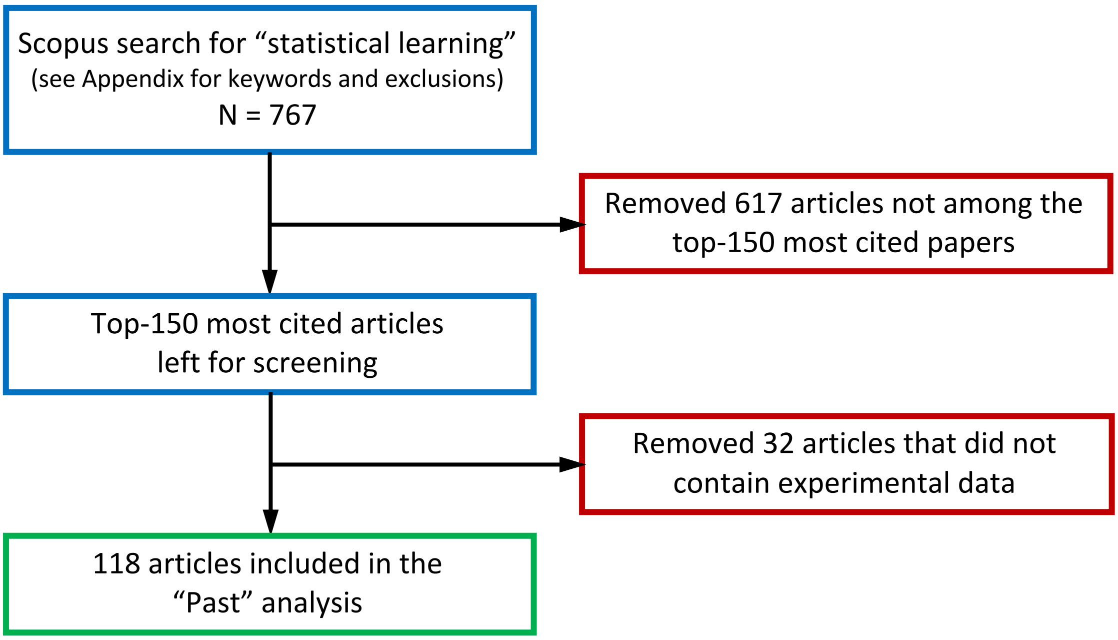

Figure 2. PRISMA flowchart for the Past SL research literature search.

Screening: Our search returned 767 papers, out of which 628 had been cited at least once, with a total number of 16902 citations[6]. We then manually inspected the 150 most highly cited articles in this set to ensure that they indeed relate to SL broadly construed. Together, these articles had 14032 citations. In other words, the articles that we focus on account for 83% of the total number of citations to the SL literature. These articles had an average of 94 citations each (min = 22, max = 549, excluding the original paper by Saffran et al.). We should emphasize that our aim in setting a cutoff by citations was to obtain insights regarding what has made a given study impactful within and outside the SL research community. Admittedly, overall number of citations is correlated with years since publications, creating some disadvantage for the very recent papers. However, because there is no clear mathematical algorithm regarding how to factor in number of years since publications for measuring impact, and given that we devote a full section of the paper to

[6] Citation statistics in this part of our discussion are based on the Scopus database accessed on July 11th, 2017. Numbers reflect relative impact at this time point.

analyze and discuss recent SL research (see Part 2 on “Present directions in SL”), our cutoff regarding citations served as an adequate screening procedure for assessing the impact of past research. Finally, given that our focus was on empirical research, we filtered the 150 impactful articles to require that they would have at least some experimental component (see the PRISMA flow chart in Figure 2). This led us to set aside 32 review, opinion, or modeling papers.

We take the final set of 118 articles to be broadly representative of the first two decades of empirical research performed by the SL research community (see supplementary material for the full listing of articles). We may not have selected the full population of papers that were generated in these two decades of research on SL, but we have assembled an unbiased corpus that allowed us to adequately characterize the main advances in the field.

2.1. Scope of research

Following Saffran et al. (1996), debates have elevated SL to be a substantial theoretical construct in cognitive theory. While at the onset it was taken to provide a viable explanation for identifying word boundaries, with time it has been expanded to cover learning regularities in many areas of cognition, extending well beyond language. It would be fair to say that in the many hundreds of studies that followed the original auditory TP learning task by Saffran et al. (1996), researchers often tailored the task’s parameters to address closely related questions. For example, the task was imported into the visual modality virtually as is, with shapes replacing syllables (e.g., Kirkham, Slemmer, & Johnson, 2002; Siegelman & Frost, 2015; Turk-Browne, Junge, & Scholl, 2005). A somewhat more significant change involved presenting regularities in terms of spatial location in a grid, rather than a temporal location in the stimulus stream (Fiser & Aslin, 2001). Rather than focusing on adjacent regularities such as AB, researchers have studied sequences of the form AxB, where x is a randomly selected stimulus (Gómez, 2002; Newport & Aslin, 2004; Onnis, Christiansen, Chater & Gómez, 2003). Instead of studying TPs of 1, sensitivity to lower TPs has been investigated (e.g., Bogaerts, Siegelman, & Frost, 2016). Instead of learning one stream of regularities, participants have been exposed to two sets, either within (e.g., Gebhart, Aslin, & Newport, 2009; Karuza et al., 2016), or between (e.g., Emberson, Conway, & Christiansen, 2011; Mitchel & Weiss, 2011; Weiss, Poepsel & Gerfen, 2015) modalities. Instead of testing human infants or adults, researchers have studied monkeys (e.g., Hauser, Newport, & Aslin, 2001), rodents (e.g., Toro & Trobalón, 2005), and birds (e.g., Lu & Vicario, 2014; see Santolin & Saffran, 2018, for a review of SL across species). Rather than testing normally developing children or adults, researchers have used the SL task with various special populations such as SLI or autism spectrum disorder (e.g., Evans, Saffran, & Robe-Torres, 2009; Hsu, Tomblin, & Christiansen, 2014; Obeid et al., 2016), and dyslexics (Gabay, Thiessen, & Holt, 2015; see Lammertink, Boersma, Wijnen, & Rispens, 2017, for a meta-analysis).

We should note that the TP learning task of Saffran et al. (1996) was not the only game in town. A parallel line of research employed the original paradigm offered by Reber (1967) for studying implicit artificial grammar learning (AGL). Here participants were typically presented with sequences of stimuli generated by a miniature grammar, and then asked to classify a new set of sequences according to whether they were derived from the grammar or not (e.g., Altmann, Dienes, & Goode, 1995). Although the AGL task was originally taken to tap implicit learning, it permeated into SL research (e.g., Conway & Christiansen, 2005, 2006; Tunney & Altmann, 1999). Whereas the task was originally taken to reflect rule learning, it is well accepted today that performance in the AGL task may be explained by overall judgments of statistically-related surface similarity between "grammatical" items that were presented during familiarization and those presented at test (e.g., Conway & Christiansen, 2005; see Pothos, 2007, for a review). Thus, similar to the TP learning task, participants are provided with a relatively brief exposure to repeated regularities, after which learning is assessed through a two-alternative forced choice (2AFC) test phase.

Without doubt, each of these lines of research has provided a greater understanding of how the kind of learning demonstrated by Saffran et al. operates in a somewhat broader range of circumstances. This has led to significant theoretical advances that cannot be overestimated. Importantly, replications that track down the nature of effects with small variants of paradigms and materials are critical for advances in science. On the other hand, constructive advances in science are characterized by a state of affairs in which large and diverse sets of data converge to carve out a given theoretical construct. This is because any one type of evidence will necessarily be imperfect or lacking in some respect, providing only partial constraints on the theory. In this sense, the relationship between data and theory is akin to a pyramid wherein a broad empirical foundation supports a specific theoretical claim. The major theoretical appeal of SL is that it hinted at a potentially overarching explanation of learning regularities in a general sense, covering deep and thorny issues such as how language is learned, how generalizations are made, how discrimination occurs, how categories are carved—in essence, impacting almost the whole scope of cognitive capacities. It is therefore important to evaluate to what extent the first two decades of SL research and empirical findings support these ambitious theoretical goals. Our analysis below provides a summary of the distribution of key design features in our representative sample of past SL studies from 1996 to 2016. Here we highlight a few illuminating observations:

- As shown in Figure 3, 60% of all the empirical papers on SL used a variation of the original task by Saffran et al. (1996), embedding sequences of auditory or visual stimuli with different TPs in continuous stream of input (below we refer to these as the "*TP papers*"). The rest mostly used a variation of the AGL task

(16%) or investigated cross-situational learning (11%). This suggests that the field was primarily made up of results from three closely related tasks.

- Out of the *TP papers* examining auditory SL, 84% used syllables as linguistic units, similar to the original Saffran et al. (1996) study.
    - 24% of the papers using syllabic units, included exactly the same "words" that Saffran et al. selected for their original study.
- Considering the patterns of regularity investigated, 82% of *TP papers* embedded either triplets or pairs of stimuli in the input stream.
    - Over 90% of these papers used TPs of 1.0, that is, perfect regularity between elements within a pattern.
- Considering the number of patterns that are the object of learning, 59% of *TP papers* employed 8 patterns or less; nearly 50% of all these experiments used four patterns (or less) as in the original Saffran et al. study.
- 86% of these studies used patterns that were uniform in size (i.e., either all trigrams or all bigrams).
- 89% of all empirical investigations used a familiarization stream that did not exceed 30 minutes, while 61% of studies settled on 10 minutes of familiarization or less.
- In 72% of all papers, participants were given passive exposure to an input stream, which we contrast with a (minimally) active task where the learner is doing something other than watching or listening to the input, or orienting to an attention-grabbing stimulus in infant studies.
- 51% of all studies monitored SL performance via a 2AFC test following familiarization. Similarly, an additional 30% targeted infants using preferential looking methods.

- 96% of all studies dealt with humans.

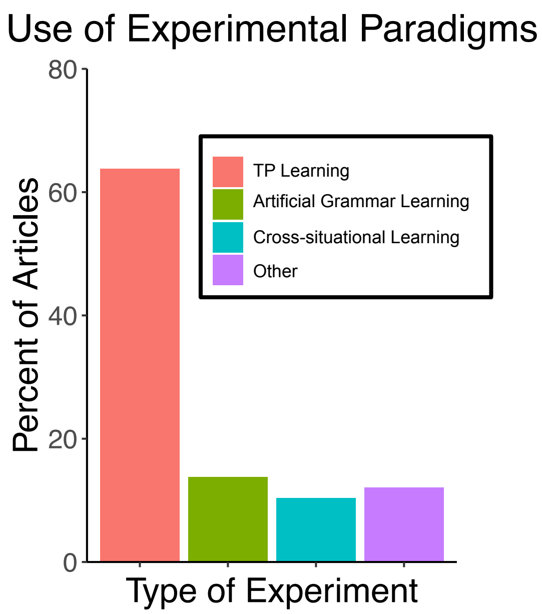

Figure 3. Frequency of use of different experimental paradigms in the most frequently cited SL articles.

These statistics indicate a substantial uniformity in the first two decades of SL research. We should note that within the large set of 767 papers, one can identify specific studies that have broken this mold (we discuss examples of such papers later on). However, our analysis shows these papers to be the exceptions rather than rule, and most studies were constrained to relatively homogeneous methodologies. This state of affairs often occurs when a groundbreaking experimental finding and methodology spurs on an entire field of research and is by no means unique to SL. A parallel situation, for example, occurred in the domain of reading, where the lexical decision task (Meyer & Schvaneveldt, 1971) has been used in thousands and thousands of experimental papers, with time eventually leading to a partial merging of theories of visual word recognition with theories of lexical decision per se. Because the original finding of Saffran et al. (1996) seemed to speak to a wide range of theoretical questions and view-points, the task itself was adopted by a diverse set of laboratories across adjacent fields to address a very wide scope of theoretical questions. In that sense it is understandable why a large proportion of the experimental work derived from the seminal task reported by Saffran et al. (1996) has key design features in common.

However, this has also led to a situation wherein the theoretical claims regarding the broad relevance of SL to cognitive science has outpaced the accumulated empirical support, which has remained relatively narrow in scope and confined to a restricted range of methodologies. Here we outline a number of examples of this phenomenon. First, although in the domain of speech several cues for segmentation (e.g., stress) have been considered, many of the original SL experiments have focused on an existence proof that a given population can extract high-probability TPs from an input stream. Regularities in the environment do not consist, however, only of TPs, and are not confined to high TPs. Second, the recurring patterns—the object of learning—were in most cases either pairs or triplets of visual or auditory stimuli. Regularities are typically significantly richer in terms of the number of elements involved, and are more abstract, often involving some level of generalization. Third, the individual elements were typically very uniform (e.g., syllables or tones of the same length; visual figures of the same kind and size), whereas real-world regularities often consist of a heterogeneous set of inputs, where instances of the same element may vary along a variety of dimensions (e.g., the same syllable will have different acoustic realizations depending on contexts and speakers; visual elements will occur across different backgrounds, etc.). Fourth, learning has been confined to relatively short durations, where participants might see each regularity 8-30 times over the course of a 5- to 15-minute familiarization phase. Learning regularities in the real world, however, spans a much larger period of time, mostly without consecutive repetitions. Fifth, learning typically has been assessed in a subsequent test-phase comprised of a series of 2AFC questions, which contrasts pairs or triplets that follow or violate the regularities in the input stream. This does not tap into how learning occurs and accumulates on a step-by-step basis, and may provide a distorted view of what exactly has been learned (see Christiansen, 2019; Siegelman, Bogaerts, Christiansen, & Frost, 2017; Siegelman, Bogaerts, Kronenfeld, & Frost, 2017, for extensive discussion).

We will return to these issues while examining the "present", to see whether and how field had changed recently, leading us to the discussion regarding what has to be done in the future.

2.2 Perspectives on SL mechanism(s).

The unitarian view of SL

As described above, much of past SL research has focused on providing an existence proof that a range of regularities can be learned. To a first approximation, this research has revealed commonalities across different domains. Sensitivity to TPs in the input stream was found not just with spoken syllables as Saffran et al. (1996) originally showed, but also with non-linguistic auditory material such as pure tones (e.g., Creel, Newport, & Aslin, 2004; Saffran, Johnson, Aslin, & Newport, 1999) and computer sound effects (e.g., Gebhart, Newport, & Aslin, 2009; Siegelman & Frost, 2015). In the visual modality, evidence for TP sensitivity was found with abstract visual shapes (e.g., Glicksohn & Cohen, 2013; Turk-Browne et al., 2005), colored simple shapes (Kirkham et al., 2002), faces (Emberson et al, in press), real-world scenes (e.g., kitchen scenes, Brady & Oliva, 2008), cartoon aliens (Arciuli & Simpson, 2011), natural visual scenes (e.g., landscapes, Schapiro, Gregory, & Landau, 2014), and fractal patterns (Schapiro, Kustner, & Turk-Browne, 2012). Once existence proofs of SL have been established across a range of domains, and in the absence of a widely accepted neurocomputational theory of how SL operates, verbal theorizing about the commonalities that have been discovered has often led to the assumption that basically the same abstract computations occur across the range of domains. In most studies this has not been taken as an explicit well-defined presupposition. Rather, it was typically taken as a loose working metaphor, defining SL as “a (or the) mechanism with which cognitive systems discover the underlying structure of the input”.

Here we argue that focusing on commonalities alone, although useful in some respects, may nevertheless lead to a theoretical emphasis on an overly abstract and underspecified common denominator among a large set of findings. When the theory is vague and underspecified, it can essentially be interpreted to be consistent with many data patterns, and it is unable to generate specific a priori predictions to guide future research. In contrast, focusing on *differences* in performance has the promise of providing important constraints regarding the viability of a unitary theory, leading a clear path regarding in what way the theory is incorrect and should be revised (see Evans & Levinson, 2009, for a similar argument regarding linguistic universals and the putative universal grammar). The focus on commonalities in a range of SL experiments has often led SL researchers to assume that SL is akin to a central device that learns regularities across a range of perceptual stimuli. Performance in the small handful of tasks was taken to be a good proxy of the device’s capacity. There is substantial evidence, however, that is inconsistent with a strong unitary theory of SL even though it has driven a substantial part of past SL research.

Evidence for a pluralist view of SL

We argue that SL, across different domains and modalities, is performed by partially overlapping yet distinct networks. Thus, on the one hand, brain areas dedicated to processing specific sensory information (visual, auditory, or somatosensory) are tuned to the statistical properties of the input stream (e.g., Hasson, 2017). On the other hand, the output of these sensory areas serves as input for other higher-order brain areas (e.g., MTL: Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Striatum: Lieberman, Chang, Chiao, Bookheimer, &

Knowlton, 2004). What is learned, therefore, is *the product of the interactions between modality-specific and higher-order brain areas*. In a nutshell, the brain includes a range of mechanisms that contribute to the perception and learning of patterned regularities. Consequently, to predict and explain a specific SL phenomenon one cannot simply focus on the computations performed by a unitary device (see Frost, Armstrong, Siegelman, & Christiansen, 2015; Siegelman et al., 2017, for discussion).

From a behavioral perspective, studies examining individual performance in SL tasks do not lend support for a unitary view of SL. First, although SL performance in a given modality is relatively stable within an individual (Siegelman & Frost, 2015; Siegelman et al., 2016), it does not reliably predict his/her ability in learning regularities in another modality. As Siegelman and Frost (2015) showed, performance in a visual statistical learning (VSL) task with abstract shapes does not correlate with performance in an analogous auditory statistical learning (ASL) task, with spoken syllables (but see further discussion of this point and additional recent findings in Part 3). The latter also does not correlate with performance in a similar ASL task with computer sounds rather than syllables. In the same vein, performance in any of these SL tasks does not correlate with performance in an SRT task, measuring implicit sequence learning. Since individual performance in one task across two timepoints using similar experimental settings would be expected to be highly correlated (Siegelman et al., 2015, 2018; Erikson et al., 2016), shared computations across modalities should have resulted in at least some correlations in performance. Evidence from the AGL task is not compatible with a unitary theory either. Conway and Christiansen (2006) have shown that learning two grammars can proceed without interference as long as they are implemented in two modalities. In the same vein, transfer of learning has been shown to be very limited across modalities (e.g., Redington & Chater, 1996; Tunney & Altmann, 1999). Taken together, these behavioral data do not fit with a simple architecture centered on a unitary SL device. From another perspective, recent evidence suggests a very different developmental trajectory for visual vs. auditory SL: whereas VSL performance linearly improves with age, ASL does not change much across development in school-aged children (Raviv & Arnon, 2017) though it does appear to change during early development (Emberson et al., in press). Such modality-specific developmental differences are not consistent with a unitary system for SL.

Admittedly, all these behavioral data and conclusions, at a first blush, stand in contrast with recent evidence from cognitive neuroscience, neuroimaging studies, and computational modeling of the hippocampus. The main evidence stemming from these studies is that the hippocampus (or one of its sub-regions, for example, CA1) is activated in various SL tasks (e.g., Turk-Browne, Scholl, Chun, & Johnson, 2009, Schapiro et al., 2014; Schapiro et al., 2017), suggesting that it is akin to a central device for all SL computations. However, the same studies also showed activation in modality specific areas (see Frost et al., 2015, for a review). Schapiro et al. (2014) reported a case of an amnestic patient with hippocampal damage, who exhibited no SL abilities, arguing for the necessity of the medial temporal lobe system for SL. In contrast, Covington, Brown-Schmidt and Duff (2018) showed that patients with hippocampal damage were not uniformly at chance, and demonstrated above-chance performance in some SL task variants. Importantly, a range of studies implicated the striatum in AGL (e.g., Liberman et al., 2004), and the left inferior frontal gyrus in ASL (Karuza et al., 2013). For example, using AGL, Knowlton, Ramus and Squire, (1992) have shown that while amnesic patients, the majority of which had confirmed or suspected damage to the hippocampus, had poor recognition of the grammatical exemplars presented during familiarization, they could nevertheless discriminate between grammatical and ungrammatical exemplars at the test phase, similar to controls. On the other hand, Christiansen Kelly, Shillcock and Greenfield (2010) found that agrammatic aphasics with damage to the left frontal areas were unable to discriminate between test items in an AGL task, despite being able to complete the training task at the same level as matched healthy controls. This suggest that left frontal areas may also play a role in AGL, similar to TP learning.

Within this context we should emphasize that cognitive neuroscience as a field is increasingly moving in the direction of structural and functional connectivity analyses. Underlying these advances is the growing appreciation that the mere activation of a given brain region cannot be interpreted as evidence of its unique computational role as it is typically densely interconnected with many other brain areas. From this perspective, deeper understanding the neurobiological underpinning of SL may require to also consider functional connectivity evidence in a range of SL tasks[7].

To summarize, at least at present, there is no unequivocal demonstration that all learning of statistical regularities requires hippocampal computations, nor is there neurobiological evidence supporting SL as a unitary device. Although it is currently unclear whether TP learning and AGL rely on the same or different brain areas, it is nonetheless possible that despite both being concerned with the learning of regularities, they may be tapping different forms of computations (however, admittedly, to our knowledge there is no experiment that tested this directly by combining the two tasks together within individuals). This leads to our conclusion that the overall neurobiological and coordinated behavioral evidence does not favor a unitary view of SL.

The cost of the unitary view to SL research

The main cost incurred by the unitary view comes from its inherent stranglehold on the development of SL as a theoretical construct. If SL is a componential and complex ability,

[7] We are indebted to Lizz Karuza for making this point.

then research should map its possible components, providing a testable theory of the different set of computations that each component employs, specifying in what ways they differ or overlap with other components' computations, and importantly, how these components interact. Although some initial work has been done on this front, much more extensive theoretical, empirical and computational work is needed to flush out these aspects of SL theory.

The unitary approach also had negative consequences in the area of individual differences (e.g., Arciuli & Simpson, 2012; Christiansen et al., 2010; Conway, Bauernschmidt, Huang, & Pisoni, 2010; Frost, Siegelman, Narkiss, & Afek, 2013; Shafto, Conway, Field, & Houston, 2012). In this branch of studies, researchers aimed to tie SL abilities to other cognitive abilities, selecting a given SL task without an a priori theory regarding why the chosen task was selected, rather than another (see Siegelman et al., 2017, for a critical discussion). In essence, such individual-difference studies treated SL as a "black box", without specifying what exactly has driven an obtained correlation between performance in some SL task and some cognitive ability. This approach also runs the risk that researchers will serially search for SL tasks that "work", in terms of predictive power, without overt discussions regarding why other tasks are not as predictive of a given cognitive ability.

Construing SL as a unitary device also had impact on the computational work on SL. It has motivated modelers to develop computational accounts of how one or two basic computations such as tracking distributional frequencies, calculating TPs, or chunking of frequently occurring patterns, explain the range of SL phenomena (e.g., French, Addyman & Mareschal, 2011; Perruchet & Vinter, 1998; Thiessen, Kronstein, & Hufnagle, 2013). In principle, the development of such domain-general models built on basic computations has had substantial merits, as these models offer explicit evidence of how learning input regularities could occur. The models also offered testable predictions to sharpen our understanding of how regularities could be extracted and represented. Nevertheless, they were mainly inspired by the qualitative commonalities in SL phenomena, offering yet again, an existence proof that regularities can be learned, rather than simultaneously focusing on how fine-grained differences in learning outcomes emerge for different parameters of the task (e.g., cross-modal differences, extent of familiarity with the stimuli and prior knowledge, event complexity, etc.). In this sense, the models have offered mostly coarse-grained insights.

In sum, as a metaphor, the unitary view of SL has had the important benefit of focusing research on a well-defined set of phenomena. However, metaphors in cognitive science run their course in terms of their utility. Once they have served their purpose, they should be abandoned, for if not, they will end up dominating and becoming entrenched in the ways researchers think about the empirical phenomena. Based on the empirical evidence at hand and with the benefit of hindsight, a pluralist approach to SL would appear to be a more constructive way of thinking about SL. Adopting pluralism about mechanisms would lead to a better understanding of various SL phenomena.

2.3 SL and other cognitive faculties

Given the theoretical assumption that most cognitive functions to some degree involve the learning of regularities, SL should be a fundamental facet of understanding most domains of cognition. An important criterion for assessing SL research is, therefore, whether it has indeed established deep links with research in other areas of cognition or whether it has developed as an isolated construct. In evaluating the extent of integration of SL research with other aspects of cognition, we consider two independent dimensions. The first focuses on the breadth of the temporal window of learning. This concerns the integration of learning regularities with what we know about memory systems that operate on different timescales.

We refer to this as *Timescale integration*. The second, perhaps more important dimension, refers to the extent to which evidence regarding learning of regularities in a range of domains of cognitive study permeates SL theory, and vice versa. We refer to this as *Domain integration*. As we elaborate below, integration of past SL research is lacking on both of these dimensions. We illustrate this in Figure 4 in the domain of visual SL.

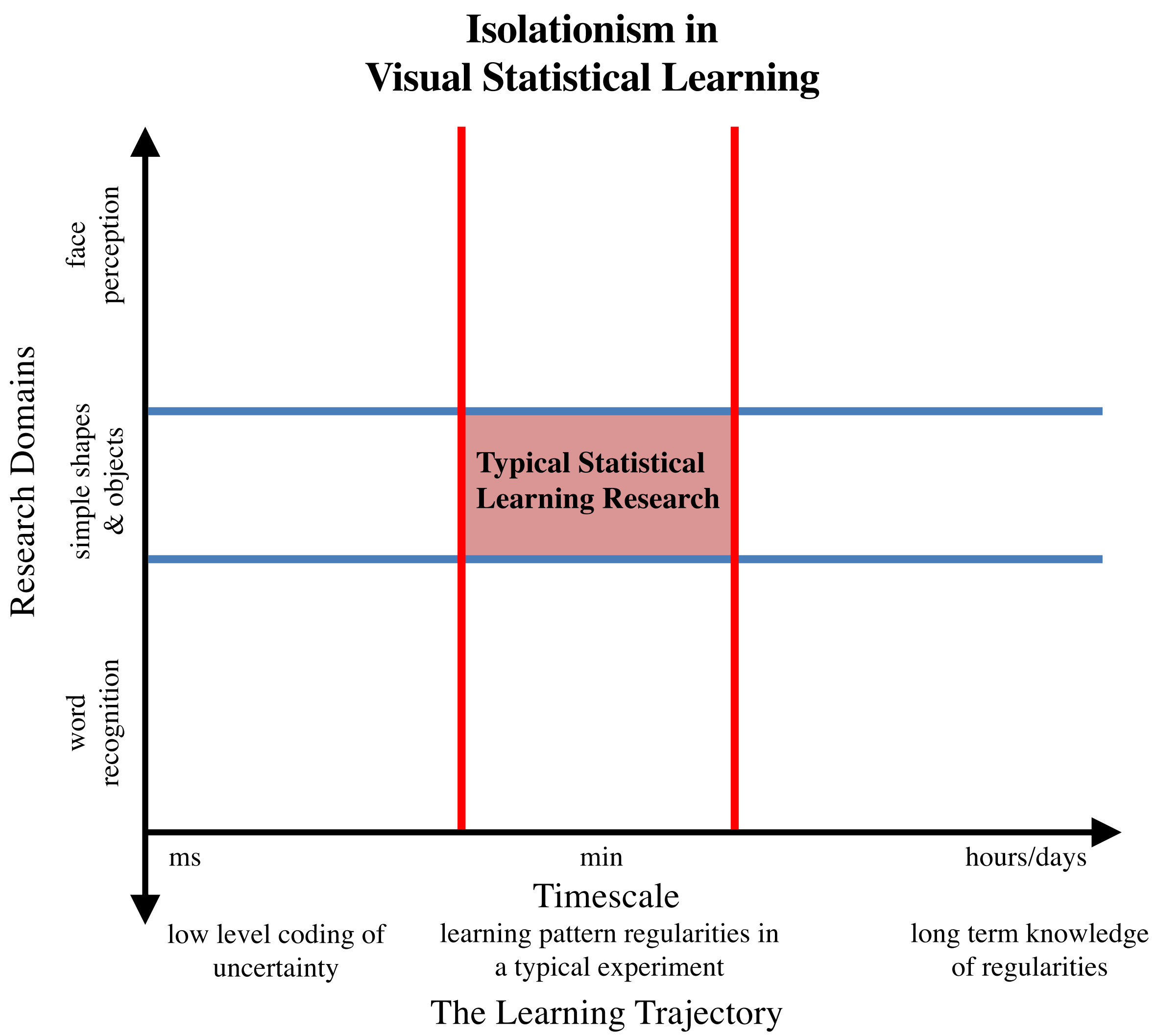

Figure 4. Timescale and Domain integration in SL research, focusing on visual SL.

Timescale integration

SL past experiments have typically considered learning on the timescale of minutes. This timescale is a derivative of 1) the type of regularities (and representations) that are to be learned (e.g., recurrent syllabic triplets, pairs of abstract shapes, etc.), and 2) the minimal time needed to reach an existence proof that the targeted pattern regularities can be learned. This has created a highly-constrained focus on a specific fraction of the continuous learning trajectory, which starts with the low-level encoding of uncertainty, and ends in long-lasting accumulated knowledge of the environment. As presented by the red lines of Figure 3, SL research has typically been squashed into a small part of this learning trajectory within a given modality. Timescale integration thus concerns establishing connections between the shorter and longer timescales of this trajectory. That is, how low-level neural coding of uncertainly feeds into the computations of higher-level pattern regularities (see Hasson, 2017, for a discussion), and how pattern-level regularities consolidate and result in long-lasting representations, merging with existing knowledge of the environment (see Gómez, 2017; Coutanche & Thomson-Schill, 2015, for discussion of this problem and possible directions). This has not been the focus of most of the past SL research.

Domain Integration

Domain integration concerns overcoming the artificial split of learning phenomena into separate research areas, aiming to achieve a level of constructive interaction between these areas. For example, contextual cueing, scene perception, visual word recognition, and face perception are all concerned, one way or another, with the learning of regularities by the visual system. For SL theory to achieve its initial promise and become an important building block in a wide range of cognitive functions, evidence from all these research areas should permeate SL research and vice versa. This, however, does not seem to be the case, as we illustrate with the following prominent examples.

To begin, consider reading research. Of the thousands of studies concerned with literacy acquisition and determinants of proficient reading performance, very few have considered SL research, looking into how computations of regularities in the visual system lead to high-quality orthographic representations, shaping visual word processing abilities. A recent vision of reading by Grainger, Dufau and Ziegler (2016; see also the recent OB1 model of reading by Snell, van Leipsig, Grainger, & Meeter, 2018), for example, acknowledges that progress in this research area has been hampered by limited cross-fertilization. Nevertheless, this account of skilled reading centers on visual constraints such as crowding and visual acuity, ignoring how SL mechanisms shape orthographic representations and letter processing to eventually determine performance (see Frost, 2012, for a discussion). This is in spite of substantial evidence linking reading performance to visual SL abilities (e.g., Arciuli & Simpson, 2012; Chetail, 2017; Frost et al., 2013).

Another example is research on memory. Although SL clearly involves memory at different levels—both short- and long-term—there has been little interaction between the two fields of research (though see Brady et al., 2009). Indeed, when Chekaf, Cowan and Mathy (2016) conducted a study of how repeated exposure to sequences of visual elements could be compressed into pairs (chunks) based on their features (shape, color, size), there was no mention of SL. Strikingly, they even predicted behavioral patterns that closely resemble those observed in SL experiments involving TP learning: "within-chunk transitions would more often be made correctly than between-chunk transitions" (Chekaf et al., 2016: p. 101). Likewise, past work on SL has rarely made direct connections with the memory literature (though see Schapiro et al. 2012, 2014, for exceptions, and Christiansen, 2019; Isbilen, McCauley, Kidd, & Christiansen, 2017, for current perspectives).

To be clear, the split between research areas is a typical product of historical divisions of research communities into predefined research areas. It is not a characteristic unique to SL research. In that sense, just as SL research is insulated from adjacent research paradigms, the reverse is also true. However, in the case of SL, this isolation is particularly problematic because it stands in the way of SL playing a stronger and more expansive role in theories of cognition, as it should. Importantly, the split between domains has led researchers to investigate the learning of regularities without considering the specific roles they subserve in the different cognitive functions. Consider, for example, two sub-domains of language, speech perception and orthographic processing. Both of these linguistic functions undoubtedly require SL, but markedly differ in the type of statistical information that is the target of learning. Speech consists of a continuous unfolding input, whereas print has critical spatial characteristics. Words in speech are co-articulated, so that their boundaries have to be extracted through e.g., TPs or chunking, whereas word boundaries in print are given for free by blank spaces in most languages. Efficient print processing requires representations of sublexical letter combinations with some letter-position invariance (i.e., quickly registering *ing* in *knowing,* and *knowingly,* see Frost, 2012, for review), whereas speech does not. These are just few examples demonstrating that there is little gain in discussing "SL computations" in a vacuous general context, without tying them to the specific cognitive operations and especially to the nature of representations that are characteristic of a given domain. As such, the problem of a domain split is particularly problematic in the context of SL research relative to other domains of cognition because, in a sense, SL research is supposed to tell us something fundamental about virtually every domain of cognition, but without deep integration in other domains it is unable to do so.

The two aforementioned examples provide an illustration of our concerns regarding the integration of SL with other fields. It is important, however, to quantify the overall integration of SL research objectively. One possible approach to do so is to examine the ratio of citations of SL research by other research communities. We thus focused on the proportion of citations from within the field (as defined by our literature search), and from outside the field. Figure 5 plots the results. It shows that whereas the first decade of SL papers had an even distribution of citations from within and outside the field, the last decade has seen a sharp drop of external citations. Although part of this change in proportion likely is a product of the expansion of the SL community, in general references to the experimental findings of SL research seem to be increasingly confined to the SL community alone, characterizing a pattern of growing isolation from other research communities. Although an increase in within-field citations are to be expected as a research community grows—as ideas and methods are being refined—the dramatic drop in external citations since 2008 is nonetheless cause for concern.

In summary, several different perspectives converge to indicate that SL research is relatively isolated with regards to research on other areas of cognition. Because SL as a theoretical construct has been taken to provide a viable and parsimonious explanation of how regularities are learned across domains of cognition, the isolation of SL research is disadvantageous not only for advancing SL theory but also for advancing theories in other relevant domains.

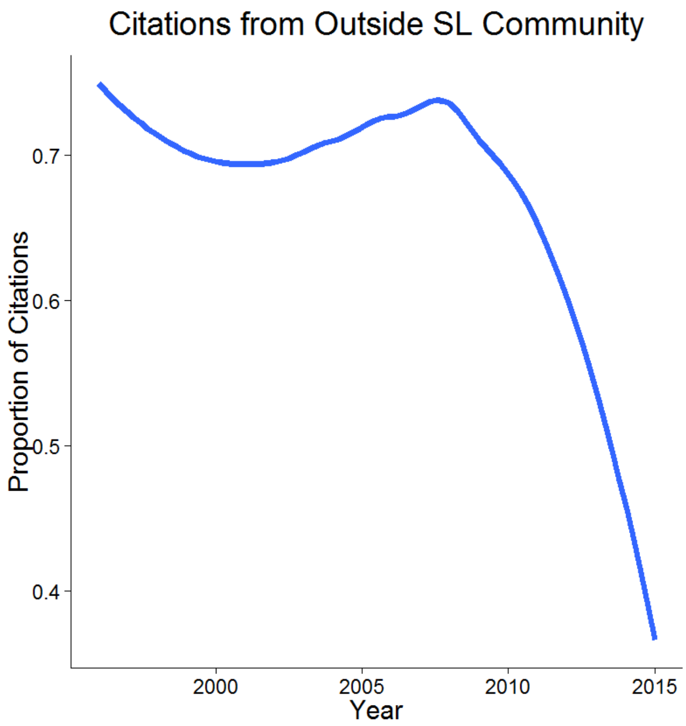

Figure 5. Proportion of citations that originate from outside the SL community, for each of the top 150 most cited articles in our database. Year refer to citations to SL research within that year. The "SL community" is defined as all the articles in our database.

2.4 The degree of specification in SL theory

Our starting point is that productive advances in any research field rest on developing precise operational terms that lead to fine-grained distinctions and well-specified predictions. This is because with abstract sketches, individual researchers might use a given term to mean very different things, obscuring how specific findings relate to one another. Without a precise language for scientific discourse, researchers may have different assumptions and intuitions regarding key questions without putting these issues on the table explicitly.

Consider for example the question of what is learned when patterns of regularities are embedded in a continuous input stream, such as the stream of syllables in the Saffran et al. (1996) experiment. Initially, there was no detailed account of exactly what was learned and it was unclear whether different researchers had similar views of how SL occurs. Only when researchers began to be more explicit about this question through computational formalisms (e.g., Endress & Mehler, 2009; Perruchet & Poulin-Charronnat, 2012; but see Perruchet & Vinter, 1998, for an earlier example) did clear and fundamental differences in perspective became apparent. The consequent discussions revolved around whether TPs alone could lead to word-like chunks, or whether other cues such as prosodic information underlie stream segmentation. Without taking a stand on this specific matter, it exemplifies the critical value of specification.

Similarly, a recent model of SL in the hippocampus (Schapiro et al., 2017) offers an explicit and testable theory of the central role of the hippocampus in SL. This stems from making precise claims regarding the nature of representations in different parts of the hippocampus, as well as the computations performed in each distinct neuroanatomical area. Such work offers specific novel predictions that can open this account to falsification and refinement through coordinated empirical studies. Hence, only when researchers are clear and precise about what they hypothesize regarding learning representation and processing, can contrasting views be revealed and resolved empirically.

The above examples illustrate, in many ways, the successful consequences of developing well-specified accounts of particular questions relevant to the SL research community. Considering the first two decades of SL research, important questions nevertheless have remained without precise answers. To name a few: What are the regularities in the environment that are the object of perception in a given domain? How are these regularities represented following learning? What exactly is the learning mechanism(s) for various types of regularities? What is the relevant timescale for different learning situations? What are the processes that constrain the learned representations? What are their learning outcomes? How does the measurement of performance in a task interact with the learning process? Obviously, answers to these questions would depend on providing more specific descriptions of what exactly is learned and how. A considerable portion of past SL research, however, has been relatively vague about these issues, mainly reverting to abstract verbal sketches, and principally concluding that a domain-general mechanism has led to learning the regularities in the experiment.

This vagueness has led to a paucity of intense debates in SL past research. It contrasts quite strikingly with research in adjacent fields with links to SL such as memory, attention, and perception, which are characterized by intense controversies sparked by well-specified theories and models. For example, is there a domain-specific neurobiological module dedicated to processing faces? (see Gauthier & Tarr, 1997; Plaut & Behrmann, 2011); is attention object based or location based (see Chun & Jiang, 1998; Logan, 1996; Roelfsema, Lamme, & Spekreijse, 1998)?; is the structure of semantic memory determined by statistical regularities or innate constraints (e.g., Caramazza & Sheldon, 1998; Rogers et al., 2004)? These debates are a direct consequence of developing very detailed theories and have contributed to advancing our understanding of the aforementioned domains by making assumptions explicit and by providing testable a priori predictions for evaluating these assumptions. These more specific accounts have also forced researchers to be more precise in their discourse, preventing findings from being taken to conform to fuzzy verbal theories, which in turn makes falsification unlikely. There is every reason to expect that being similarly explicit and detailed in the development of SL theory would also lead to similar large advances in our collective understanding of how regularities are learned. Linking back to the previous section of the paper, it also seems obvious that a well-specified theory has even greater potential for deep integration with other adjacent fields of cognition.

One salient symptom of abstract sketching is how SL has been defined. A common occurrence in the field was merging the theoretical construct of SL with the experimental task that is supposed to tap into it: If what participants do in the task is SL, then SL is what participants do in the experiment. We refer to this circularity as *Tautologism*. The problem with Tautologism is self-evident: If the mechanism underlying the theoretical construct is explained by describing what participants do in the task that is taken to tap into it, then the theoretical construct does not stand by itself, and is bound to the description of task performance. With this state of affairs, little can be said about its internal structure, and a theory of SL is no more than a redescription of the data. A similar phenomenon has occurred in the domain of intelligence measurement, where there was no agreement regarding an independent definition of human intelligence, mainly because of issues related to cultural bias in measuring intelligence. Eventually, the solution was to define IQ by reverting to Tautologism: “IQ is what IQ tests measure”. However, whereas the research community on intelligence has acknowledged this problem explicitly (see for example, Mackintosh, 1998), Tautologism in SL research has been pervasive, and typically implicitly embedded in the research assumptions. Here are but a few quotes illustrating this, including one of our own:

— “The best-known example of this statistical learning ability is the use of the conditional relation between speech sounds” (Thiessen, 2011).

— “An individual’s capacity for SL can be measured in a number of ways. For instance, it can be assessed by asking a participant to watch a continuous stream of evenly paced, individually presented items on a monitor.” (Arciuli & Simpson, 2012).

— “The rationale of such approaches is to show that some measure of statistical learning ability, *as assessed in tasks requiring implicitly learning relations among probabilistic sequences*, is correlated with performance on one or more tasks involving language (Onnis, Frank, Yun, & Lou-Magnuson, 2016; italics added).

—"We hypothesized that if a general statistical-learning ability underlies learning to read in a new language that is characterized by a novel set of statistical regularities, then relative success in learning the transitional probabilities of random visual shapes would predict the speed and success of learning to read a new language." (Frost et al., 2013).

These quotations show how the ability of SL is explicated by describing what participants do in a narrow set of tasks focusing on the learning of TPs in continuous input. Tautologism is a consequence of underspecification and lack of preciseness because it treats SL as a black-box device. Without a precise description of candidate representations and computations operating upon them, the explanation for SL is no more than a redescription of performance in SL tasks. Implicit Tautologism conveys the false impression that the mechanisms underlying SL are understood to a first approximation, and all that remains is to sharpen our understanding of SL by tweaking the parameters of the task to work out the details. Moreover, an underspecified description of potential SL representations and computations may lead researchers to oversimplify the learning problems, thereby reducing ecological validity.

2.5 Assessing ecological validity

The original interest in whether children could parse an input stream based on statistical regularities alone was well motivated in and of itself, providing groundbreaking insights. However, as revealed in our literature review, past SL research has typically focused on tasks wherein only a very restricted type of statistical regularities is available for learning in the input stream, and participants were passively exposed to these inputs. We elaborate below on how each of these trends has impacted the ecological validity of what we know about SL and its role in a range of cognitive operations.

Let us consider first the types of statistical information. In the initial work by Saffran and colleagues (1996), the focus was on whether a continuous stream of four artificial three-syllable words could be segmented based solely by learning the differences in TPs within versus between items. The use of a small set of artificial nonwords had the benefit of providing a powerful and transparent demonstration that *in principle* the continuous stream can be parsed solely by attending to differences in TPs alone. Since that initial study, a number of studies have provided evidence that the original findings generalize across different types of stimuli and domains. In this vein, Pelucchi, Hay and Saffran (2009) replicated the typical TP finding, but with richer stimuli based on a natural language, presenting infants with child-directed speech in Italian. Similarly, Schapiro et al. (2012) have used highly complex fractal visual objects to examine the learning of their co-occurrence. However, in terms of ecological validity, learning the regularities in the environment rarely involves learning TPs alone. In the domain of language, for example, Chinese readers learn that for 80 percent of logographs, the semantic radical appears on the left side, whereas the phonetic radical appears on the right side. In Spanish, speakers learn that words cannot end with the phoneme /m/. In English, native speakers learn that the bigram LT tends to appear in word-final position. In Semitic languages, speakers learn the constraint of obligatory contours: roots can have the form of ABB but not of AAB—the doubling consonants can only occur at the second and third root position (e.g., Berent, Everett, & Shimron, 2001). Similarly, Marcus et al. (1999) has shown how infants can learn regularities at a higher of abstraction than simple TPs, such as AAB (generalizing this reduplicative pattern to novel stimuli). All these examples are not easily captured by an exclusive focus on TPs alone, but may be captured by mechanisms sensitive to other type of regularities, though this has received scant attention (but see Christiansen, Conway, & Curtin, 2005).

In the same vein, most past studies have used a fixed value of TPs throughout the stream, often with TPs of 1.0 within the repeated units (i.e., fully deterministic regularity). While learning such a simple regularity is an ideal starting point, the statistical regularities governing patterns in the real world span a wide range of values. While at some domains TPs can be exceedingly high, in others, such as language, they can be exceedingly small. The process of learning a large set of low probability regularities over time necessarily involves additional memory processes related to long-term memory and consolidation (see Gómez, 2017). This creates a rift between the experimental simplification and the ecological equivalent it is supposed to reflect (Yang, 2004; see Bogaerts et al., 2016, for manipulation of TPs).

Similarly, as our database shows, past work focused to a large extent on presenting patterns composed of the same number of elements (i.e., pairs, triplets). However, if all patterns composing the stream have the same length, the problem of segmenting the stream into its constituents is vastly simplified. To be concrete, if the stream is composed of N patterns, and all patterns are composed of K elements, finding the boundaries of one single pattern removes all remaining uncertainty regarding the identity of the remaining patterns in the stream. Indeed, there have been suggestions that some perceptual cues in the stream drive the segmentation procedure (e.g., Endress & Mehler, 2009). Obviously, if the stream was composed of patterns varying in length, say K=1-5 elements per pattern, as all languages are (no language has words of a fixed syllable length), segmentation would be a much more challenging problem to solve, and may require additional mechanisms. Although the leading computational models of SL (e.g., PARSER, Perruchet & Vinter, 1998; SRN, Elman, 1990; TRACX, French et al., 2011) are set to deal with non-uniform continuous streams, at present there is little experimental evidence regarding learning performance of complex streams, and what the underlying mechanisms and computations for such learning might be.

Finally, SL research has almost exclusively focused on methods in which participants are passively exposed to an input stream, where the only learnable information is that which is contained in the stream. Such an approach implicitly adopts an apathetic perspective of the learner, taking organisms to be automatic absorbers of environmental regularities. That some pattern regularities can be learned by mere exposure is not contested. Indeed, children and adults have been shown to automatically segment a continuous input stream, even while engaging in a secondary covert task, such as drawing computer illustrations (Arciuli, Torkildsen, Stevens, & Simpson, 2014; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Nevertheless, just because learning can easily occur incidentally in such passive circumstances does not mean that SL typically is a passive process where regularities are automatically detected, registered, and learned. Indeed, in a more ecologically valid setting, this type of "pure" statistical learning is rarely the case.

Consider for example the question of how children learn to map the spoken forms they hear into the objects they see. Given the extensive uncertainty regarding the correct mapping, this is a clear SL problem. Two recent lines of research that have explored how this is achieved have reached similar conclusions: children are not passive learners but are actively shaping the learning process by constraining the information to which they are exposed. For example, Smith and her colleagues (e.g., Clerkin, Hart, Rehg, Yu, & Smith, 2017; Smith, Yu, Yoshida, & Fausey, 2015) have shown that the manner by which children focus on objects throughout development determines what is in the center of their visual field and for how long, thereby reducing significantly the extent of ambiguity regarding the correct mappings of object-label pairs. Breaking into language through SL is, thus, determined by an intricate set of specific interactions of the learning child with his/her environment. In a different related line of work, Frank and his colleagues (e.g., Frank & Goodman, 2014; Yurovsky & Frank, 2015) have shown that children consider a variety of social cues to actively seek out additional constraints beyond the information presented to them, so as to try to resolve ambiguity during learning (see also Goldstein & Schwade, 2008). Taken together, these studies demonstrate that *a good SL theory is one which considers and focuses on the interactions of the organism with the environment* (see also Dale & Christiansen, 2004).

Overall, patterns in the natural environment are vastly less constrained than in typical SL experiments, are characterized by more subtle and varied statistical regularities, and the learning situations are different than those tested in typical statistical learning tasks. Naturally, initial SL experimental work intentionally distilled the learning situations into easily tractable pieces to obtain a set of existence proofs that learning can occur in principle. Nevertheless, with time, the lack of methodological expansion of SL research has led to reduced ecological validity.

To summarize Part 1, the first two decades of self-identified SL investigations have formed a large research community that extensively examined the learning of regularities in the auditory, visual, and tactile modalities. An important part of this research was harnessed to provide an existence proof that humans and non-humans are sensitive to the statistical properties of the input, focusing to a large extent on transitional statistics. This was done by using variations of a relatively narrow set of experimental tasks. We have outlined the important merits and promise of this methodological approach but also its potential weaknesses and limitations in making SL an important theoretical construct in cognitive science. We now move on to examine the most recent SL research, aiming to provide a perspective regarding the trajectory that this field has been taking most recently.

3. Part 2: Present directions in SL

Our aim in this part is to examine whether the initial characteristics of past SL research have undergone changes in recent years, and if so in what direction. Having this goal in mind, we focused on SL papers published in the period between 2016 to 2018. Our search used the same criteria as before, targeting all journal articles that contained the term "statistical learning" in their abstract, title, or keyword list. Given the brief period since publication, the number of citations could not serve as a reliable criterion. Our only requirement was therefore that papers would be cited at least once. Our search returned 151 such papers. Manual inspection revealed that 5 of these papers were not related to SL, and 16 additional papers centered on theoretical reviews, corpus analyses, computational modeling or description of statistics in various domains, which left us with 130 experimental papers—a sample that has about the same size as the one the served the "past" analysis (see Figure 6 for a PRISMA flowchart).

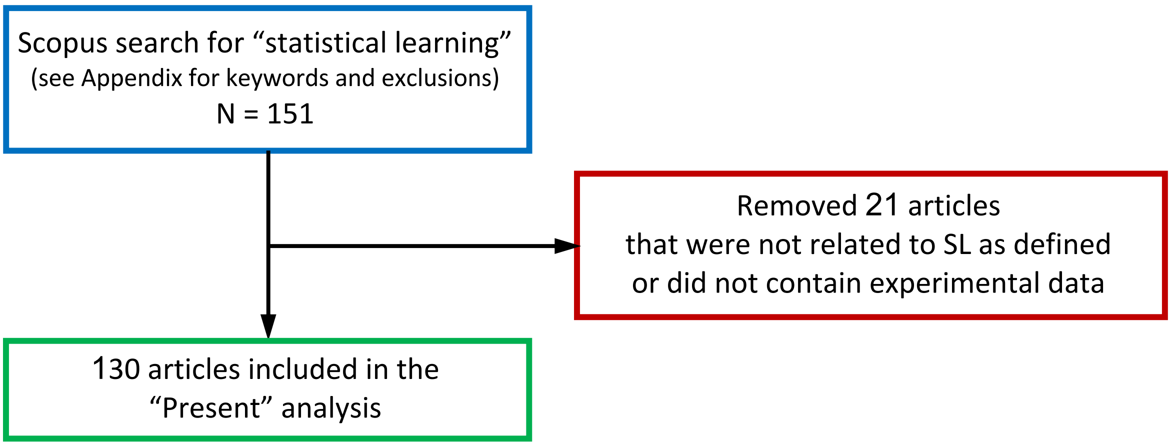

Figure 6. PRISMA flowchart for the Present SL research literature search

Our analysis of "present" research followed then similar criteria as "past" research. Thus, we first focused on whether the scope of methodologies and research questions have widened. Our findings are presented in Figure 7. The figure reveals less uniformity in the key design of studies, suggesting that present SL research moves towards greater expansion in research questions and methods. Here we highlight some important observations:

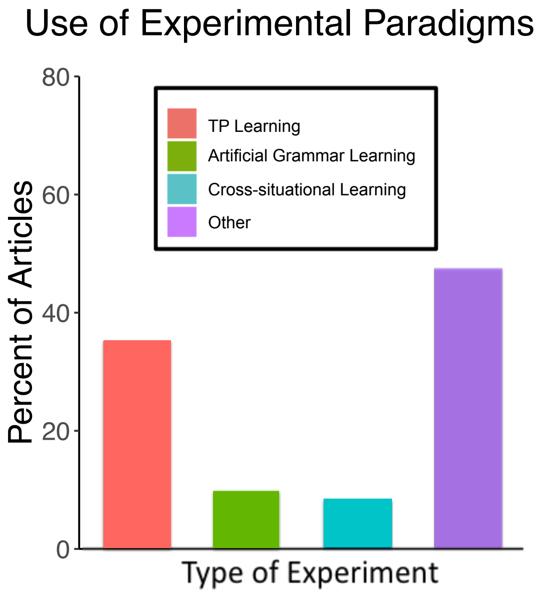

Figure 7. Frequency of use of different experimental paradigms in SL articles 2016-2018.

- The traditional SL paradigms are still dominant. However, the original task reported by Saffran et al. (1996), which constituted over 60% of past experimental research, constitutes 35% overall of the present studies, with AGL and cross-situational learning accounting for 10% and 9% of the distribution, respectively.
- Although most experiments that involved familiarization with a continuous input in the auditory modality used syllables as linguistic units (about 60%), 40% extended research to music pitch, beat, or linguistic tones.
- About 26% of all empirical studies involved neurobiological measures, such as EEG recording, BOLD activation, connectivity, and neural oscillations (see our

next section on novel approaches to SL). This reflects the permeation of neuroscience into SL research.

- 7% of all studies focused on special populations such as dyslexics, aphasics, children with SLI, Williams syndrome, and autism.
- Considering the structural properties of experimental designs, out of the papers that employed the typical task of presenting "words" embedded in a continuous input stream, 90% used TPs of 1.0, and 98% used patterns that were uniform in size (though see Trecca et al., 2019, for an exception). These were, as a rule, triplets or pairs, where number of patterns range from 4-9.
- About 72% of studies involving SL or AGL tasks used a 2AFC test following familiarization, which was typically brief (in the range of minutes).
- As before, only 6% of studies examined species other than humans.

These statistics reveal that recent experimental work, self-identified as "SL research", does appear to be undergoing some important changes. First, studies more frequently move away from established proof of concept to present research exploring the learning of regularities in a range of domains. To name a few examples these (labeled 'Other' in Figure 7) include, social learning in infancy (Crivello, Philips, & Poulin Dubois, 2016); learning melodic structure (Rohrmeier & Widdess, 2017); pitch (Daikoku, Yatomi, & Yumoto, 2016); face discrimination (e.g., Altvater-Mackensen, Jessen, & Grossmann, 2016; Dotsch, Hassin, & Todorov, 2016); action prediction (e.g., Monroy, Gerson, & Hunnius, 2017; Schuwerk, Sodian, & Paulus, 2017); orthographic regularities (Chetail, 2017; Hex & Tong (2016); and natural images (e.g., Denison, Sheynin, & Silver, 2017).

Our label of "Other" SL methods spans a range of experimental approaches. Here, again, we offer a few characteristic examples. To identify the neural correlates of audiovisual SL in musicians and non-musicians, Paraskevopoulos et al. (2018) used complex streams that involved statistical regularities in four dimensions, colors, shape, pitch, and timbre, in an oddball paradigm. Sloan and Johnson (2018) tracked eye-movements of infants in a spatial array of shapes appearing on a screen in 5 locations, examining the learning of illusory and embedded visual sequences to assess whether they are represented as chunks or not. Giorgio et al. (2018) employed brain imaging to track the functional brain networks implicated in SL by presenting probabilistic sequences using Markov chains of stimuli. Yu and Zhao (2018) investigated how SL shapes object representation by exposing subjects to objects in structured and non-structured streams, showing how pairing objects in space, impacts the judged distance between them.

These few examples suggest important methodological advances. We first note the increased introduction of neurobiological measures for tracking learning. Aside of having the promise of advancing towards a mechanistic explanation of SL, such measures often provide information regarding how learning actually unfolds (see for example, Farthouat et al., 2016). Following this goal, behavioral online measures that provide information regarding the time-course of learning have also been introduced (e.g., Siegelman et al., 2017). This is a significant step forward as it holds the promise of assessing learning more reliably (see Siegelman et al., 2017, for discussion).

However, in spite of this expansion our analysis above shows that, to some extent, SL research is still often focused on probabilistic predictions, often tapping full regularities where TPs are 1.0. Similarly, patterns are uniform in size (mostly pairs or triplets), limited in number (mostly 4-8), with brief familiarization streams (typically in the range of minutes). Importantly, the unitarian view of SL still dominates research methods, where SL is often tacitly assumed to be akin to a domain-general “black box”. A given task is thus typically chosen as a proxy for this unified ability, without much discussion of the targeted computations. From this perspective, underspecification is still the rule rather than the exception in present studies. This leads us then to the final part of our discussion where we discuss possible future directions for SL research.

4. Part 3: Towards a more pluralistic approach to SL

Our aim in delineating the current limitations of SL research was to outline possible principles for a novel framework for SL research that would enhance its overall impact in cognitive science. In this part of our discussion, we offer possible directions for future research. To be clear, our suggestions for future directions should not be taken to undermine the important contributions of SL research so far. Rather, we argue that to achieve its initial promise, SL research should now adopt different working assumptions, set novel goals, ask different questions, and consider different methodologies. In what follows, we describe the basic tenets underlying our approach, and outline our proposal for a novel research agenda for SL.

4.1 A realistic view of the learning environment

Two decades of research have produced an irrefutable proof of concept: humans and non-humans are able to perceive and learn the range of spatial or sequential regularities that experimenters typically embed in the sensory input. However, from the stand-point of external validity, the question at hand is whether the experimental environment resembles the ecological environment in which SL is hypothesized to occur. Here to re-emphasize: we take it as self-evident that experimental designs are inevitably constrained and necessarily constructed to focus on a limited set of independent variables to avoid experimental confounds. Thus, we do not take issue with the inherent procedures imposed by the rigorous nature of scientific investigations; nor do we take issue with the limitation of scope of any given SL study. Our claim is that the ecological environment in which learning typically occurs is still distant from current SL research to an extent that our understanding of how organisms learn the full range of regularities of their environment is, at best, very partial, and at worst, inaccurate. Importantly, the typical experimental designs of SL research have constrained the range of questions that have been asked about learning regularities, with some critical issues being missed as a result.

The first fact to consider is that organisms are bombarded by a virtually infinite range of regularities in the environment (see Saffran & Kirkham, 2018, for a similar argument within the domain of language). Whereas in a given experiment, participants are passively presented with a relatively simple input containing, in most cases, one type of regularity (e.g., K "words" with TPs of p within elements, and q between words), human babies, zebra finches, or cotton-top tamarins are continuously exposed to a myriad of regularities in all sensory modalities but learn only a subset of these. What mechanisms lead species to focus on a given range and type of regularities, disregarding or being insensitive to others?[8] Can multiple regularities be learned at the same time? If so, what are the constraints regarding the capacity of simultaneous assimilating multiple regularities? If not, is there a priority for learning one type of regularity over another? What determines these priorities? These are but a handful of questions to which we, to date, have no clear answers. Yet, they are fundamental for our understanding of SL, once a realistic ecological view of the learning environment is adopted.

The challenge of multiple regularities

[8] Here we focus on regularities that, in principle, could be perceived by an organism given its neurobiological endowment, yet, they are not.

A main theoretical shift, in the present context, is to consider the learner as "active" in the sense of him/her focusing on a specific range of regularities, and allocating priorities regarding what will be assimilated at a given time and what will not. In contrast to the typical lab setting where the regularities are selected for participants (whether humans or nonhumans), organisms are faced continuously with a multitude of visual and auditory regularities but do not learn them all. How is all this orchestrated? For a given species, what cues determine which specific streams of regularities should be attended to and learned, and which should be ignored?

Some initial insights regarding possible mechanisms can be gained from two recent studies. In the first, Ferguson and Lew-Williams (2016) investigated children's ability to learn patterns such as ABB (generalizing *le-di-di* to *ko-ga-ga*, see Marcus et al., 1999). Marcus, Fernandes, and Johnson (2007) have shown that when children hear speech sounds, they learn the patterns, but when they hear non-speech sounds such as sine-wave tones, they do not. This seems like an innate mechanism of selection, and indeed this finding was originally taken to suggest that speech is "special" given the unique human capacity for language (see for example, Liberman & Mattingly, 1985). Children, it was concluded, are hardwired to attend to speech. However, Ferguson and Lew-Williams (2016) demonstrated that if children are previously exposed to a video of two persons communicating in tones (a communicative context), learning does occur for tones, just as it occurs for speech sounds. Hence, it is not the speech signal that matters but the information regarding the communicative value of the signal.

Two important conclusions can be drawn from this example. First, that for children, and presumably for any organism, there are some preferences regarding what regularities should be attend to, and what regularities should be initially ignored. A system that tracks all possible regularities in the environment will simply not work, because memory is limited.

Second, one important source of priority is informativeness—what in the environment carries important information for a given species. Communication is critical for many if not for most species, be they humans or zebra finches. Consequently, learning regularities that subserve communication within species will be a primary filter for selection among the infinite regularities presented in the environment. However more generally, if the informativeness is indeed an important constraint for shaping organisms' sensitivity to specific patterning in the environment, then a viable theory of SL should first focus on mapping what types of patterning carry what information for a given species. This will enable researchers to draw clear, testable predictions regarding what would be learned easily, and what would not.

The second study concerns preferences to attend to regularities at a specific range of complexity. In a recent study, Kidd, Piantadosi, & Aslin (2012) demonstrated that infants prefer to attend to events that are neither highly unpredictable nor highly predictable. This "Goldilocks effect", was explained by Kidd and her colleagues as a characteristic of immature members of any species, that must be highly selective in sampling information from their environment in order to learn efficiently. Kidd et al. (2012) were clearly targeting a new and overlooked aspect of SL in arguing that children must avoid learning from events that are too simple or too complex. However, this constraint is not restricted to immature members of species, it is a prerequisite of efficient learning even when organisms mature. Learning abilities do indeed increase with development, but they are always limited. Selection of relevant regularities should, consequently, always be a primary mechanism for shaping SL at any age.

The findings of Kidd et al. (2012) resonate also with what we know about the neurobiology of tracking uncertainty. Recent neuroimaging studies have identified brain systems that track uncertainty in a curvilinear U-shaped function, in both the visual and auditory cortices (Nastase, Iacovella, & Hasson, 2014; and see Hasson, 2017, for a review).

Thus, for these systems, full randomness or full regularity are alike in terms of informativeness (or lack thereof), and they are tuned to the *quasi-regularities* in the environment, whether visual or auditory. Note that current SL research typically puts the demarcation line between random and non-random streams, implicitly assuming that everything that is not random is, in principle, the target of learning. However, the two lines of research we have reviewed here suggest that this assumption may be an oversimplification.

The challenge of complex streams.

Independent of the issue of learning priorities discussed above, the cascade of regularities in the environment is almost never as simple as in laboratory experiments. Thus, even if the sensory streams would carry the right amount of informativeness, and have adequate levels of uncertainty for tracking the statistical information, typically they would be far more complex than those employed in current laboratory settings. Visual and auditory inputs often contain multiple regularities. A typical example would be infants in bilingual environments, where speakers may switch from a first language to a second language without any cue that a switch was about to occur. How are regularities in these environments processed?

The current evidence regarding how learners deal with multiple regularities is relatively meager and mixed. For example, Gebhart et al. (2009) reported a primacy effect where, in the absence of a contextual cue, the first set of structural regularities in two sequentially presented streams was learned, whereas the second set was not. In contrast, tracking learning online, Siegelman et al. (2017) showed that the second set of regularities is learned as well (and see Bulgarelli & Weiss, 2016, for similar conclusions). Weiss, Gerfen and Mitchel (2009) demonstrated that when two streams overlap in their statistics, a contextual cue (e.g., change of voice) is required to learn both. In the same vein, Mitchel and Weiss (2010) showed that learning of two speech streams is facilitated when it is accompanied by coherent visual information. Using an AGL task, Conway and Christiansen (2006) found that the simultaneous learning of two sets of different regularities is possible when the sequences from two separate grammars are presented in different modalities. Moreover, Vuong, Meyer and Christiansen (2016) demonstrated that it is possible to learn both adjacent and nonadjacent dependencies using an experimental paradigm in which an AGL is embedded within an SRT task. This set of findings raises intriguing questions regarding how the learning of complex streams is orchestrated in ecological settings. Cleary, evidence regarding this critical issue is scarce.

To complicate matters further, words in continuous speech vary in length, and vary in the distribution of TPs within and between elements. This complicates learning significantly. Indeed, Hoch, Tyler and Tillmann (2013) demonstrated that inserting units of different length into the auditory stream, hinders learning (though partial learning is possible, Trecca et al., 2019). In the same vein, as reported above, Lew-Williams and Saffran (2012) showed that previous exposure to disyllabic words hinders infant performance in streams containing trisyllabic words, and vice versa. If SL is to play a major role in explaining language acquisition, a comprehensive theory of SL should specify the relevant bootstrapping mechanisms and the range of cues that are utilized for processing complex streams, as well as how they interact.

The substantial impact of the findings by Saffran et al. (1996) was in demonstrating that word boundaries can be perceived given differences in TPs within and between words, so that relatively simple learning mechanisms can potentially account for language learning. Yet, the linguistic environment is exceedingly complex, morphological structure often concatenates an array of phonological units that differ in size and structure, differences in frequencies of co-occurrence are typically very subtle, regularities are intermixed with irregularities, and correlations are remarkably small. Whether the mechanisms revealed in common SL experimental settings can be taken as a proxy for how language acquisition proceeds, is still an outstanding question (though some promise may be found in providing multiple cues to the relevant structure; Van den Bos, Christiansen & Misyak, 2012).

To summarize, given the infinite complexity of the environment, an ecological theory of SL should focus on unravelling empirically the series of constraints that predict what will be learned, what will not, and why it is so. Importantly, the theory should explain how learning proceeds when the stream to be learned is complex and not uniform in terms of sizes of units and the statistics of their co-occurrence.

4.2 Adopting a more realistic view of the learner

Organisms learn the regularities of their environment continuously from birth (see James, 2010, for earlier fetal learning). Hence, learning of regularities typically involves the *updating* of existing representations to facilitate subsequent processing, rather than establishing entirely novel ones. In other words, learners typically come to the task of learning having being exposed to the distributional properties of sensory events in their environment, so that any novel learning occurs against the backdrop of prior experience with similar or related input. The learner, then, is not a tabula rasa—a blank slate—upon which SL can work. From an evolutionary perspective, efficient processing requires that novel regularities in the incoming input should be weighed against what the organism has already learnt about its environment. Simple bottom-up processing of the input would not do. Indeed, why evolve mechanisms for aggregating the distributional properties of the environment, as has been demonstrated in a wide set of studies (e.g., Clerkin et al., 2017), if this accumulated learning is not used online for improving processing? The critical questions for investigation, therefore, are: How does prior exposure to the incoming signal influence the learning process? How, in the long term, is the novel information assimilated to facilitate the organisms' future behavior in that environment?

Current SL research has very few answers to these basic questions. The main reason is that most SL studies implicitly consider learning to be a process of assimilating novel regularities. Siegelman, Bogaerts, Arciuli, and Frost (2018) label this "the tabula rasa assumption" of SL research (see Christiansen, Conway, & Curtin, 2000; Christiansen & Curtin, 1999, for an earlier version of this criticism). The "tabula rasa" assumption considers the learning outcomes of an experiment to reflect only the input structure set by the experimenter alone. In typical experiments of visual or auditory SL, the relevant factors would be, for example, the number of patterns in the stream, the TPs within and between patterns, the similarity of test items to foils in the subsequent 2AFC test phase, etc. Underlying this approach is the assumption that the patterns as well as the foils embedded in the stream were unknown to the participants at the start, so whatever is acquired (or not) during the familiarization session reflects the net efficiency of SL computations. The "tabula rasa" assumption may largely be valid in experimental designs when the learned material is very novel (e.g., abstract shapes, Turk-Browne et al., 2005; fractal stimuli, Schapiro et al., 2014), and importantly, when there is no prior knowledge regarding co-occurrences of elements in the stream. It is clearly false, however, when the learned material is not novel, such as many studies involving linguistic material. For example, humans hear speech continuously from birth, and accumulate knowledge about the distributional properties of speech sounds in their native language every day of their lives. This means that they already have expectations regarding the co-occurrence of speech sounds in their language. The critical question is, therefore, how the well-established representations regarding probabilistic co-occurrences of speech sounds in one's native language determine the outcome of subsequent learning.

There is ample evidence showing that prior linguistic exposure affects performance in ASL tasks, such as the one originally offered by Saffran et al. (1996). For example, pre-exposing participants to isolated words or part-words before the beginning of the familiarization stream has a substantial effect on ASL performance, which can either facilitate (Cunillera, Camara, Laine, & Rodriguez-Fornells, 2010; Lew-Williams, Pelucchi, & Saffran, 2011), or hinder (Perruchet, Poulin-Charronnat, Tillmann, & Peereman, 2014; Poulin-Charronnat, Perruchet, Tillmann, & Peereman, 2016) learning. Some studies have shown that phonotactic cues characteristic to a particular language can drive segmentation of the speech input (e.g., Finn & Hudson Kam, 2008; Mersad & Nazzi, 2011; Onnis, Monaghan, Richmond, & Chater, 2005). In a similar vein, native language background can also affect if and how learners might segment an artificial language (e.g., Caldwell-Harris, Lancaster, Ladd, Dediu, & Christiansen, 2015; Toro, Sebastián-Gallés, & Mattys, 2009; Trecca et al., 2019). Together, these findings suggest that differences in prior linguistic experience lead to different results in the ASL task. They also show that the entrenchment of the statistics of one's native language inevitably produces biases towards probable co-occurrences of speech elements, influencing the patterns of subsequent learning. That is, ASL performance does not simply reflect the learning of the artificial patterns in the task, as was originally assumed. Rather, performance reflects how these new patterns fit with the statistics of prior language exposure—and this holds for learners of all ages.

The impact of entrenchment in the auditory SL task was recently demonstrated by Siegelman et al. (2018) by considering the internal consistency of the "words" employed in the task. When there is no prior knowledge whatsoever, and thus no possible predictions regarding the co-occurrence of elements in the stream, then all patterns—"words"—are by definition equal in terms of the learner's prior expectations, and this results in high correlation in performance between patterns. In contrast, if items are not entirely novel, and implicate some prior knowledge, then the “words” in the stream are not equal in terms of what they impose on the learner, and consequently substantial variance between patterns inevitably emerges. Siegelman and his colleagues have shown that whereas visual SL with abstract shapes always displays high internal consistency, the classical ASL task always displays low internal consistency. Thus, learning, for example, “balogi” in a continuous input stream, does not predict learning another “word”, such as “gupati”. In this context, our finding in Part 1, that about 24% of studies with ASL used the same set of words employed by Saffran et al. (1996) carries then critical significance. If different “words” were used in the different studies, a significant variability in the experimental outcomes would have probably been the result[9].

A realistic view of the learner requires a major shift in SL research, taking into account learning that interfaces with prior knowledge and learning that does not. In the domain of language, therefore, the main focus should be on how prior linguistic exposure might affect SL task involving linguistic stimuli (see e.g., Caldwell-Harris et al., 2015; Trecca et al., 2019). For example, understanding the impact of statistical entrenchment in speech perception would require mapping the cues that could, in principle, impact speech segmentation, and then assessing the relative weight of each of these cues and their possible interactions with one another. Another possibility is to study SL “in the wild”—directly measuring sensitivity to statistics in natural language (for example, as measured via corpus analyses) to see how these might relate to language processing (see McCauley, Isbilen & Christiansen, 2017). Similarly, a theory of orthographic SL would require weighing the relative frequency of individual letters and various possible letter combinations, including bigrams, trigrams, etc., given their position within words in the language, Relatedly, it would

[9] Some SL researchers, perhaps implicitly aware of this issue, have used two different input streams to avoid potential idiosyncratic effects of a single set of stimuli. However, since the two streams were often created so that words of one stream served as foils for the other, they remain closely related sharing many of the same entrenchment effects.

also need to consider visual factors such as crowding, visual acuity, and other constraints on the visual system (see Lerner, Armstrong, & Frost, 2014; Grainger et al., 2016, for discussion), then taking into account the correlation of orthographic forms with phonological, semantic, and other patterns. Having identified the cue weightings associated with these various regularities, the different statistics could then be evaluated in targeted laboratory experiments.

4.3 Integrating SL into cognition

The important role of SL in cognitive science stems from the wide range of processes it subserves. As we have argued in Part 1, understanding the learning of regularities requires researchers to define, as a first step, the specific domain of learning, whether it is speech, visual scenes, objects, faces, grammar, or print, to name a few. Each of these domains is characterized by different types of regularities, and different types of computations. Common to them all is only a very abstract and vacuous notion of learning "patterns", per our initial definition of SL. While it is possible that there may be something common to all pattern learning, very little can be said about it, mainly that the learning focuses on patterns in the environment. The increased focus of SL research on a specific type of patterning, that of co-occurrence of elements in a stream (either through TP statistics or through AGL), has led, to some extent, to the conception of SL as a cohesive and independent domain of research in its own right, concerned with mapping the constraints of such learning. Here we would like to argue that, in the long term, SL research should be incorporated into the different research programs of each of the above domains. We label this form of incorporation *Domain integration*, per the logic of Figure 4. It contrasts with *Timescale integration* which requires SL research to converge with what we know about the general faculties that subserve cognition such as memory, attention, or executive functions. In the following, we discuss both.

Domain integration

To outline what domain integration could involve, let us consider two major faculties that involve statistical regularities—literacy acquisition and face perception. To assess the range of regularities that are assimilated during literacy acquisition, we should consider the bulk of established effects reported in the domain of visual word recognition and text reading. For example, proficient readers name words with regular spelling-to-sound correspondence (i.e. *punt*) faster than words with irregular spelling-to-sound correspondence (i.e. *pint*, e.g., Cortese & Simpson, 2000). They are faster to name words with a consistent body such as *mint* than words with an inconsistent body like *pint* **(**e.g.**,** Jared, McRae, & Seidenberg, 1990). They automatically decompose morphologically complex words like *farmer* into stem *farm,* and suffix *er,* but also pseudo-complex words like *corner* into *corn+er* (e.g., Rastle, Davis, & New, 2004). They are sensitive to the sequential co-occurrence of root letters within different words that correlate in meaning in Semitic languages (e.g., Frost, Forster, & Deutsch, 1997). They learn that there is a high probability that the semantic radical of Chinese words will be on the left side, whereas the phonetic radical will be in the right side (e.g., Lee, 2011). They know that in French, *n* could probably follow *r* but not follow *c*, thereby affecting perceptual processing of bigrams in rapid serial visual presentation (e.g., Chetail, 2017). They learn that words in English that end with the sound /es/ and are printed *us* are most probably nouns rather than adjectives (Ulicheva, Ahronoff, & Rastle, 2018).

These findings highlight the types of regularities that are the object of learning in this domain, driving the range of behavioral and neurobiological phenomena involved in lexical decision, naming, priming, semantic judgements, or eye movements. These regularities concern correlation between letters or letter sequences and sound or sound sequences of the language, correlations between short letter sequences such as suffixes and prefixes and semantic meaning, regularities regarding spatial location of graphemes and lexical status, regularities regarding the probabilistic co-occurrences of letters within isolated words and probabilistic co-occurrences of words in a sentence, the regularity between spelling patterns and syntactic class, and this is not an exhaustive list. Proficient literacy is, thus, a form of SL expertise, related to assimilating a range of statistical regularities that reflect the dimensions of language—orthography, phonology, morphology, and meaning. What aspects of SL take part in the acquisition of this skill? How exactly do domain-general SL contribute to establishing orthographic representations and lexical organization? These types of questions, to which we presently have too few answers given that SL and reading research proceed in parallel lines, are the basis for our suggestion for domain integration of SL into literacy acquisition research (see for example, Arciuli, 2018, for discussion).

Consider now face perception, another human ability that involves expertise. Within just 100 ms of exposure, people can form inferences regarding the trustworthiness or aggressiveness of unfamiliar faces (Willis & Todorov, 2006). These inferences emerge from perceived emotion, facial maturity, or perceived gender that, in turn, are correlated with a range of consistent cues such as the distance between eyes and eyebrows, the size of eyes, or the ratio of width to height of the face, etc. (e.g., Oosterhof & Todorov, 2008). We know that humans are better at memorizing faces of their own race than other races (the other-race-effect, e.g., Tanaka, Kiefer, & Bukach, 2004), suggesting that experience and learning determine performance. Recently, Dotsch, Hassin and Todorov (2016) have shown how SL shapes face evaluation. By generating a statistical distribution of facial features through sampling of a large number of real faces, Dotsch et al. demonstrated that the location of a face on the statistical distribution determines its evaluative inference, the more distant it is from the mean tendency of the distribution, the more negative the inference. Finally, in a recent study, Zwebner, Sellier, Rosenfeld, Goldenberg and Mayo (2017) have shown that participants examining an unfamiliar face within their own culture, are above chance in selecting the person's true name from a list of several names. Zwebner et al. argued that social expectations how a person with a specific name should look like (e.g., hairstyle, etc.), eventually influence his/her facial appearance, resulting in some regularity. The manner by which such statistical information is perceived by participants probably underlies the effect.

This brief review leads to the conclusion that, similar to word perception, learning of regularities underlies important aspects of face perception. However, the precise nature of the regularities that drive the above list of effects remains to a large extent obscure. What is the object of learning that leads to face perception expertise? What are the culturally-bound internal representations that develop with experience, and underlie emotion inference? What aspects of SL are implicated in this form of regularity learning? Only by integrating work on SL within face perception research can significant advances be achieved.

Timescale integration

We now turn to integrating SL theory and research with the general abilities that subserve cognition, focusing, as an example, on attention and memory.

SL and attention

Our analysis of past and present reveals that only few papers have directly target SL and attention (but see recent work by Wang & Theeuwes, 2018a,b,c), in the timescale where attention determine the learning process. Admittedly, the role of attention in modulating cognitive capacities is under-researched across cognitive psychology in general. However, the important role of selective attention as a theoretical construct in understanding learning should lead to greater integration of attention research with SL. Indeed, the implicit learning literature has investigated the impact of attention on learning for quite some time (for reviews, see e.g., Perruchet & Vinter, 2002; Shanks & St. John, 1994). However, the extensive work on explicit vs. implicit learning has not been integrated into SL research, reflecting another aspect of lack of integration (see for example Perruchet & Pacton, 2006, for a discussion). Possibly, once it was established that SL can be incidental, not requiring overt attention (e.g., Saffran et al., 1997), integrating theories of attention with theories of SL was not a major concern. However, the extent to which selective attention determines SL performance is an unresolved issue. For example, in contrast to Saffran et al. (1997), Toro, Sinnett and Soto-Faraco (2005) found that speech segmentation in the ASL task is compromised without attention allocation. In visual SL, Turk-Browne et al. (2005) offered a nuanced discussion of the role of attention, suggesting that attention is required for selecting the relevant stimulus properties, while the learning of regularities occurs without intent (and see Baker, Olson, & Behrmann, 2004, for the impact of attention on perceptual grouping in visual SL). However, using a similar experimental approach, Musz, Weber and Thompson-Schill (2015) have recently observed that visual SL is not reliably modulated by attention.

Attention, however, is a highly complex theoretical construct, and simply splitting cognitive processes into those "requiring overt attention" and those that do not, may miss important aspects of it. Indeed, attention is consistently discussed in the literature on contextual cueing (CC). In the CC paradigm (e.g., Chun & Jiang, 1998), participants search for a letter target (T) within a spatial configuration of many distractors (L), when half of the configurations are repeated and half are novel. The CC effect is defined as the faster detection of the letter T in the repeated configurations vs. the novel ones. CC, therefore, is a clear SL phenomenon (see Goujon, Didierjean & Thorpe, 2015, for discussion), although few researchers explicitly label it as such according to our literature search. It revolves around the implicit learning of spatial contingencies (see Goujon et al., 2015, for discussion), quite similar to the seminal study of Fiser and Aslin (2001) with spatial grids. The typical account for the CC effect is that learning the regularities regarding the location of the letter T in the grid, results in deployment of visual attention towards the right location, leading to faster search time. In that sense, CC is a pure attentional effect (e.g., Chun, 2000), demonstrating how visual SL learning leads to patterns of attention deployment, even though learning is incidental—participants are at chance in recognizing the repeated configurations. Hence, incidental learning does not necessarily mean lack of attention deployment.

The interplay of SL and attention has been recently demonstrated by Wang and Theeuwes (2018a,b,c). These studies employed the singleton task, where participants search for a salient shape (e.g., a green diamond surrounded by green circles), and are required to ignore a distractor that stands out (e.g., a red circle). The typical finding is that the time to locate the target increases if a distractor is present in the display, in spite of explicit instructions to ignore it, because attention is captured automatically (e.g., Theeuwes, 1992). The singleton paradigm thus monitors to what extent attentional selection can or cannot be controlled. In their series of studies, Wang and Theeuwes have shown that if the color distractor is presented in one location at a high probability, its hindering impact in term of capturing attention decreases, and this is independent of participants' awareness of the statistical regularities. Thus, the extent of involuntary capture of attention is modulated by the learning of the statistical information regarding distractor location, whether learning is incidental or not. They further explored how SL interacts with intentional top-down suppression, suggesting an intricate interaction between SL computations and attention allocation. In the same vein, Tummeltshammer, Mareschal and Kirkham (2014) have shown that the ability to shift attention away from a distractor stimulus to learn a cue–reward regularity, changes over the course of development. Similarly, Zhao, Al-Aidros, & Turk-

Browne (2013) reported that, in general, regularities bias spatial attention so that visual search is facilitated at locations that involve temporal regularities, irrespective whether these regularities predicted target location. Further, Zhao and Luo (2017) showed that statistical regularities in local vs. global scale prioritize local vs. global processing. Taken together, these studies demonstrate an attentional priority to statistically structured sources of information.

Discussions of attention in the domain of SL often intermix attention (or the lack thereof) with "intent", "automaticity", "awareness", or "explicitness". Here the problem of underspecification becomes perhaps more acute. Integrating SL with theories of attention requires a well-specified definition of what aspect of attention is the target of research, and what attentional mechanism(s) undergo experimental scrutiny. At present, we know that SL can occur largely automatically, without intent, without conscious awareness, and that it is often implicit and incidental. Nevertheless, we do not know exactly how mechanisms of attention determine what regularities will be attended to and what will not, how they modulate learning outcomes, how they change over the course of development to impact SL behavior, and how they interact with memory systems to determine whether learning is long lasting or not. This should be a primary concern for future SL research.

SL and memory

For SL to underlie basic functions of cognition, such as language, visual perception, or semantic categorization, the continuous perception of regularities in the environment has to be assimilated into stable long-term representations. However, when the time course of these processes is considered, an apparent paradox emerges. Whereas learning the co-occurrences of elements in an input stream is exceedingly fast and effortless (neonates right after birth already display sensitivity to frequency of co-occurrences of syllables in an auditory stream presented for 15 minutes during sleep, see Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), assimilating the regularities in the environment seems to be exceedingly slow. Consider the case of language. In general, learning a first (or second) language is a slow and effortful process. Thus, there is a striking contrast between how early and fast SL is compared to how slow the process of learning language regularities is. Aside from the very simplified, extremely limited nature of SL stimuli compared to the noisy input of real-world language (as discussed in our section on the realistic view of the learning environment), several factors affect the relative slowness of language learning. One relates to how new experience is affected by and subsequently impacts on existing patterned regularities learned from prior exposure to language (for related discussion, see Armstrong, Dumay, Kim, & Pitt, 2017). However, another potential factor is the interaction between experience and the gradual maturation of different neural systems. Gomez (2017) argued that different memory-related neural systems with different encoding and retention capacities emerge over the course of development, and this determines what will be retained, and at what speed. In the same vein, Santolin and Saffran (2017) provide an extensive review of cross-species learning abilities, highlighting differences in their memory systems. In a nutshell, since the perceived distributional properties of the input have to be assimilated long term to impact behavior, understanding the learning of regularities in the environment requires a focus on how mechanisms of SL interact with the different neural systems of a given species at a given developmental phase to produce stable learning.[10]

[10] We note here that our discussion of memory-related neural systems does not imply that we endorse a strict separation of memory and processing. Indeed, recent neuroimaging results suggests that memory is not separated from but rather intrinsic to processing (see Hasson, Chen & Honey, 2015, for a review). Similarly, it has been proposed that learning simply involves becoming better at processing in both SL (Christiansen, in press) and natural language (Christiansen & Chater, 2016), pointing to an integrated account of learning, memory, and processing.

The above discussion implies that studies that focused on existence proofs of SL abilities with different populations showed at best that, in principle, these populations display sensitivity to regularities, not that they can and do assimilate and retain the perceived statistical information. Indeed, direct evidence about retention of SL is scarce (though see e.g., Kim, Seitz, Feenstra, & Shams, 2009). To have ecological validity, SL research should therefore consider the maturation trajectory of the different neural systems (e.g., the CA1 and CA3 hippocampal regions, the cortico-striatal networks, the neocortex, etc.), the maturation trajectory of their interconnecting pathways, what we know about processes of consolidation (see Gomez, 2017, for a review), how but mostly, how experience interacts with maturation over various timescales.

4.4 Opening the door to novel approaches to SL

As our summary of critical features of the Past SL research reveals, over 51% of studies that presented regularities in a familiarization stream measured SL performance via a 2AFC test following familiarization. Present research reveals a similar ratio. Thus, the majority of SL research and theory hinges on tapping the number of correct responses to a relatively short series of test questions regarding the structure of the input stream relative to chance. We have detailed the psychometric and methodological shortcomings of this offline measure elsewhere (Christiansen, 2019; Isbilen et al., 2017; Siegelman et al., 2016, 2017), and we will not reiterate them here. However, aside from these inherent methodological limitations, a fundamental shortcoming of the 2AFC measure of SL is that it asks participants to reflect on what they have learned, rather than tapping more directly into the underlying system doing the learning with a processing-based measure (for discussion, see Christiansen, 2019). This makes the current dominant measure of SL deeply impoverished: It does not provide any information regarding the time-course of learning or its trajectory (e.g., how fast is learning, is it incremental or abrupt?); it does not capture learning and its relation to broader cognitive abilities such as memory or attention; it does not address the neural underpinnings of SL. Adopting processing-based measures is critical for reaching a mechanistic understanding of how SL proceeds, stabilizes, and is integrated with prior learning across cognitive systems.

There are several lines of emerging research that can help address the problem of underspecification in SL research using processing-based measures. Some of this work relates to understanding the time course of learning. Whereas the SRT task has been used for some time to investigate the time course of learning fixed sequences (e.g., Nissen & Bullemer, 1987), it has more recently been used to study SL by incorporating AGL into this task (e.g., Misyak, Christiansen, & Tomblin, 2010). Closer to the classic paradigm monitoring the learning of triplet in a continuous stream, Siegelman et al. (2017) demonstrated that tracking the extent of speeded RTs to predictable stimuli throughout the experimental session holds the promise of revealing novel information regarding when learning occurs and how it proceeds. This simple behavioral online measure comes almost for free by simply asking participants to advance through the sequence in a self-paced manner rather than watching the shape sequence in a passive manner. However, like most RT measures it is inherently noisy, and therefore will benefit from being supplemented by other convergent measures of learning.

Another line of research seeks to link SL more closely to the cognitive mechanisms that it subserves, such as memory. The rationale is that sensitivity to statistical patterns should improve memory recall through the familiar process of chunking: coherent statistical patterns should be easier to chunk and thus result in improvement of memory performance. Intriguingly, very early work on AGL explicitly used a classic memory task—serial recall—to demonstrate effects of learning (Miller, 1958; Reber, 1967; see Christiansen, 2019, for a historical overview). More recently, serial recall has been used to study both VSL (Karpicke

& Pisoni, 2004) and ASL (Conway et al., 2010), demonstrating how sensitivity to distributional regularities facilitates short-term memory performance. Relatedly, Isbilen et al. (2017) used the recall task to obtain a reliable measure of TP learning, capturing sensitivity to patterned regularities hypothesized to be relevant for language learning. Importantly, such statistically-induced facilitation of recall should be observable not only in the context of experiments with artificial language stimuli, but also in studies involving real-world natural language statistics as demonstrated by McCauley et al. (2017).

Interestingly, as an example of lack of integration, in parallel to SL research there is currently extensive work on the neurobiological basis of prediction focusing on neural oscillations. Neuronal oscillations reflect rhythmic fluctuations in the inhibition/excitation balance of neuronal populations (e.g., Buzsáki & Wang, 2012; Haider & McCormick, 2009) and have been proposed to be instrumental to account for memory formation and attentional selection of inputs. They provide, therefore, an efficient mechanism to amplify the neuronal responses to behaviorally relevant events (e.g., Schroeder, et al., 2010). These mechanisms have been shown to support the detection of predictable events given statistical regularities. Evidence indicates that such rhythmic processing may be achieved by a phase entrainment of oscillatory activity at delta (1-5 Hz) and beta (15-30Hz) frequencies, which follows the temporal structure of the continuous stream of individual elements and patterns embedded therein (e.g., Lakatos et al., 2008). Learning regularities implies that events in any input stream would differ in terms of their predictability. This should be reflected in specific oscillatory activity at a given time point. Whereas Lakatos et al. (2008) were concerned with simple temporal expectations, beta-range oscillations have been associated with expectation in a SL paradigm (Pearce et al., 2010). Another approach is to track patterns of synchronization of EEG activity. It has been shown that in an auditory stream of syllables where "words" are embedded, EEG activity synchronizes first with syllable presentation, but then synchronizes with "word" rate (e.g., Batterink & Paller, 2017; Buiatti, Peña, & Dehaene-Lambertz, 2009). Pinpointing the time by which synchronization diverges, thus provides evidence regarding the time course of learning. Hence, using such range of neural measures has the promise of advancing SL research significantly towards a better understanding of its underlying mechanisms.

This discussion leads us to the potential merit of tracking individual performance in SL. Reaching a more precise mechanistic theory of SL, and mapping its componential facets will benefit from a move from aggregate measures of learning at the group-level to investigating differences in individual performance (see Frost et al., 2015, for discussion, and Kidd, Donnelly & Christiansen, 2018, for similar arguments about language). This line of research holds the promise of teasing apart different aspects of SL, examining their relation with one another, as well as their relations with specific cognitive abilities. Indeed, substantial evidence against the unitarian view of SL has been provided by studies that focused on individual performance (e.g., Misyak & Christiansen, 2012; Siegelman & Frost, 2015). Although recent years have seen a growing interest in such research, little is known to date about the precise componential structure of SL, what its independent facets are, and to what extent these facets predict specific abilities (see Siegelman et al., 2017, for discussion). Precise investigation of individual performance, however, requires shaping novel methodologies that are sensitive enough to track how learning proceeds within single participants.

Finally, computational models of SL can serve as a major research tool in investigating the process of updating representations when prior knowledge dominates learning of novel regularities, and when learning involves complex streams of information. As we have argued in our discussion of the impact of the unitarian view of SL, this requires a major change of focus in computational approaches to SL. Rather than providing proof of concept that SL can proceed through one or two types of computations, modelling work should be harnessed to provide sources of constraints regarding how learning of regularities proceeds, and how fundamental learning, representation, and processing principles interact with the statistical properties of a sensory input, to capture, explain, and predict a wide range of empirical phenomena (see e.g., McCauley & Christiansen, 2019, for statistically-based computational model that captures early language acquisition across multiple languages).

5 Summary and conclusions

This article has focused on the important accomplishments of self-identified SL research, but also on an apparent gap between the promise of SL as a theoretical construct, and the actual advances that this field of research has achieved so far. The working hypothesis of SL research has been that it is applicable to all functions related to distributional analyses of environmental input, and would thus provide adequate descriptive and explanatory foundations for a wide range of cognitive abilities. However, research on SL has been hampered by some critical limitations, preventing it from achieving its original promise: the imbalance between the breadth of theoretical claims and the actual empirical evidence supporting them; taking SL to be a unitary central device, overlooking evidence concerning its componential aspects; studying SL in isolation from the cognitive systems it subserves and interacts with, while focusing on very narrow timescales; often being too vague and imprecise regarding actual representations, processing mechanisms, and learning outcomes; taking the dominant experimental paradigm to be the explanatory mechanism, and explaining the mechanism by describing behavior in the experimental paradigm; ignoring the complexity of learning situations in the environment, focusing on relatively impoverished learning conditions which lack ecological validity; considering the learner as an apathetic passive absorber of regularities, missing their active role in shaping the learning parameters.

Taken together, these limitations have led to the situation where SL research is often engaged in manipulating a narrow set of parameters within a too small set of experimental procedures.

These limitations led us to conclude that a change of focus is required for future research, so that SL would achieve its original promise. The first step is to consider and adopt a more realistic and ecologically valid view of the learning environment, and of the organisms that are continuously learning from it. This, however, requires asking a novel set of research questions. A valid theory of SL has to provide adequate answers for how and why organisms focus on a specific subset of regularities from an infinite range of patterns in the environment, how they perceive and assimilate multiple regularities embedded in sensory input, how they learn patterns that are not uniform in size and vary in their distributional properties, and how they overcome the substantial noise characteristics of sensory inputs. A realistic view of the learner requires stressing that learning seldom involves assimilating completely novel representations. Rather, the learning of regularities is a continuous process where prior knowledge affects the learner's expectations, determining the learning outcome to a large extent. This suggests that the multiple existence proofs of SL accumulated in the last two decades have centered to a large extent on situations that are quite distant from the ecological learning environment we typically face. Going forward, a valid theory of SL should refocus to not only consider simple existence proof experiments but also to provide an adequate account of how learning accumulates and stabilizes into long-term representations, given what we know about the developmental trajectories of other cognitive skills, such as memory and attention.

This new research agenda requires a departure from current experimental paradigms, adopting novel methodologies and approaches. As a first step, measures of learning should be processing-based rather than reflection-based as in much of the past SL work. They should be refined and expanded to consider learning trajectories, learning stability, integrated with cognitive systems, and allow for the merging together of reliable behavioral and neurobiological signatures of learning, and drawing upon constraints provided by computational learning models. This, in fact, reflects the typical advances in all domains that focus on assimilating expertise, in vision or audition, whether in processing orthographic information, understanding speech, recognizing objects, or analyzing visual scenes. Hence, it is possible that in time, SL may eventually outlive its purpose as an independent field of research, and instead have become integrated into the study of these different domains.

Acknowledgment:

This paper was supported by the European Research Council (ERC) Advanced grant (project 692502-L2STAT) under the Horizon2020 research and innovation program, and by the Israel Science Foundation (Grant 217/14), awarded to RF. BCA was supported by a Natural Sciences and Engineering (NSERC) Discovery Grant (2017-06310). MHC was supported in part by the Danish Council for Independent Research (FKK-grant DFF-7013-00074). We thank Noam Siegelman, Louisa Bogaerts, and Lizz Karuza for their comments and very helpful discussions. We are grateful for the assistance of librarians Angela Hamilton and Sarah Guay in conducting our literature search.

References

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 899–912. https://doi.org/10.1037/0278-7393.21.4.899

Altvater-Mackensen N., Jessen S., & Grossmann T. (2017). Brain responses reveal that infants' face discrimination is guided by statistical learning from distributional information. *Developmental Science, 20*, e12393, 10.1111/desc.12393

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science, 14*, 464–473. https://doi.org/10.1111/j.1467-7687.2009.00937.x

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*, 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Arciuli, J., von Koss Torkildsen, J., Stevens, D. J., & Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00747

Arciuli, J. (2018). Reading as statistical learning. *Language, Speech, and Hearing Services in Schools, 49*, 634–643. https://pubs.asha.org/ss/rights_and_permissions.aspx

Armstrong, B. C., Dumay, N., Kim, W., & Pitt, M. A. (2017). Generalization from newly learned words reveals structural properties of the human reading system. *Journal of Experimental Psychology: General, 146(2)*, 227-249. http://dx.doi.org/10.1037/xge0000257

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9(4)*, 321–324. https://doi.org/10.1111/1467-9280.00063

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science, 15(7)*, 460–466. https://doi.org/10.1111/j.0956-7976.2004.00702.x

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex, 90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language, 83*, 62–78.

https://doi.org/10.1016/j.jml.2015.04.004

Berent, I., Everett, D. L., & Shimron, J. (2001). Do phonological representations specify variables? Evidence from the obligatory contour principle. *Cognitive Psychology, 42(1)*, 1–60. https://doi.org/10.1006/cogp.2000.0742

Bogaerts, L., Siegelman, N., & Frost, R. (2016). Splitting the variance of statistical learning performance: A parametric investigation of exposure duration and transitional probabilities. *Psychonomic Bulletin & Review, 23*, 1250–1256. https://doi.org/10.3758/s13423-015-0996-z

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138(4)*, 487-502.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105(38)*, 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes. *Psychological Science, 19(7)*, 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage, 44(2)*, 509–519. https://doi.org/10.1016/j.neuroimage.2008.09.015

Bulgarelli, F., & Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(10)*, 1621–1631. https://doi.org/10.1037/xlm0000263

Buzsáki, G., & Wang, X.-J. (2012). Mechanisms of gamma oscillations. *Annual Review of Neuroscience, 35*, 203–225. https://doi.org/10.1146/annurev-neuro-062111-150444

Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., & Christiansen, M. H. (2015). Factors influencing sensitivity to lexical tone in an artificial language. *Studies in Second Language Acquisition, 37(2)*, 335–357. https://doi.org/10.1017/S0272263114000849

Caramazza, A., & Sheldon, J.R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10*, 1-34.

Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition, 155*, 96–107.

https://doi.org/10.1016/j.cognition.2016.05.024

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition, 163*, 103–120. https://doi.org/10.1016/j.cognition.2017.02.015

Chetail, F. (2017). Statistical learning and visual word processing: Development and impact. Meeting of COOL: Conscious and unconscious aspects of learning - Satellite workshop of ESCoP 2017. Potsdam, September.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press.

Christiansen, M.H. (2019). Implicit-statistical learning: A tale of two literatures*. Topics in Cognitive Science, 11*, 468-481. https://doi.org/10.1111/tops.12332

Christiansen, M.H. & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences, 39*, e62. https://doi.org/10.1017/S0140525X1500031X

Christiansen, M. H., Conway, C. M., & Curtin, S. (2000). A connectionist single-mechanism account of rule-like behavior in infancy. In *Proceedings of the 22nd Meeting of the Cognitive Science Society* (pp. 83–86). Cognitive Science Society.

Christiansen, M. H., Conway, C. M., & Curtin, S. (2005). Multiple-cue integration in language acquisition: A connectionist model of speech segmentation and rule-like behavior. In J.W. Minett & W.S.-Y. Wang (Eds.), *Language acquisition, change and emergence: Essays in evolutionary linguistics* (pp. 205-249). Hong Kong: City University of Hong Kong Press.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes, 27(2)*, 231–256. https://doi.org/10.1080/01690965.2011.606666

Christiansen, M. H., & Curtin, S. (1999). Transfer of learning: rule acquisition or statistical learning? *Trends in Cognitive Sciences, 3(8)*, 289–290. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10431257

Christiansen, M. H., Kelly, M.L., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition, 116*, 382–393. https://doi.org/10.1016/j.cognition.2010.05.015

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*, 170-178. https://doi.org/10.1016/S1364-6613(00)01476-5

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28–71. https://doi.org/10.1006/cogp.1998.0681

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711)*, 20160055. https://doi.org/10.1098/rstb.2016.0055

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114(3)*, 356–371. https://doi.org/10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 31(1)*, 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychological Science, 17*, 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Cortese, M. J., & Simpson, G. (2000). Regularity effects in word naming: What are they? *Memory & Cognition, 28(8)*, 1269–1276. https://doi.org/10.3758/bf03211827

Coutanche, M.N. & Thompson-Schill, S.L. (2015). Rapid consolidation of new knowledge in adulthood via fast mapping. *Trends in Cognitive Sciences, 19*, 486-488.

Covington, N.V., Brown-Shmidt, S., & Duff, M.C. (2018). The Necessity of the Hippocampus for Statistical Learning. *Journal of Cognitive Neuroscience, 30*, 680-697. doi: 10.1162/jocn_a_01228

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(5)*, 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

Crivello, P., Philips, S., & Poulin Dubois, D. (2018). Selective social learning in infancy: looking for mechanisms. *Developmental Science, 21*, e12592, 10.1111/desc.12592

Cunillera, T., Cimara, E., Laine, M., & Rodiguez-Fornells, A. (2010). Words as anchors: Known words facilitate statistical learning. *Experimental Psychology, 57(2)*, 134–141. https://doi.org/10.1027/1618-3169/a000017

Daikoku T., Yatomi Y., Yumoto M. (2017). Statistical learning of an auditory sequence and reorganization of acquired knowledge: A time course of word segmentation and ordering. *Neuropsychologia, 95*, 1-16. 10.1016/j.neuropsychologia.2016.12.006

Dale, R. & Christiansen, M.H. (2004). Active and passive statistical learning: Exploring the role of feedback in artificial grammar learning and language. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 262-267). Mahwah, NJ: Lawrence Erlbaum.

Denison R.N., Sheynin J., Silver M.A. (2016). Perceptual suppression of predicted natural images. *Journal of Vision, 16*, 1-15. 10.1167/16.13.6

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19(1)*, 158–164. https://doi.org/10.1038/nn.4186

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour, 1(1)*, 1. https://doi.org/10.1038/s41562-016-0001

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology, 64*, 1021–1040. https://doi.org/10.1080/17470218.2010.538972

Emberson, L.L., Misyak, J.B., Shwade, J.A., Christiansen, M.H., & Goldstein, M.H. (in press). Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism(s). *Developmental Science.* https://doi.org/10.1111/desc.12847

Endress, A. D., & Langus, A. (2017). Transitional probabilities count more than frequency, but might not be used for memorization. *Cognitive Psychology, 92*, 37–64. https://doi.org/10.1016/j.cogpsych.2016.11.004

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*, 351–367. https://doi.org/10.1016/j.jml.2008.10.003

Evans, J., Saffran, J., & Robe-Torres, K. (2009). Statistical learning in children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 52*, 321–335.

Evans, N., & Levinson, S. C. (2009). With diversity in mind: Freeing the language sciences from Universal Grammar. *Behavioral and Brain Sciences, 32(5)*, 472.

https://doi.org/10.1017/S0140525X09990525

Farthouat, J., Franco, A., Alison, M., Delpouve, J., Wens, V. Op de Beek, M., de Tiege, X., Peigneux, P. (2017). Auditory magnetoencephalographic frequency-tagged responses mirror the ongoing segmentation processes underlying statistical learning. *Brain Topography, 30*, 220-232. https://doi.org/10.1007/s10548-016-0518-y.

Ferguson, B., & Lew-Williams, C. (2016). Communicative signals support abstract rule learning by 7-month-old infants. *Scientific Reports, 6*, doi: 10.1038/srep25434

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108(2)*, 477–499. https://doi.org/10.1016/j.cognition.2008.04.002

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*, 499–504. https://doi.org/10.1111/1467-9280.00392

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(4)*, 680–698. https://doi.org/10.1037/0278-7393.10.4.680

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review, 118,* 614–636. http://dx .doi.org/10.1037/a0025255

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences, 35(5)*, 263–279. https://doi.org/10.1017/S0140525X11001841

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19(3)*, 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*, 829–856. https://doi.org/10.1037/h0090337

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24(7)*, 1243–52. https://doi.org/10.1177/0956797612472207

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research, 58*, 934–945. https://doi.org/10.1044/2015

Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. *Vision Research, 37(12)*, 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). changing structures in midstream: learning along the statistical garden path. *Cognitive Science, 33(6),* 1087–1116. https://doi.org/10.1111/j.1551-6709.2009.01041.x

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review, 16*, 486–490. https://doi.org/10.3758/PBR.16.3.486

Getzmann, S., & N????t??nen, R. (2015). The mismatch negativity as a measure of auditory stream segregation in a simulated “cocktail-party” scenario: Effect of age. *Neurobiology of Aging, 36(11)*, 3029–3037. https://doi.org/10.1016/j.neurobiolaging.2015.07.017

Giorgio J., Karlaftis V.M., Wang R., Shen Y., Tino P., Welchman A., & Kourtzi Z. (2018) Functional brain networks for learning predictive statistics. *Cortex, 107*, 204-219. 10.1016/j.cortex.2017.08.014

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review, 20*, 1161–1169. https://doi.org/10.3758/s13423-013-0458-4

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological science, 19(5)*, 515-523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436. https://doi.org/10.1111/1467-9280.00476

Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 372(1711)*. http://rstb.royalsocietypublishing.org/content/372/1711/20160054

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences, 19(9)*, 524–533. https://doi.org/10.1016/j.tics.2015.07.009

Grainger, J., Dufau, S., & Ziegler, J. C. (2016). A vision of reading. *Trends in Cognitive Sciences, 20*, 171-179. https://doi.org/10.1016/j.tics.2015.12.008

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*, 14-23. https://doi.org/10.1016/j.tics.2005.11.006

Haider, B., & McCormick, D. A. (2009). Rapid neocortical dynamics: Cellular and network mechanisms. *Neuron, 62*, 171–189. https://doi.org/10.1016/j.neuron.2009.04.008

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304-313. http://dx.doi.org/10.1016/j.tics.2015.04.006

Hasson, U. (2017). The neurobiology of uncertainty: implications for statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711*), 20160048. https://doi.org/10.1098/rstb.2016.0048

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78(3)*, B53–B64. https://doi.org/10.1016/S0010-0277(00)00132-3

He X., Tong S.X. (2017). Quantity matters: Children with dyslexia are impaired in a small, but not large, number of exposures during implicit repeated sequence learning. *American Journal of Speech-Language Pathology, 26*, 1080-1091. 10.1044/2017_AJSLP-15-0190

Hoch, L., Tyler, M. D., & Tillmann, B. (2013). Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. *Psychonomic Bulletin & Review, 20(1)*, 142–147. https://doi.org/10.3758/s13423-012-0309-8

Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology, 5*, 1–10. https://doi.org/10.3389/fpsyg.2014.00175

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology. General, 130*, 658–680. https://doi.org/10.1037/0096-3445.130.4.658

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2017). Testing statistical learning implicitly: A novel chunk-based measure of statistical learning. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 564–569). Austin, TX: Cognitive

Science Society.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology, 19(18)*, 1543–1548. https://doi.org/10.1016/j.cub.2009.07.051

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences, 109(19)*, 7241–7244. https://doi.org/10.1073/pnas.1200155109

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29(6)*, 687–715. https://doi.org/10.1016/0749-596X(90)90044-Z

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1–13. https://doi.org/10.1016/j.cognition.2015.07.009

Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span. *Memory & Cognition, 32(6)*, 956–964. https://doi.org/10.3758/BF03196873

Karuza, E. A., Li, P., Weiss, D. J., Bulgarelli, F., Zinszer, B. D., & Aslin, R. N. (2016). Sampling over Nonuniform Distributions: A neural efficiency account of the primacy effect in statistical learning. *Journal of Cognitive Neuroscience, 28(10)*, 1484–1500. https://doi.org/10.1162/jocn_a_00990

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127*, 46–54. https://doi.org/10.1016/j.bandl.2012.11.007

Kidd, E., Donnelly, S. & Christiansen, M.H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences, 22*, 154-169. https://doi.org/10.1016/j.tics.2017.11.006

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One, 7(5)*, e36399. https://doi.org/10.1371/journal.pone.0036399

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*, 145–149. https://doi.org/10.1016/j.neulet.2009.06.030

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy:

evidence for a domain general learning mechanism. *Cognition, 83(2)*, B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances. *Psychological Science, 3(3)*, 172–179. https://doi.org/10.1111/j.1467-9280.1992.tb00021.x

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*, 110–113. https://doi.org/10.1126/science.1154735

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 60*, 3474-3486. https://doi.org/10.1044/2017_JSLHR-L-16-0439

Laska, M., & Metzker, K. (1998). Food avoidance learning in squirrel monkeys and common marmosets. *Learning & Memory, 5(3)*, 193–203. https://doi.org/10.1101/lm.5.3.193

Lee, C.Y. (2011). The statistical learning perspective on Chinese reading. In P. McCardle, J. R. Lee, O. J. L. Tzeng, & B. Miller (Eds.), *Dyslexia across languages: Orthography and the brain-gene-behavior link*. Baltimore, MD: Brookes Publishing. pp. 44-61.

Lerner, I., Armstrong, B. C., & Frost, R. (2014). What can we learn from learning models about sensitivity to letter-order in visual word recognition? *Journal of Memory and Language, 77*, 40-58. http://dx.doi.org/10.1016/j.jml.2014.09.002

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14(6),* 1323–1329. https://doi.org/10.1111/j.1467-7687.2011.01079.x

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition, 122(2),* 241–246. https://doi.org/10.1016/j.cognition.2011.10.007

Liberman, A.M., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience, 16(3)*, 427–438. https://doi.org/10.1162/089892904322926764

Logan, G. D. (1996). The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychological Review, 103(4)*, 603–49. https://doi.org/10.1037/0033-295X.103.4.603

Lu, K., & Vicario, D. S. (2014). Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proceedings of the National Academy of Sciences, 111(40)*, 14553–14558. https://doi.org/10.1073/pnas.1412109111

Mackintosh, N.J. (1998). *IQ and human intelligence*. Oxford: Oxford University Press.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18*, 387–391. https://doi.org/10.1111/j.1467-9280.2007.01910.x

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80. https://doi.org/10.1126/science.284.5416.875a

Mareschal, D., & French, R. M. (2017). TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711)*, 20160057. https://doi.org/10.1098/rstb.2016.0057

McCauley, S.M. & Christiansen, M.H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review, 126*, 1-51. http://dx.doi.org/10.1037/rev0000126

McCauley, S. M., Isbilen, E. S., & Christiansen, M. H. (2017). Chunking ability shapes sentence processing at multiple levels of abstraction. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2681–2688). Austin, TX: Cognitive Science Society.

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

Mersad, K., & Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory & Cognition, 39*, 1085–1093. https://doi.org/10.3758/s13421-011-0074-3

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words:

Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90(2)*, 227–234. https://doi.org/10.1037/h0031564

Miller, G.A. (1958). Free recall of redundant strings of letters. Journal of Experimental Psychology, 56, 485-491.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning, 62*, 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1*, 31. https://doi.org/10.3389/fpsyg.2010.00031

Mitchel, A. D., & Weiss, D. J. (2010). What's in a face? Visual contribution to speech segmentation. *Language and Cognitive Processes, 25*, 456-482. https://doi.org/10.1080/01690960903209888

Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: Cross-modal effects in multisensory statistical learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*, 1081–1091. https://doi.org/10.1037/a0023700

Monroy C., Gerson S., Hunnius S. (2017). Infants' motor proficiency and statistical learning for actions. *Frontiers in Psychology, 8*, 2174. https://doi.org/10.3389/fpsyg.2017.02174.

Musz, E., Weber, M. J., & Thompson-Schill, S. L. (2015). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Attention, Perception, & Psychophysics, 77(1)*, 78–96. https://doi.org/10.3758/s13414-014-0757-5

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42(4)*, 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Nastase, S., Iacovella, V., & Hasson, U. (2014). Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Human Brain Mapping, 35*, 1111–1128. https://doi.org/10.1002/hbm.22238

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48(2)*, 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

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

Onnis, L., Christiansen, M.H., Chater, N. & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 886-891). Mahwah, NJ: Lawrence Erlbaum.

Onnis, L., Frank, M. C., Yun, H., & Lou-Magnuson, M. (2016). Statistical learning bias predicts second-language reading efficiency. In A. Papafragou, D. Grodner, & D. Mirman (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2105–2110). Austin, TX: Cognitive Science Society.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53(2),* 225–237. https://doi.org/10.1016/j.jml.2005.02.011

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105(32)*, 11087–11092. https://doi.org/10.1073/pnas.0805664105

Paraskevopoulos E., Chalas N., Kartsidis, P., Wollbrink, B., & Bamidis, P. (2018). Statistical learning of multisensory regularities is enhanced in musicians: An MEG study. *NeuroImage, 175*, 150-160. 10.1016/j.neuroimage.2018.04.002

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage, 50(1)*, 302–313. https://doi.org/10.1016/j.neuroimage.2009.12.019

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Pelucchi, B., Hay, J.F. & Saffran, J.R. (2009). Statistical learning in a natural language by 8-months old infants. *Child Development, 80*, 674-685. doi: 10.1111/j.1467-8624.2009.01290.x.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604-607.

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of*

Memory and Language, 66(4), 807–818. https://doi.org/10.1016/j.jml.2012.02.010

Perruchet, P., Poulin-Charronnat, B., Tillmann, B., & Peereman, R. (2014). New evidence for chunk-based models in word segmentation. *Acta Psychologica, 149*, 1–8. https://doi.org/10.1016/j.actpsy.2014.01.015

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*, 246–263. https://doi.org/10.1006/jmla.1998.2576

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences, 25(3),* 297-330. https://doi.org/10.1017/S0140525X02000067

Plaut, D., & Behrmann, M. (2011). Complementary neural representations for faces and words: A computational exploration. *Cognitive Neuropsychology, 28*, 251–275. https://doi.org/10.1080/02643294.2011.609812

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*, 227–244. https://doi.org/10.1037/0033-2909.133.2.227

Poulin-Charronnat, B., Perruchet, P., Tillmann, B., & Peereman, R. (2016). Familiar units prevail over statistical cues in word segmentation. *Psychological Research, 81*, 990–1003. https://doi.org/10.1007/s00426-016-0793-y

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review, 11(6)*, 1090–1098. https://doi.org/10.3758/BF03196742

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. *Developmental Science, 21*, e12593. https://doi.org/10.1111/desc.12593

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology. General, 125(2)*, 123–138. https://doi.org/10.1037/0096-3445.125.2.123

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395(6700)*, 376–381. https://doi.org/10.1038/26475

Rogers, TT., Lambon Ralph, M.A., Garrard, P., Bozeat, S., McClelland, J.L., Hodges, J.R., & Patterson, K. (2004). Structure and deterrioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111*, 205-235. https://dx.doi.org/10.1037/0033-295x.111.1.205

Rohrmeier M., & Widdess R. (2017). Incidental learning of melodic structure of north indian music. *Cognitive Science, 41*, 1299-1327. 10.1111/cogs.12404

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology, 44(4)*, 1100–1105. https://doi.org/10.1016/j.jesp.2007.12.001

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, *274(5294)*, 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70(1)*, 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8(2)*, 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Saffran, J., & Kirkham, N.Z. (2018). Infant statistical learning. *Annual Review of Psychology, 69*, 181-203. http://www.annualreviews.org/doi/10.1146/annurev-psych-122216-011805.

Santolin, C., & Saffran, J. R. (2017). Constraints on statistical learning across species. *Trends in Cognitive Sciences, 22*, 52-63. https://doi.org/10.1016/j.tics.2017.10.003

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

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

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711)*, 20160049. https://doi.org/10.1098/rstb.2016.0049

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology, 20*, 172–176.

https://doi.org/10.1016/j.conb.2010.02.010

Schuwerk T., Sodian B., & Paulus M.(2016). Cognitive mechanisms underlying action prediction in children and adults with autism spectrum condition. *Journal of Autism and Developmental Disorders, 69*, 800-816. 10.1007/s10803-016-2899-x

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599-1603.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy, 17*, 247–271. https://doi.org/10.1111/j.1532-7078.2011.00085.x

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17(3)*, 367–395. https://doi.org/10.1017/S0140525X00035032

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711)*, 20160059. https://doi.org/10.1098/rstb.2016.0059

Siegelman, N., Bogaerts, L., Arciuli, J. & Frost, R. (2018). Statistical entrenchment: prior knowledge impacts statistical learning performance. *Cognition, 177*, 198-213. https://doi: 10.1016/j.cognition.2018.04.011.

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2017). Redefining “learning” in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science, 42*, 692-727. https://doi.org/10.1111/cogs.12556

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Sigurdardottir, H. M., Danielsdottir, H. B., Gudmundsdottir, M., Hjartarson, K. H., Thorarinsdottir, E. A., & Kristjánsson, Á. (2017). Problems with visual statistical learning in developmental dyslexia. *Scientific Reports, 7*, 606. https://doi.org/10.1038/s41598-017-00554-5

Sloan, L.K. & Johnson, S.P. (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

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development, 16(3)*, 407–419. https://doi.org/10.1080/15248372.2014.933430

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological review, 125*, 969-984.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18(6)*, 643–662. https://doi.org/10.1037/h0054651

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93,* B1-B9. https://doi.org/10.1016/j.cognition.2003.09.011

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials*. BMC Neuroscience, 10(1)*, 21. https://doi.org/10.1186/1471-2202-10-21

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51(6)*, 599–606. https://doi.org/10.3758/BF03211656

Thiessen, E. D. (2011). Domain general constraints on statistical learning. *Child Development, 82(2)*, 462–470. https://doi.org/10.1111/j.1467-8624.2010.01522.x

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin, 139*, 792–814. https://doi.org/10.1037/a0030801

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12(12)*, 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Toro, J. M., Sebastián-Gallés, N., & Mattys, S. L. (2009). The role of perceptual salience during the segmentation of connected speech. *European Journal of Cognitive Psychology*, *21*(5), 786–800. https://doi.org/10.1080/09541440802405584

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25-B34. https://doi.org/10.1016/j.cognition.2005.01.006

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*, 867–875. https://doi.org/10.3758/BF03193539

Trecca, F., McCauley, S.M., Andersen, S.R., Bleses, D., Basbøll, H., Højen, A., Madsen, T.O., Ribu, I.S B., & Christiansen, M. H. (2019). Segmentation of highly vocalic speech via statistical learning: Insights from a cross-linguistic study of Danish, Norwegian, and English. *Language Learning, 69,* 143–176*.* https://doi.org/10.1111/lang.12325

Tummeltshammer, K. S., Mareschal, D., & Kirkham, N. Z. (2014). Infants' selective attention to reliable visual cues in the presence of salient distractors. *Child Development*, *85*(5), 1981–1994. https://doi.org/10.1111/cdev.12239

Tunney, R. J., & Altmann, G. T. M. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(5), 1322–1333.

Turk-Browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology-General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–45. https://doi.org/10.1162/jocn.2009.21131

Ulicheva, A., Harvey, H., Aronoff, M. & Rastle, K. (2017). *Statistical information encoded in English writing*. Spoken presentation at the Experimental Psychology Society meeting, July 14, 2017, Reading, England.

Van den Bos, E., Christiansen, M.H. & Misyak, J.B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language, 67*, 507- 520*.* http://dx.doi.org/10.1016/j.jml.2012.07.008

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent learning of adjacent and nonadjacent dependencies. *Language Learning, 66*, 8-30. https://doi.org/10.1111/lang.12137

Wang, B., & Theeuwes, J. (2018a). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception & Performance, 44, 13-17.*

Wang, B., & Theeuwes, J. (2018b). How to inhibit a distractor location? Statistical learning versus active, top-down suppression. *Attention, Perception, & Psychophysics, 80,* 860–870.

Wang, B., & Theeuwes, J. (2018c). Statistical regularities modulate attentional capture independent of search strategy. *Attention, Perception, & Psychophysics, 80*, 1763–

1774.

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech Segmentation in a Simulated Bilingual Environment: A Challenge for Statistical Learning? *Language Learning and Development*, *5*(1), 30–49. https://doi.org/10.1080/15475440802340101

Weiss, D.J., Poepsel, T., & Gerfen, C. (2015) Tracking multiple inputs: The challenge of bilingual statistical learning. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 167-190). John Benjamins Press.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Yang, C. D. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, *8*, 451–456. https://doi.org/10.1016/j.tics.2004.08.006

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics: Research article. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yorovsky, D., & Frank, M.C. (2015). An integrative account of constraints on cross-situational learning. *Cognition, 145*, 53-62.

Yu, R.Q. & Zhao, J. (2018). Object representations are biased toward each other through statistical learning. *Visual Cognition, 26*, 253-267. doi: 10.1080/13506285.2018.1435596

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science, 24*, 667-677. doi: 10.1177/0956797612460407

Zhao, J., & Luo, Y. (2017). Statistical regularities guide the spatial scale of attention. *Attention, Perception, & Psychophysics, 79*, 24-30. DOI 10.3758/s13414-016-1233-1

Zwebner, Y., Sellier, A.-L., Rosenfeld, N., Goldenberg, J., & Mayo, R. (2017). We look like our names: The manifestation of name stereotypes in facial appearance. *Journal of Personality and Social Psychology*, *112*(4), 527–554. https://doi.org/10.1037/pspa0000076.