Is there such a thing as a “good statistical learner”?

Louisa Bogaerts[1], Noam Siegelman[2], Morten H. Christiansen[2,3,4]
& Ram Frost [2,5,6]

[1] Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands
[2] Haskins Laboratories, CT 06511 New Haven, USA
[3] Cornell University, NY 14850 Ithaca, USA
[4] Aarhus University, 8000 Aarhus, Denmark
[5] The Hebrew University of Jerusalem, 91905 Jerusalem, Israel
[6] Basque Center for Cognition, Brain, and Language, 20009 Donostia, Spain

Word count: 4422

Keywords: individual differences, statistical learning, cognitive abilities

Abstract

A growing body of research investigates individual differences in the learning of statistical structure, tying them to variability in cognitive (dis)abilities. This approach views statistical learning (SL) as a general individual ability that underlies performance across a range of cognitive domains. But is there a general SL capacity that can sort individuals from “bad” to “good” statistical learners? Explicating the suppositions underlying this approach, we suggest that current evidence supporting it is meager. We outline an alternative perspective that considers the variability of statistical environments within different cognitive domains. Once we focus on learning that is tuned to the statistics of real-world sensory inputs, an alternative view of SL computations emerges with a radically different outlook for SL research.

Individual differences in statistical learning

Recent years have seen a growing body of research tying variation in a range of cognitive capacities to success or failure in assimilating the statistical structure of the input. This reflects an increased appreciation that our environment—be it perceptual, cognitive, or social—is saturated with statistical regularities that are the target of learning and processing. The neurocognitive mechanism for detecting and assimilating the range of regularities in the input has been labelled "statistical learning" (SL) [1–4]. Although the impact of **statistical regularities** (see **Glossary**) on cognitive processing had been previously recognized, the interest in SL surged after the seminal paper by Saffran and colleagues on speech segmentation [1]. The concept of SL has subsequently permeated many other cognitive **domains** (e.g., visual perception, music, social cognition, attention, etc.; see [5] for review), because they all involve statistical structure.

With this new perspective on cognition came a novel prediction: That **individual differences** in these various domains are fundamentally linked to SL capability. As a result, the last decade has seen a growing body of work targeting SL as a general individual ability for perceiving and assimilating regularities in the input. The main premise of this research is that individuals range from "good" to "bad" statistical learners, and that "good" statistical learners are expected to have better skills across the wide range of cognitive functions that require the assimilation of statistical structure (e.g., reading [6–8], early language development [9,10], syntactic processing [11,12], object and scene perception [13,14], music [15,16] etc.). Many recent studies, ours included [7,17,18], have consequently assessed correlations between performance in laboratory SL tasks and cognitive abilities in a variety of domains, in normal and special populations. A few studies, in particular those investigating language and literacy acquisition, have tested more narrow and nuanced predictions about the predictive power of individual differences in SL, for example by linking the sensitivity to orthography-to-phonology regularities to early reading skills [19], or by establishing a relation between infants' knowledge of their native language's sound structure and their vocabulary size [20]. However, most studies have selected a given SL task, assuming that performance on the chosen task is sufficiently representative of one's general SL capacity to be predictive of the targeted cognitive ability (or disability), be it reading, musicality, or social skills, to name a few. Although results have not been unequivocal [18,21–23], and although effect sizes are often small, most published work has reported significant positive correlations between SL performance and performance in multiple cognitive functions (see **Table 1**). Typically, null effects within this research line have been discussed in terms of insufficient variability in performance [24] and poor task reliability more generally [18,23]. Importantly, underlying this experimental approach is the (typically implicit) supposition that an individual has a general, unitary ability for discovering regularities which assists the learning of any type of statistical structure. In some studies this supposition is formulated explicitly, as can be seen in the following quotes:

> Parks et al. [25] state "We are interested in how the ability to learn patterns overall is related to language and social competency skills […]. It is therefore expected that auditory and visual statistical learning will contribute similarly […] given that both tasks assess the ability to learn statistical patterns in general." (p. 3)

> Kirkham, Slemmer and Johnson [26] write "These results are consistent with the existence of a domain general statistical learning device that is available to even very young infants […]." (p. 40)

Table 1. Examples of studies tying individual differences in the learning of statistical structure to variance in cognitive abilities.

Predicted cognitive ability	Cognitive measure	Statistical learning task, learning measure(s)	Stimuli of statistical learning task	Sample (age)	Main findings	Reference
Literacy	Sentence reading Word and nonword reading	Auditory triplet learning, acceleration of target detection times during familiarization & 2-Alternative Forced Choice (2-AFC) familiarity test	Pure tones	Adults (18–34 years) Children (8–16 years)	Full sample: positive correlation between 2-AFC measure and sentence reading, null findings with acceleration measure Children: positive correlations between the acceleration measure and word and nonword reading, null findings with 2-AFC	[27]
		Visual triplet learning, acceleration of target detection times during familiarization &	Alien figures		Full sample: positive correlation between 2-AFC measure and sentence reading, null	

		2-AFC familiarity test			findings with acceleration measure Children: null findings	
	Word and nonword reading Morphological priming	Visual triplet learning, 2-AFC familiarity test	Abstract shapes	Adults (18–34 years), native English speakers learning Hebrew	Positive correlation with all reading measures	[7]
	Word reading	Visual triplet learning, 2-AFC familiarity test	Alien figures	Adults (18–34 years) Children (6.4-12.5 years)	Adults: Positive correlation Children: Positive correlation	[28]
	Word and nonword reading Spelling test	Visual triplet learning, self-paced measure during familiarization & Pattern completion test & 2-AFC familiarity test	Alien figures	Children (8.3–11.2 years) with and without a dyslexia diagnosis	Null findings: no evidence of a relationship between any of the SL measures and reading or spelling skills above and beyond participant-level variables	[23]
		Serial reaction time task	4 locations			
Oral language processing	Lexical-processing efficiency Vocabulary size	Auditory pair learning, 2-AFC familiarity test with head-turn preference measure	Syllables Words	Infants (15-16 months)	Positive correlations with lexical-processing efficiency, null findings for vocabulary size	[9]
	Vocabulary size & growth	Auditory non-adjacent dependency learning, 2-AFC familiarity test with head turn preference measure	Syllables	Infants (15.5-17.5 months, tested at multiple time points till the age of 30 months)	Positive correlation with vocabulary size (at multiple time points), null findings for vocabulary growth	[10]
	Syntax comprehension Vocabulary	Visual triplet learning, 2-AFC familiarity test	Alien figures	Children (6.1–8.1 years)	SL independently predicts comprehension of passives and object relative clauses, null findings for other grammatical structures and vocabulary	[24]
Music skills	Melody discrimination & Rhythm discrimination	Auditory triplet learning, 2-AFC familiarity test	Pure tones	Children (M = 10.3 years) with and without musical training	Positive correlation with general music score	[29]

	(combined in a general music score)	Visual triplet learning, 2-AFC familiarity test	Alien figures		Positive correlation with general music score	
Social competency	Social competency questionnaire Autistic traits questionnaire Receptive and expressive language abilities	Visual triplet learning, psychometrically optimized familiarity test	Abstract shapes	Young adults (16-21 years)	Positive correlation with receptive language and social competency abilities, null finding for relation with autism symptomatology	[25]
		Auditory triplet learning, 2-AFC familiarity test	Syllables		Null findings for receptive language and social competency abilities, positive correlation to autism symptomatology	
Feature-comparison skills	Visual comparison performance	Visual distributional statistical learning, psychometrically optimized familiarity test & frequency estimates	Abstract shapes	Young adults (17-26 years), trained forensic examiners and novices (informed, uninformed and misinformed)	Informed novices: Positive correlation between familiarity measure and visual comparison performance, null findings for frequency estimates measure Null findings for all other groups	[30]

From a historical perspective, this approach to individual differences in SL resonates with research into other general cognitive capacities, such as the study of human intelligence, with its G-factor, or memory, with its general working memory factor (see **Box 1**). It assumes that a general SL capacity determines individual performance in regularity learning across domains, resulting in something akin to a "general SL-factor." As the qualification "general" has also been used in the context of discussing domain-specificity vs. domain-generality (see [31] for discussion), we should clarify that a "general SL capacity," as used here, implies that SL is a domain-general ability, whereas domain-generality does not require the existence of a unitary SL capacity. Rather, "domain-generality" in the context of SL research reflects the recognition that sensitivity to regularities is found across all cognitive domains, and extends beyond the original finding of sensitivity to trisyllabic patterns in a continuous speech stream [1] (see [5] for discussion). The idea of a general SL capacity is a more specific claim: It presupposes that individuals differ in their general ability for learning regularities, whatever those regularities are, and that this general capacity contributes to their learning in any domain. As such, sensitivity to statistical regularities is taken to be a major cognitive construct, subserving basic and higher-order cognitive functions, thus impacting human performance across the board. Importantly, this unitary view assumes that there is something common to the computation of statistical regularities across modalities and domains, leading to some shared variance in individuals' performance in assimilating regularities across cognition.

The possibility of a general SL-factor, common to learning regularities across domains, has far-reaching theoretical and practical implications. It suggests that a general computational device assimilates the wide range of regularities in the environment, and that individuals differ in its efficiency. Even more importantly, since performance in multiple SL tasks was found to be independent of intelligence, working memory, and executive functions [32,33], a general SL-factor has the promise to account for a substantial portion of unexplained variance in cognitive performance. If a general SL-factor could be comprehensively assessed through a validated and normed test battery, similar to the G-factor, a general SL score could provide a reliable estimate of an individual's SL capacity relatively to the population distribution. Then, this single general SL score could predict, at least to some extent, an individual's performance in a given cognitive function over and above intelligence or memory. Because SL is an important building block of virtually all current theories of cognitive processing, this could revolutionize research on individual differences in cognitive science.

In this paper, we evaluate this intriguing prospect and outline some of the challenges it might face. We start by discussing what a general SL-factor would imply as a theoretical construct, before considering evidence for the notion of a "good statistical learner." Next, we outline a broader ecological perspective on the variety of statistics that need to be accommodated and consider existing challenges for the notion of "good statistical learners." We then outline an alternative view of SL computations and discuss its implications for future research.

What would a general statistical learning factor imply?

Theoretical constructs should be well-defined so that they can be empirically validated. We thus start by outlining three implicit suppositions that underlie the concepts of "good statistical learners" and a general SL-factor.

First, and foremost, there is the supposition of nesting and sharing. A general SL-factor implies that all **modality**- and domain-specific SL abilities (e.g., detecting word-boundaries, learning spatial contingencies, etc.) are nested within it, just as vocabulary, comprehension, and visual-spatial abilities are nested within intelligence. Nesting could be hierarchical or not [34], but it necessarily entails a relation of whole and parts between the general factor and its components. Nesting leads inevitably to sharing. Given that statistical regularities vary in sensory modality, material, type of contingencies, etc., recent studies have argued that SL is a componential ability spanning an array of dimensions [35–39]. However, if all SL dimensions are nested within a general SL-factor, they should share some variance, which reflects the commonality of all SL computations. Sharing could result from all facets having some positive load on the general factor (as Spearman originally postulated for intelligence [40]), and/or from some facets partially overlapping because they implicate shared computations. We note that sharing does not preclude the possibility that some (additional) shared variance in performance is due to factors external to SL per se (e.g., attention); we clarify, however, that the sharing assumption refers to common variance originating specifically from shared SL computations rather than from an external third factor.

The next two suppositions are related to the possibility of assessing individuals' SL ability as ranging from "good" to "bad." First, tying "low," "mid-range" or "high" scores in a cognitive function to "low," "mid-range" and "high" scores in an SL task (as done in the studies of **Table 1**) assumes that SL performance displays monotonicity. Monotonicity implies that given valid and reliable measurements, higher scores would reflect better SL performance, pointing to "good" statistical learners, whereas lower scores would reflect worse SL performance, pointing to "bad" statistical learners. Monotonicity by no means implies linearity; it simply requires an ordinal scale. It is worth noting that monotonicity could still hold even in the absence of a general factor, if performance in different SL systems displays a monotonic continuum. However, the backbone of the concept of a "good statistical learner," as it currently appears in the literature, is that individuals can be differentiated along a unified continuum, once their ability is reliably and validly measured. Second, from a psychometric perspective, the alluring prospect of assessing individuals' general SL ability using a single score through a test battery requires aggregability: Performance across the range of SL dimensions could, in principle, be aggregated (potentially with weighting, so that some facets contribute more than others), to give rise to a single score reflecting the general factor, similar to the aggregation of subtest of intelligence to provide a general score of intellectual ability.

Now that the basic suppositions underlying the notion of “good statistical learners” are laid out, we examine to what extent they withstand empirical and theoretical scrutiny.

Evidence in favor of a general statistical learning ability

Several studies (listed in **Table 1**) have found significant positive correlations between performance on a SL task and a range of cognitive skills. Importantly, some of these correlations were observed when the same task predicted different functions in different modalities (e.g., a similar visual **embedded pattern learning task** with alien-like figures correlates with both reading abilities [28] and musical skills [29]). This suggests that a given SL task reflects a general ability for learning regularities, so that it can simultaneously predict performance across different cognitive domains. In the same vein, a given cognitive function (e.g., reading skill) was predicted by two different SL tasks, one involving abstract shapes [7] and one involving auditory tones [27]. The finding that two different SL tasks in different sensory modalities both have predictive value for individual differences in a given domain, suggests that they at least partially represent the same general ability. Another piece of evidence for shared computations across modalities comes from work that revealed shared variance between visual and auditory SL tasks. For example, a study using non-linguistic auditory materials, which do not implicate learners' prior language knowledge, obtained a significant correlation between SL performance in the visual and the auditory modality [41]. Further, from a neurobiological perspective, imaging studies consistently report activation of the same subset of brain regions in SL tasks across modalities and stimuli (see [42] for review). These domain-general regions seem to point to common neurocircuitry involved in processing statistical regularities regardless of specific input characteristics. Taken together, all these findings coincide with the claim that the variety of SL tasks taps a common factor, presumably related to a general ability to register statistical regularities across domains.

We argue, however, that these findings should be interpreted with caution. The correlations between visual and auditory SL tasks might be driven by the significant similarity in the statistical patterns they employ (e.g., pairs or triplets within a continuous sensory stream).

Thus, finding similarities in learning embedded pairs or triplets of syllables, musical tones, natural sounds, shapes, aliens figures, or objects may speak to the uniformity of the artificial tasks that are typically employed to tap SL, not to capturing the statistics of real-world sensory environments. Furthermore, most of these studies use a two-alternative forced-choice paradigm to test knowledge of regularities and thus all require meta-cognitive decision processing [43], which may contribute to the observed correlations. In the same vein, the domain-general neurocircuitry that is activated in these tasks (mainly the Medial Temporal Lobe (MTL) memory system, [44,45]) may reflect the inevitable hippocampal involvement in learning a limited set of embedded patterns in the artificial stream, and does not necessarily speak to the long, continuous process of assimilating the statistical distributions characteristic of the real-world environment. As to the reported correlations between SL and cognitive outcomes, they are generally weak—significantly weaker than those reported in the domain of general intelligence and memory, and furthermore, there are multiple reports of null results (see **Table 1**). Even when observed, the weak correlations could have been driven by a range of mediating factors and overlap in task demands. For example, typical SL tasks engage sustained attention and require fast intuitive judgments [33,46–48], hence inter-individual differences in these capacities could similarly impact performance in the SL tasks and the measured cognitive outcome, leading to the observed small correlations (see [49] for discussion).

An ecological perspective

Our starting point is that SL mechanisms are meant to assimilate the statistics of the real-world environments, be it print, spoken language, objects, or visual scenes. As such, an adequate SL account of a given domain should consider the rich and idiosyncratic scope of the statistical regularities that characterize it. When this approach is adopted, it becomes apparent that the statistical patterns that need to be assimilated for different cognitive functions differ, and can vary quite dramatically. In light of these differences in input structure across domains, a key question is whether there are overarching SL computations that are involved regardless of the nature of the input, and if so, what are they?

Computational models of SL have mainly focused on co-occurrence learning and the segmentation of continuous, patterned input streams. For example, models such as PARSER [50] and TRACX [51,52] have proposed chunk extraction as an alternative learning mechanism. A recent biologically inspired neural network model offered an architecture where a hippocampal monosynaptic pathway drives the learning of regularities [45]. This model can simulate the learning of simple patterns in an artificial SL task, and also more complex statistics (e.g., small ‘community structures’ [53]). However, since these computational accounts focus on the specific issue of how boundaries are extracted from continuous input, they are limited in their explanatory scope when it comes to explaining the learning of the large set of real-world regularities. It remains an open question whether a single **computational mechanism** can deal with them all.

To exemplify this issue, we consider two well-studied cognitive functions as test cases: reading and visual object perception. We show that on a conceptual level, the to-be-learned regularities vary substantially even within two domains that both involve the visual modality, suggesting that uncovering a common computational principle might not be an easy task. Finding common computational principles across all domains and modalities is likely to be even more challenging.

Reading. Readers are sensitive to a range of statistical regularities, including frequency of letters and words [54,55], the co-occurrence of letters [56,57], correlations between letters and speech sounds [58], between letter combinations and stress patterns [59], and between letters and semantic meaning through morphological structure [60,61]. Readers are also affected by the likely position of letters within words (e.g., ‘er’ being a likely word ending), the morphological information letters convey given their location (e.g., ‘er’ anywhere but in final position is probably not a morpheme [62]), the predictability of words in sentences [63,64], the contextual similarity among alpha-numeric characters in text [65], the sequential order of potential word lengths in sentences [66], syntactic and semantic plausibility [67,68], and this is not an exhaustive list. All of these different types of regularities are “statistical” in nature, and thus fall under the general label of “statistical learning.” However, the computations that they implicate are potentially quite different from one another. To exemplify, the computational solutions for learning the correlations of letters with sounds and meaning do not have clear overlap with the computational solution for predicting, say, the length of an upcoming word given the previous word-lengths. Importantly, as detailed in the next section, these statistical computations are even more distant from those that subserve efficient visual object recognition and scene perception.

Visual object and scene perception. Our visual world is complex in nature, but rarely presents randomness. Humans are sensitive to both the physical and contextual regularities that characterize our visual environment. One striking example is that vertical and horizontal orientations occur much more frequently than oblique orientations in both man-made and natural environments [69]. Indeed, participants have been found to perceive vertical and horizontal orientations better then oblique orientations, suggesting a tuning of the perceptual system to real-world statistics. Similarly, light usually comes from above [70] and this results in a strong perceptual prior to interpret the source of light as such [71]. Further, different scene and object categories (e.g., forests, beaches, streets, natural objects vs. man-made objects, portraits, etc.) were found to have characteristic spectral signatures which can be determined by averaging hundreds of images of the same category [72]. These summary statistics seem to aid perception of objects which are congruent with the category [73], suggesting that our perceptual system tracks spectral statistics. Recent work further suggests that if visual objects regularly co-occur in time, their representations within the MTL become more similar one to the other [44], so that the system is tuned to track temporal co-occurrence statistics. In the same vein, the perceptual system also tracks co-occurrence in space, so that objects that appear together in a given spatial composition engage attention as if they are a single object [74,75]. Another source of statistical regularities concerns the typical location of objects in specific types of scenes. Object identification has been shown to be facilitated by presentation in congruent context scenes (e.g., a teapot in a kitchen rather than at a beach), and within a typical scene structure (e.g., a computer mouse positioned on the table next to the computer rather than on a computer screen [76,77]; see [78] for review).

Similar to reading, this very brief summary outlines the wide range of statistical regularities that are computed by the visual system in the domain of object and scene perception. Merging the two overviews together, it is clear that printed texts and scenes are characterized by a range of probabilistic regularities creating structure, and that readers and scene perceivers assimilate these. Shared variance in SL performance across these two cognitive functions would only emerge if they rely on mechanisms that share input representations or computations. Finding a computational common denominator for such distinct domains is perhaps possible, but does not seem an easy task.

Challenges for "good statistical learners"

We now consider whether computations of statistics of real-world sensory environments display the three implicit suppositions underlying the concepts of "good statistical learners" and a general SL-factor. When it comes to our test cases of reading and visual object perception, the existence of nesting and variance sharing remains an open question. Do efficient readers also perceive objects and scenes faster or better? Do participants who more rapidly identify objects in a congruent context and within a typical scene structure [77] also show, say, higher predictability effects of words in a sentence [64]? The assumption of good general statistical learners implicates that some individuals are proficient at picking up the statistical structure of the environment across all cognitive domains, while others are relatively poor across all domains. However, to our knowledge, there is no empirical evidence that speaks to this issue. In **Box 2** we outline specific types of evidence that are predicted by a general SL device.

The concept of "good statistical learners" faces additional challenges when considering monotonicity. Some environments are characterized by stable statistics while other are characterized by constant change. For example, the statistics of the visual world are more or less constant, whereas the characteristics of printed material change across different genres of text [79,80]. In fact, even at a given time and a given developmental phase, the statistical environment of one text may be quite different than that of another (e.g., different novels written in different periods, etc.). Recent evidence suggests that readers adapt to the statistical properties of a particular novel (for example, the sequential combinations of word-lengths in sentences and characteristic syntactic structures), and this results in more efficient ocular movements, as reflected by reduced viewing time [66]. Hence, for optimal reading performance, SL computations should be optimally flexible; not too flexible, so as to preserve the accumulated reading experience, but not too rigid, to allow efficient adaptation to novel statistics. Such a "sweet-spot" in the sensitivity and the attention to regularities in the input [81] challenges the monotonicity assumption. It further suggests that what it takes to be a good statistical learner may be quite different across domains. In some domains, a good statistical learner displays high sensitivity to statistical properties as well as high rigidity, relying strongly on long-term statistics. In others domains, a good statistical learner is characterized by more flexibility, relying more heavily on recent experiences (see also [82] for an implementation in a Bayesian framework). One could propose a definition specifying that individuals are good statistical learners when they optimally weight both long-term statistics and recent experiences, depending on the task at hand and the stability of the relevant input, but the definition of a good statistical learner then becomes domain-dependent. In all laboratory SL tasks as currently used, the more novel patterns are assimilated, the better learning is considered to be. However, moving to real-world sensory environments with some domains implicating monotonicity and some not, the aggregability assumption, too, faces significant challenges. The goal of aggregating SL abilities across domains in the hope of converging on a general score may thus be intractable.

An alternative approach: Statistical learning from an ecological perspective

To understand how individual differences in SL might contribute to variation in cognition, we need a different perspective. **Figure 1** illustrates the general SL ability account, contrasting it with an alternative theoretical approach which posits multiple specific SL-computations in different cognitive domains. At the cognitive level, the difference between the two accounts is most prominent in the presence of a single SL construct involved in assimilating a range of regularities in different environments vs. independent SL computations which are bound to a given domain. At the mechanistic level, a general ability implies one common computational mechanism that assimilates the range of regularities across different types of environments independently of the statistics involved. The figure exemplifies this through one recent candidate model where the extraction of regularities across domains relies on computations in the hippocampus [45]. Whereas in this example both the computations and the neurobiological substrate are unified, this is not a necessity. In principle, a shared set of SL computations could be carried out by different neural substrates (i.e., either because computations are distributed across multiple separate substrates or because multiple separate substrates each perform the same set of computations on different representations [42]), yet when resulting in shared variance this would per our view still be a general SL ability. For the alternative account, on the other hand, sensitivity to specific regularities is an emergent property of different mechanisms that process input in particular domains (e.g., reading [83], syntactic processing [84], object perception [85]) given their differing computational constraints.

Contrasting the two accounts: Empirical implications

Generic laboratory tasks (e.g., focusing on the ability to extract pair/triplet patterns based on transitional probabilities between individual stimuli) have helped establish SL as a powerful form of implicit learning. They have shown that the learning of statistical structure is possible across a variety of sensory modalities and domains (e.g., [11,37,86–88] and see [5] for a review), throughout the human lifespan [36,89–91], across species [92–94], and does not require instruction, reinforcement, or feedback [95]. However, changing the focus to SL mechanisms which are tuned to the complex range of statistical regularities characterizing real-world sensory environments (rather than the simple statistics of typical SL tasks) leads to a radically different course of future SL research, where such generic tasks no longer suffice. **Box 3** outlines the blueprint for such future research program.

It is evident that to determine whether there is such a thing like a "good statistical learner," a deeper understanding of the statistical environments that characterize a range of cognitive domains is required as a first step. This research should be complemented by empirical evidence regarding which of the revealed statistical regularities are the target of learning ([73], see also [97–99]), and modulate behavior. In addition, future advances in computational models are needed, to explicitly connect the statistical regularities learners actually assimilate in different domains to the cognitive and neural mechanisms that are responsible for learning them. Once theories and models of the statistical computations across cognitive domains are formulated, the viability of a general SL ability can be assessed through the study of individual differences, neurological patients and special populations with hypothesized deficits in SL. Research on impaired populations is particularly informative for this debate. The general SL capacity perspective predicts that impaired SL would result in difficulties acquiring sensitivity to statistical structure across the board. The alternative account of multiple specific SL abilities is, in contrast, consistent with domain-selective impairments [100].

Concluding Remarks

If a general SL-factor exists and a methodology for its comprehensive assessment can be developed, the practical and theoretical implications would be far-reaching. However, as we have argued above, the existence of SL as a general individual ability faces significant challenges. We have suggested an alternative perspective according to which sensitivity to statistical regularities in different domains is more likely grounded in different computational mechanisms. In this perspective, what all SL computations have in common is a very abstract notion of dealing with some sort of "regularity." Current evidence on individual differences in SL performance has severe limitations in determining which model should be favored. Most experimental SL paradigms mimic one another in terms of the statistical patterns they employ, rather than mimicking the statistical regularities that are the object of learning in different domains. To contrast theoretical approaches to SL, future work should focus on characterizing the different statistical environments in a multitude of cognitive domains (see **Outstanding Questions**). Without evidence from tasks that tap regularities characteristic of real-world environments in different domains, research that ties individual differences in a cognitive function to a general SL capacity stands on shaky theoretical grounds.

Box 1. A short history of salient general factors in cognitive science

Higher-order latent variables have been proposed across a wide range of cognitive abilities. Here we outline two examples of impactful general factors.

Intelligence. More than a century ago, Charles Spearman demonstrated that different measures of intelligence tend to correlate with each other to various degrees—known as the positive manifold. He proposed the two-factor theory of intelligence, stating that intellectual abilities are comprised of two kinds of factors: (1) a general ability labeled the G-factor, and (2) a number of specific abilities (S-factors), all having some load on the general factor [40]. Whereas conceptualizations of intelligence have since further evolved, the G-factor is still omnipresent and has been validated cross-culturally [101]. The current version of the Wechsler Adult Intelligence Scale (WAIS-IV, [102]) still provides a broad IQ-score to summarize general intellectual ability, which results from aggregating performance across a range of specific tasks and is taken to predict a wide range of cognitive functions.

Working memory. Together with intelligence, working memory has been one of the most frequently studied constructs in cognitive science. Working memory has been suggested to modulate a range of cognitive abilities (e.g., reading, mathematics [103]). Some work employing confirmatory factor analyses has supported the concept of a general, higher-order working memory capacity factor and hence the view that a broad set of tasks that use different working memory contents (e.g., verbal, visuospatial) and tap different processing demands (e.g., maintenance, updating) all purportedly engage a higher-level capacity [104–106].

Box 2. Hypothetical empirical evidence in support of a general SL ability

- Systematic positive correlations between sensitivity to regularities across domains: e.g., individuals who are more sensitive to spelling patterns in print are also more sensitive to chord and note co-occurrence in music, to conditional probabilities of objects in visual scenes, and to correlations between facial features and emotions. These correlations are found between tasks that tap sensitivity to these statistical structures but not with other tasks, demonstrating that they are not driven by factors external to SL.
- A similar developmental trajectory of sensitivity to statistical regularities across different domains, mirroring the developmental stages of the general learning device.
- Evidence from special populations (individuals with brain damage or neurodevelopmental disorders) of hindered learning of regularities across domains,

which can be traced back to an impairment of the general SL device, and cannot be explained by general cognitive factors such as memory, attention, etc.

- A unified (neuro)computational model architecture, implemented in different domains and operating on different inputs, can successfully learn different real-world regularities, from vision to language to social cognition.

Box 3. Proposed blueprint for future research

Developing and testing an **ecologically-valid** theory of regularity learning could proceed along the following sequence of 3 steps:

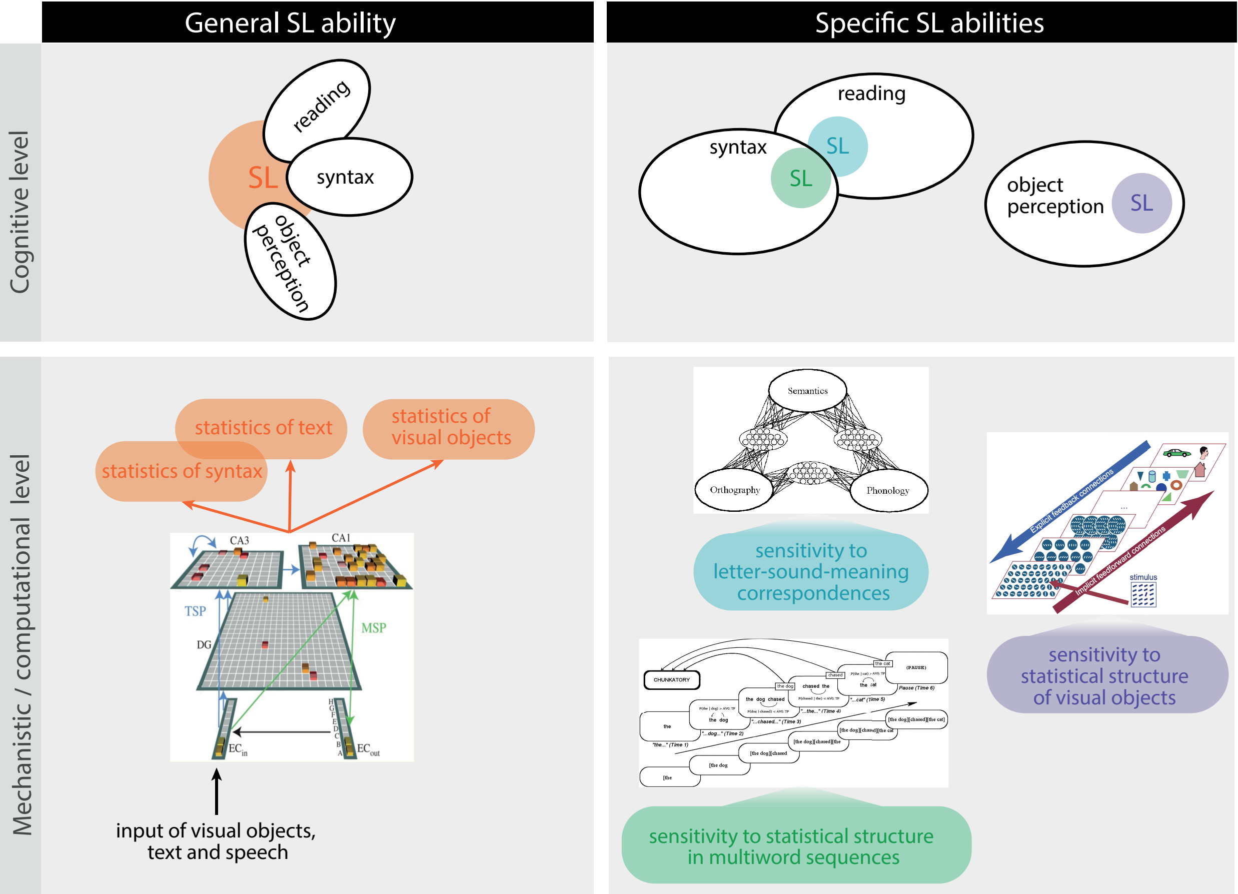
1. Map the domains of cognition that are characterized by significant structure (e.g., speech, print, syntax, music, objects, scenes, etc.) to identify the range of statistical regularities that characterize a given domain and could be the target of learning. Corpus analyses are an important tool in revealing the statistical regularities that exist in a domain [72,107,108]. In identifying domain-bound regularities, an important consideration is the experience of the learner, and how it is shaped over time and across development. "Big data" capturing everyday environments from the learner's point of view are therefore of great value (see [109] for discussion).
2. Use computational modeling to elucidate the possible computations that can account for the learning of different regularities within a given cognitive domain. This modeling should involve datasets that capture the environment within which real-world learning takes place, to show whether and how the relevant statistical information can be utilized.
3. Provide empirical evidence regarding which of the revealed real-world statistical regularities are indeed perceived and learned by individuals (at different stages of development), as well as the role they play in assisting processing in a given domain.

Glossary

- **Computational mechanism:** defined by the representations that are being processed, and by the transformation(s) applied to the input to generate the output (i.e., the assimilated regularities).
- **Domain:** cognitive performance can be conceptualized in terms of different domains of functioning (e.g., language, visual perception, attention, social cognition).
- **Embedded pattern learning task:** a classic task used to measure SL ability. It involves the presentation of a continuous visual or auditory stream made up of embedded patterns, followed by a test that assesses the preference for the embedded patterns (over foil patterns).
- **Ecological validity:** concerns the ability to generalize from the data observed under experimental settings to the state of affairs and natural behaviors in the world.
- **Individual differences:** in the context of SL, individual differences typically refer to quantitative differences in learning outcomes between learners, but could in principle refer to any quantitative or qualitative inter-individual variance (differences in the speed and trajectory of learning, individual variation in the adaptability to changing environments, etc.).
- **Modality:** the sensory mode of stimuli (e.g., vision, audition, touch). Note the dissociation between modality and domain: e.g., music and language are both in the same (auditory) modality but constitute separate domains.
- **Statistical regularities**: here we focus on the wide range of constancies in the input that provide information regarding patterning (in time and/or space) in the environment.

Figure legend

Figure 1. Depiction of the general SL ability vs. multiple specific SL-abilities. The upper panels illustrate the differences at the cognitive level, the bottom panel illustrates these differences at a mechanistic/computational level. The selected models of domain-general SL computation [45], reading [83] (as depicted by [110]), syntactic processing [84] and object perception [85] are taken as figurative examples of computational implementations of SL and the specific domains.


General SL ability
Specific SL abilities
Cognitive level
Mechanistic / computational level
reading
SL
syntax
object perception
reading
syntax
SL
SL
object perception
SL
statistics of text
statistics of visual objects
statistics of syntax
CA3
CA1
TSP
MSP
DG
input of visual objects, text and speech
Semantics
Orthography
Phonology
sensitivity to letter-sound-meaning correspondences
Explicit feedback connections
Implicit feedforward connections
stimulus
sensitivity to statistical structure of visual objects
CHUNKATORY
sensitivity to statistical structure in multiword sequences

Acknowledgement

This article was supported by the European Research Council Advanced Grant Project 692502-L2STAT and the Israel Science Foundation (ISF) Grant Project 705/20, awarded to R. Frost. N.S. received funding from the ISF, grant number 48/20. We thank Merav Ahissar, Erin Isbilen and Felicity Frinsel for their feedback on an earlier version of the paper and Blair Armstrong for the helpful discussions we had regarding computational mechanisms.

References

1 Saffran, J.R. *et al.* (1996) Statistical learning by 8-month-old infants. *Science (80-.).* 274, 1926–1928

2 Kirkham, N.Z. *et al.* (2002) Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* 83,

3 Gebhart, A.L. *et al.* (2009) Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* 16, 486–90

4 Sherman, B.E. *et al.* The prevalence and importance of statistical learning in human cognition and behavior. , *Current Opinion in Behavioral Sciences*, 32. 01-Apr-(2020) , Elsevier Ltd, 15–20

5 Frost, R. *et al.* (2019) Statistical learning research: A critical review and possible new directions. *Psychol. Bull.* DOI: 10.1037/bul0000210

6 Arciuli, J. (2018) Reading as statistical learning. *Lang. Speech. Hear. Serv. Sch.* 49, 634–643

7 Frost, R. *et al.* (2013) What Predicts Successful Literacy Acquisition in a Second Language? *Psychol. Sci.* 24, 1243–1252

8 Chetail, F. (2015) Reconsidering the role of orthographic redundancy in visual word recognition. *Front. Psychol.* 0, 645

9 Lany, J. *et al.* (2018) Infant statistical-learning ability is related to real-Time language processing. *J. Child Lang.* 45, 368–391

10 Frost, R.L.A. *et al.* (2020) Non-adjacent dependency learning in infancy, and its link to language development. *Cogn. Psychol.* 120, 101291

11 Saffran, J.R. and Wilson, D.P. (2003) From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy* 4, 273–284

12 Gerken, L. *et al.* (2005) Infants can use distributional cues to form syntactic categories. *J. Child Lang.* 32, 249–268

13 Fiser, J. and Aslin, R.N. (2005) Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 134, 521–37

14 Turk-Browne, N.B. *et al.* (2010) Implicit perceptual anticipation triggered by statistical learning. *J. Neurosci.* 30, 11177–11187

15 Daikoku, T. (2019) Statistical learning and the uncertainty of melody and bass line in music. *PLoS One* 14, e0226734

16 Tillmann, B. and McAdams, S. (2004) Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1131–1142

17 Misyak, J.B. and Christiansen, M.H. (2012) Statistical Learning and Language: An Individual Differences Study. *Lang. Learn.* 62, 302–331

18 Pavlidou, E. and Bogaerts, L. (2019) Implicit Statistical Learning across Modalities and its Relationship with Reading in Childhood. *Front. Psychol.* 10, 1834

19 Siegelman, N. *et al.* (2020) Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. *J. Mem. Lang.* 114, 104145

20 Graf Estes, K. *et al.* (2016) Finding patterns and learning words: Infant phonotactic knowledge is associated with vocabulary size. *J. Exp. Child Psychol.* 146, 34–49

21 Schmalz, X. *et al.* (2019) Is Statistical Learning Ability Related to Reading Ability, and If So, Why? *Sci. Stud. Read.* 23, 64–76

22 West, G. *et al.* (2018) The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Dev. Sci.* 21,

23 van Witteloostuijn, M. *et al.* (2021) The contribution of individual differences in statistical learning to reading and spelling performance in children with and without

dyslexia. *Dyslexia* 27, 168–186
24 Kidd, E. and Arciuli, J. (2015) Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Dev.* 87, 184–93
25 Parks, K.M.A. *et al.* (2020) Statistical Learning and Social Competency: The Mediating Role of Language. *Sci. Rep.* 10, 1–15
26 Jeste, S.S. *et al.* (2015) Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Dev. Sci.* 18, 90–105
27 Qi, Z. *et al.* (2019) Hearing Matters More Than Seeing: A Cross-Modality Study of Statistical Learning and Reading Ability. *Sci. Stud. Read.* 23, 101–115
28 Arciuli, J. and Simpson, I.C. (2012) Statistical Learning Is Related to Reading Ability in Children and Adults. *Cogn. Sci.* 36, 286–304
29 Mandikal Vasuki, P.R. *et al.* (2017) Statistical learning and auditory processing in children with music training: An ERP study. *Clin. Neurophysiol.* 128, 1270–1281
30 Growns, B. and Martire, K.A. (2020) Forensic feature-comparison expertise: Statistical learning facilitates visual comparison performance. *J. Exp. Psychol. Appl.* 26, 493–506
31 Bogaerts, L. *et al.* (2020) Integrating statistical learning into cognitive science. *J. Mem. Lang.* 115,
32 Siegelman, N. and Frost, R. (2015) Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *J. Mem. Lang.* 81, 105–120
33 Kaufman, S.B. *et al.* (2010) Implicit learning as an ability. *Cognition* 116, 321–340
34 Carroll, J.B. (1993) *Human Cognitive Abilities*, Cambridge University Press.
35 Siegelman, N. *et al.* (2017) Towards a theory of individual differences in statistical learning. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160059
36 Raviv, L. and Arnon, I. (2018) The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Dev. Sci.* 21, e12593
37 Emberson, L.L. *et al.* (2019) Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism(s). *Dev. Sci.* 22, e12847
38 Thiessen, E.D. What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. , *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372. 05-Jan-(2017) , Royal Society of London
39 Bogaerts, L. *et al.* (2021) Statistical Learning and Language Impairments: Toward More Precise Theoretical Accounts. *Perspect. Psychol. Sci.* 16, 319–337
40 Spearman, C. (1904) "General Intelligence," Objectively Determined and Measured. *Am. J. Psychol.* 15, 201
41 Siegelman, N. *et al.* (2018) Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition* 177, 198–213
42 Frost, R. *et al.* Domain generality versus modality specificity: The paradox of statistical learning. , *Trends in Cognitive Sciences*, 19. (2015) , Elsevier Ltd, 117–125
43 Christiansen, M.H. (2019) Implicit Statistical Learning: A Tale of Two Literatures. *Top. Cogn. Sci.* 11, 468–481
44 Schapiro, A.C. *et al.* (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* 22, 1622–1627
45 Schapiro, A.C. *et al.* (2017) Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philos. Trans. R. Soc. B Biol. Sci.* 372,
46 Isbilen, E.S. *et al.* (2020) Statistically Induced Chunking Recall: A Memory-Based Approach to Statistical Learning. *Cogn. Sci.* 44, 12848

47 Arnon, I. (2020) Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behav. Res. Methods* DOI: 10.3758/s13428-019-01205-5

48 Batterink, L.J. *et al.* (2015) Implicit and explicit contributions to statistical learning. *J Mem Lang* 83, 62–78

49 Ramus, F. and Ahissar, M. (2012) Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cogn. Neuropsychol.* 29, 104–122

50 Perruchet, P. and Vinter, A. (1998) PARSER: A Model for Word Segmentation. *J. Mem. Lang.* 39, 246–263

51 French, R.M. *et al.* (2011) TRACX: A Recognition-Based Connectionist Framework for Sequence Segmentation and Chunk Extraction. *Psychol. Rev.* 118, 614–636

52 Mareschal, D. and French, R.M. (2017) TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philos. Trans. R. Soc. B Biol. Sci.* 372,

53 Karuza, E.A. *et al.* (2016) Local Patterns to Global Architectures: Influences of Network Topology on Human Learning. *Trends Cogn. Sci.* 20, 629–640

54 Brysbaert, M. *et al.* (2018) The Word Frequency Effect in Word Processing: An Updated Review. *Curr. Dir. Psychol. Sci.* 27, 45–50

55 New, B. and Grainger, J. (2011) On letter frequency effects. *Acta Psychol. (Amst).* 138, 322–328

56 Cassar, M. and Treiman, R. (1997) The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *J. Educ. Psychol.* 89, 631–644

57 Chetail, F. *et al.* (2015) What can megastudies tell us about the orthographic structure of English words? *Q. J. Exp. Psychol.* 68, 1519–1540

58 Apfelbaum, K.S. *et al.* (2013) Statistical learning in reading: variability in irrelevant letters helps children learn phonics skills. *Dev. Psychol.* 49, 1348–65

59 Ševa, N. *et al.* (2009) Stressing what is important: Orthographic cues and lexical stress assignment. *J. Neurolinguistics* 22, 237–249

60 Ulicheva, A. *et al.* (2020) Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition* 195,

61 Marelli, M. *et al.* (2015) Semantic transparency in free stems: The effect of Orthography-Semantics Consistency on word recognition. *Q. J. Exp. Psychol. (Hove).* 68, 1571–1583

62 Crepaldi, D. *et al.* (2010) Morphemes in their place: Evidence for position-specific identification of suffixes. *Mem. Cogn. 2010 383* 38, 312–321

63 Ashby, J. *et al.* (2005) Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Q. J. Exp. Psychol. Sect. A Hum. Exp. Psychol.* 58, 1065–1086

64 Smith, N.J. and Levy, R. (2013) The effect of word predictability on reading time is logarithmic. *Cognition* 128, 302–319

65 Schubert, T.M. *et al.* (2020) Reading the written language environment: Learning orthographic structure from statistical regularities. *J. Mem. Lang.* 114, 104148

66 Snell, J. and Theeuwes, J. (2020) A story about statistical learning in a story: Regularities impact eye movements during book reading. *J. Mem. Lang.* 113, 104127

67 Levy, R. (2008) Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177

68 Padó, U. *et al.* (2009) A Probabilistic Model of Semantic Plausibility in Sentence Processing. *Cogn. Sci.* 33, 794–838

69 Coppola, D.M. *et al.* (1998) The distribution of oriented contours in the real world.

Proc. Natl. Acad. Sci. U. S. A. 95, 4002–4006

70 Kleffner, D.A. and Ramachandran, V.S. (1992) On the perception of shape from shading. *Percept. Psychophys.* 52, 18–36

71 Stone, J. V. *et al.* (2009) Where is the light? Bayesian perceptual priors for lighting direction. *Proc. R. Soc. B Biol. Sci.* 276, 1797–1804

72 Torralba, A. and Oliva, A. (2003) , Statistics of natural image categories. , in *Network: Computation in Neural Systems*, 14, pp. 391–412

73 Lauer, T. *et al.* (2018) The role of scene summary statistics in object recognition. *Sci. Rep.* 8, 1–12

74 Lengyel, G. *et al.* (2021) Statistically defined visual chunks engage object-based attention. *Nat. Commun.* 12, 1–12

75 Lengyel, G. *et al.* (2019) Unimodal statistical learning produces multimodal object-like representations. *Elife* 8,

76 Palmer, tephen E. (1975) The effects of contextual scenes on the identification of objects. *Mem. Cognit.* 3, 519–526

77 Võ, M.L.H. and Wolfe, J.M. (2013) Differential Electrophysiological Signatures of Semantic and Syntactic Scene Processing. *Psychol. Sci.* 24, 1816–1823

78 Võ, M.L.H. *et al.* Reading scenes: how scene grammar guides attention and aids perception in real-world environments. , *Current Opinion in Psychology*, 29. 01-Oct-(2019) , Elsevier B.V., 205–210

79 Montag, J.L. *et al.* (2015) The words children hear: Picture books and the statistics for language learning. *Psychol. Sci.* 26, 1489–1496

80 Kerz, E. *et al.* Tuning to Multiple Statistics Second Language Processing of Multiword Sequences across Registers. , *41st Annual Conference of the Cognitive Science Society*. (2019)

81 Kidd, C. *et al.* (2012) The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS One* 7, e36399

82 Lieder, I. *et al.* (2019) Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nat. Neurosci.* 22, 256–264

83 Seidenberg, M.S. and McClelland, J.L. (1989) A Distributed, Developmental Model of Word Recognition and Naming. *Psychol. Rev.* 96, 523–568

84 McCauley, S.M. and Christiansen, M.H. (2019) Language learning as language use: A cross-linguistic model of child language development. *Psychol. Rev.* 126, 1–51

85 Ahissar, M. and Hochstein, S. (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464

86 Growns, B. *et al.* (2020) The multi-faceted nature of visual statistical learning: Individual differences in learning conditional and distributional regularities across time and space. *Psychon. Bull. Rev.* 27, 1291–1299

87 Vidal, Y. *et al.* (2021) A general-purpose mechanism of visual feature association in visual word identification and beyond. *Curr. Biol.* 31, 1261-1267.e3

88 Ferrante, O. *et al.* (2018) Altering spatial priority maps via statistical learning of target selection and distractor filtering. *Cortex* 102, 67–95

89 Bulf, H. *et al.* (2011) Visual statistical learning in the newborn infant. *Cognition* 121, 127–132

90 Palmer, S.D. (2018) Statistical learning for speech segmentation: Age-related changes and underlying mechanisms. *Psychol. Aging* 33, 1035

91 Saffran, J.R. *et al.* (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52

92 Rey, A. *et al.* (2018) Regularity Extraction Across Species: Associative Learning

Mechanisms Shared by Human and Non-Human Primates. *Top. Cogn. Sci.* DOI: 10.1111/tops.12343

93 Toro, J.M. and Trobalón, J.B. Statistical computations over a speech stream in a rodent. , *Perception and Psychophysics*, 67. (2005) , Psychonomic Society Inc., 867–875

94 Menyhart, O. *et al.* (2015) Juvenile zebra finches learn the underlying structural regularities of their fathers' song. *Front. Psychol.* 0, 571

95 Aslin, R.N. (2017) Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdiscip. Rev. Cogn. Sci.* 8, e1373

96 Clerkin, E.M. *et al.* (2017) Real-world visual statistics and infants' first-learned object names. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160055

97 Lavi-Rotbain, O. and Arnon, I. (2021) Visual statistical learning is facilitated in Zipfian distributions. *Cognition* 206, 104492

98 Potter, C.E. and Lew-Williams, C. (2019) Infants' selective use of reliable cues in multidimensional language input. *Dev. Psychol.* 55, 1–8

99 Elleman, A.M. *et al.* 02-Jan-(2019) , The Role of Statistical Learning in Word Reading and Spelling Development: More Questions Than Answers. , *Scientific Studies of Reading*, 23, Routledge, 1–7

100 Bogaerts, L. *et al.* (2020) Statistical Learning and Language Impairments: Toward More Precise Theoretical Accounts. *Perspect. Psychol. Sci.* DOI: 10.1177/1745691620953082

101 Warne, R.T. and Burningham, C. (2019) Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. *Psychol. Bull.* 145, 237–272

102 Wechsler, D. Wechsler Adult Intelligence Scale--Fourth Edition (WAIS-IV). APA PsycTests. .

103 Gathercole, S.E. *et al.* (2016) How Common are WM Deficits in Children with Difficulties in Reading and Mathematics? *J. Appl. Res. Mem. Cogn.* 5, 384–394

104 Oswald, F.L. *et al.* (2014) The development of a short domain-general measure of working memory capacity. *Behav. Res. Methods* 47, 1343–1355

105 Wilhelm, O. *et al.* (2013) What is working memory capacity, and how can we measure it? *Front. Psychol.* 4, 433

106 Waris, O. *et al.* (2017) A Latent Factor Analysis of Working Memory Measures Using Large-Scale Data. *Front. Psychol.* 8, 1062

107 Christiansen, M.H. and Monaghan, P. (2016) Division of Labor in Vocabulary Structure: Insights From Corpus Analyses. *Top. Cogn. Sci.* 8, 610–624

108 Siegelman, N. *et al.* (2020) Using information-theoretic measures to characterize the structure of the writing system: the case of orthographic-phonological regularities in English. *Behav. Res. Methods* 52, 1292–1312

109 Smith, L.B. *et al.* The Developing Infant Creates a Curriculum for Statistical Learning. , *Trends in Cognitive Sciences*, 22. 01-Apr-(2018) , Elsevier Ltd, 325–336

110 Welbourne, S.R. and Ralph, M.A.L. (2005) Exploring the impact of plasticity-related recovery after brain damage in a connectionist model of single-word reading. *Cogn. Affect. Behav. Neurosci.* 5, 77–92

Outstanding questions

- What are the relevant statistical computations in different real-world perceptual and cognitive environments as encountered by learners? How does the statistical structure characterizing different domains change over time (short-term and long-term across development)?
- What are the endogenous biological factors that contribute to individual differences in sensitivity to different structural regularities? How do potential differences in genotypes interact with environmental variability to produce variation in an individual's neural mechanisms involved in the learning of different types of regularities?
- Do patients with damage to the medial temporal lobe memory system, thought of as the main neural substrate of SL, show no or strongly reduced learning of regularities in all cognitive domains?
- Does SL imply that statistics are stored as such? If so, how might this be implemented? If not, might statistical regularities instead be stored not as statistics but as cumulative weight changes in neural networks?