

The Long Road of Statistical Learning Research: Past, Present, and Future

Blair C. Armstrong^{1,2}, Ram Frost^{3,4,2}, and Morten H. Christiansen^{5,4,6}

1. Department of Psychology and Centre for French & Linguistics, University of Toronto Scarborough, Toronto, Canada
2. BCBL. Basque Center on Cognition, Brain, and Language, San Sebastian, Spain
3. Department of Psychology, The Hebrew University, Jerusalem, Israel
4. Haskins Laboratories, New Haven, Connecticut, USA
5. Cornell University, Ithaca, New York, USA
6. Aarhus University, Aarhus, Denmark.

Almost all types of learning involve, to some degree, the ability to encode regularities across time and space. Although statistical learning (SL) research initially focused on offering a viable alternative to rule-based grammars and specialized mechanisms for word learning (e.g., [1,2]), the processing of regularities embedded in sensory input extends well beyond language. SL, therefore, was taken to offer a comprehensive theory of information processing, holding the promise of advancing knowledge across various domains of cognition including visual and auditory perception, multi-modal integration, motor learning, segmentation, categorization, and generalization, to name a few.

On the theoretical level, SL has had substantial impact on the cognitive sciences, viewed as a powerful domain-general learning mechanism and often invoked to argue against nativist or domain-specific accounts of language and cognition. However, a retrospective view of two decades of SL research reveals a substantial gulf between the wide-reaching promise of SL as a theoretical construct and the actual empirical work that would support it. Since the foundational work of Reber [1], and Saffran and colleagues [2], research on SL has primarily focused on providing a proof of concept of the human ability to perceive and learn the distributional properties of visual or auditory input. This has been achieved by monitoring participants' performance in laboratory settings with a strikingly narrow set of tasks: In one paradigm, sequences of stimuli generated by some miniature artificial grammar are presented for familiarization, and then subsequent correct classification of novel grammatical and ungrammatical sequences attests for learning (i.e., Artificial Grammar Learning-AGL). In another paradigm, regularities are embedded in a sensory input (typically visual or auditory), and learning of these regularities (i.e., co-occurrence of elements, their transitional probabilities, etc.) during a relatively brief familiarization phase, usually on the order of minutes, is assessed in a subsequent test phase. Extensive research using this approach has indeed provided us with detailed information regarding performance profiles in this particular set of artificial laboratory tasks. We know, for example, that infants are able to segment artificial speech on the basis of the distributional properties of the embedded elements [2], that newborns, like adults, display remarkable sensitivity to the co-occurrence of items in a continuous stream (e.g., [3]), that this sensitivity is displayed across sensory modalities, (visual: e.g., [4–6]; auditory: e.g., [7]; tactile: e.g., [8]), for verbal as well as non-verbal stimuli (e.g., [9]), that sensitivity extends to both adjacent (e.g., [10]) and nonadjacent contingencies (e.g., [11,12]), and that learning does not require overt attention (e.g., [13]), nor explicit memory (e.g., [14]).

Although these findings represent considerable progress within the field, much of SL research has focused on relatively restricted sets of issues, often related to the types of regularities extracted from the input, the possible cues that modulate extraction, the necessary conditions for determining above chance performance in terms of rate of presentation, complexity of embedded stimuli, their similarity to previously established representations, etc. At large, the “Zeitgeist” of this research implicitly regards SL as an independent computational mechanism, akin to a device, that is specialized for extracting the distributional properties of the sensory input, where research should focus on determining its operational scope. This has naturally led to investigating SL in isolation as a separate ability from other systems. A corollary of this approach is that advancing knowledge of SL would be achieved by mapping the set of constraints on its operation.

Is this all there is to SL? From a theoretical perspective, would the full description of constraints on SL reveal its exact role across the full breadth of cognitive systems? Should the field continue along the same trajectory of the previous two decades for the next two decades?

We take it as self-evident that a full understanding of SL is not tantamount to detailing performance of children and adults in registering the structural similarity of grammatical sequences in an AGL paradigm, and/or extracting the transitional probabilities between syllables or meaningless shapes in a stream. A powerful theory of SL as a domain-general mechanism—or set of mechanisms—requires a wider perspective. If SL is a cornerstone of cognition in general, then a comprehensive theory will have to integrate and constrain SL by what we know about key cognitive faculties, such as perception, attention, and memory, what we know about their development throughout the life span or through evolution, and what we know about their neurobiological and computational instantiation.

The main goal of this special issue is therefore to place SL in its rightful role as fundamental part of learning and development across cognition. It aims to foster a transition from studying SL in isolation to studying it as an integral part of different cognitive systems. This would involve, for instance, tying early statistical sensitivities in infants to phonological structure, to broader theories of language emergence, constrained by what we know about memory, attention, and their developmental trajectories. From learning basic regularities in the visual modality, to theories of perception, visual cognition, scene segmentation, object recognition, and what we know about the neural systems that support these functions. From treating individual variation in statistical learning as noise, to emphasizing the functional significance of such variability, in relation to what we know about learning and communication abilities and disabilities. In sum, this special issue offers a way forward to understanding how SL subserves cognition.

Through this approach, what has traditionally been termed “learning” may usefully be construed as SL operating at a large scale, in coordination with the core mechanisms of other cognitive systems and abilities. This approach has the promise to offer not only a better understanding of SL, but also a better understanding of the cognitive systems it operates within. This forward-looking foundational viewpoint, however, requires stressing a different set of theoretical questions for the SL research community, allocating a central role for an interdisciplinary program that leverages the unique insights from different disciplines and methodologies. Fortunately, the seeds of this new perspective has already been sown and the time is ripe to bring these into an integrated whole.

The diverse papers of the present volume, in one way or another, exemplify this direction towards the new frontiers of SL research. Each one of them identifies fundamental questions along the lines outlined above, and offers a blueprint for addressing them. Together, the papers thus provide an exciting picture of what the future may hold for a more integrated and interdisciplinary approach to SL, viewed within its rightful place in cognition.

The volume was put together to provide a broad glimpse of the new frontiers, building from a low-level neurobiological understanding of SL and its neurocomputational instantiation, to a scaffolded

consideration of how these mechanisms connect with higher-level key cognitive systems. This understanding is achieved by drawing upon insights from evolution, development, and computational constraints on processing. The volume thus begins with Hasson's **(this issue)** critical review of the basic neural building blocks for detecting regularities or their absence. Hasson outlines areas of convergence and divergence between models of SL and models focused on the coding of uncertainty. He then derives desiderata for future neurobiological work in SL. This review sets the stage for understanding the possible neurobiological constraints for any theory of SL.

Next, Schapiro and her colleagues **(this issue)** provide a higher-level perspective on the important role of the hippocampus in extracting regularities from different sensory input streams. Through a series of neurocomputational simulations, they reveal how the hippocampal system can resolve an apparent paradox created by the need to encode distinct memories for particular events, on the one hand, and rapidly extract regularities among events, on the other. Drawing upon insights from computational modeling, their work clearly illustrates how a more integrated understanding of SL and complementary memory systems can better define the interplay between the hippocampus and the neocortex.

Gomez **(this issue)** addresses the critical gap between the rapid encoding of regularities in brief laboratory experiments, and what is required for the permanent retention of knowledge in the domain of language. This work is informed by developmental insights into the different memory systems that support initial encoding versus subsequent consolidation. Gomez, thus, specifically targets the problem of ecological validity in SL research. Whereas typical learning in the laboratory proceeds at an exceedingly rapid pace, language acquisition during infancy is known to be slow in relative terms. This discrepancy cannot be resolved without considering the constraints of the different memory systems implicated in learning, as well as their developmental trajectories. In focusing on these considerations, we gain a better understanding of what underlies the observed differences between adult and infant SL.

In a related vein, Arciuli **(this issue)**, discusses SL in the context of age-related changes and neurodevelopmental accounts of typical and impaired communication abilities, such as autism spectrum disorder. This work touches on a fundamental question: is SL a unitary mechanism or a composite ability that relies upon the close coordination of a number of separate cognitive systems such as perception, attention, and memory? Arciuli provides substantial evidence for considering SL as a multi-faceted ability, where individual differences in SL performance should be understood in terms of variability in the efficacy and relative maturation of these respective systems. This approach of deriving meaning from individual variability, as opposed to considering it as noise, not only explicates contrasting findings in SL research, but also offers a theoretical perspective for tying SL to a range of disorders.

Generalizing this perspective, Siegelman and colleagues **(this issue)** offer a formal conceptual framework for defining SL as a componential ability. By considering a range of findings from group and individual level studies, they outline potential dimensions of SL, and point to the major methodological consequences that this has for tying individual differences in SL to specific cognitive functions. This

framework offers clear blueprints for structuring future research, requiring researchers to specify a priori how and why specific SL tasks would engage particular cognitive systems. As a corollary, they explicate how some learning measures are better suited for probing certain dimensions of SL.

Of key importance to understanding SL as embedded in our broader cognitive abilities is determining the nature of input available for such learning. Clerkin and colleagues (**this issue**) adopt an ecologically-motivated approach to the development of early word learning, asking what the visual environment looks like during the first year of an infant's life. Although the visual input is very cluttered with many objects in view, the frequency distribution of particular object categories follows a power-law distribution: a very small set of objects occur repeatedly. The authors note that this frequency pattern is quite different from the uniform distribution that is typically used in SL experiments (typically under the heading of "cross-situational learning"). Nonetheless, the right-skewed distribution of objects in the child's visual field may be crucial for word learning, as suggested by the fact that the names for these visual object categories belong to the first words that are learned. This paper thus underscores the importance of incorporating ecological constraints into both experimental work and theoretical considerations about SL.

Although often implicit in the discussion of SL results, it is clear that the outcome of SL is not simply a representation of the statistics of the input. Rather, the cognitive system uses sensitivity to distributional patterns to shape its expectations and behavioral responses in an adaptive way, constrained by pre-existing biases in that system. The study by Feher and colleagues (**this issue**) provides an innovative test of this perspective in the context of self-tutored bird song learning. They record the songs of juvenile zebra finches placed in isolation and play it back to them moments later. These birds normally learn from adult males that have established categories of song elements. However, the juvenile birds themselves start out with a broadly distributed signal. Yet, the self-tutored birds quickly developed categorical signals at the same rate as birds raised with an adult tutor. These results demonstrate that SL does not simply involve recording distributional patterns, but rather reflects an active process of learning, shaped by existing perceptual and cognitive biases.

The empirical work of Shimizu and colleagues (**this issue**) extends SL research on several important fronts. First, it focuses on visuo-motor SL, thereby probing the link between perception and action. Second it shifts away from classical SL brain areas associated with SL, investigating the relatively understudied role of the cerebellum. Third, rather than using the typical design where neural activity is indirectly driven by the experimental manipulation of the input, Shimizu and colleagues manipulate neural activity itself via transcranial direct current stimulation (tDCS) to probe for commensurate changes in performance. This work not only reveals the critical role of the cerebellum in learning and generalizing regularities in the motor domain, but also raises intriguing questions regarding its role in SL across a range of domains.

By complementing neurocomputational simulations, computational modeling at the cognitive level can provide additional insights into the possible mechanisms underlying SL. Thiessen (**this issue**)

discusses recent modeling efforts situating SL within a basic memory framework. He proposes that SL may be accommodated by two distinct kinds of computational mechanisms: one that relies on chunk-based memory processes to store exemplars, and another that captures central tendencies in distributional input by integrating over prior exemplars stored in memory. A key feature of this computational account is that statistics are stored in any form—the effects of exposure to statistical patterns are instead reflected implicitly in the system’s memory traces. The paper thus provides a parsimonious way in which to understand statistical learning in the context of exemplar memory.

Mareschal and French (**this issue**) address a related question that is currently the subject of heated debate: Does the SL mechanism target the transitional probabilities between elements in the input signal, or is it simply designed to group together co-occurring elements into memory chunks? Using a variant of a connectionist autoencoder model, they show how gradual chunking of co-occurring elements within an input can potentially explain effects associated with backward and forward transitional probability learning, as well as preference for whole-words over part-words which occur with equal probability in the stream. They also show that such a model is developmentally plausible by predicting the established improvement of SL with age. This work demonstrates the critical role that explicit computational theories of SL can have in reconciling apparently discrepant findings and theoretical accounts, offering a more parsimonious explanation of a range of effects without sacrificing descriptive adequacy.

Using the domain of sentence processing as an anchor, Altmann (**this issue**), in a sense, turns SL on its head. After describing how repeated encounters with regularities in the input are the basis for generalization and abstraction in the form of semantic knowledge, he reverse engineers this process. In so doing, Altmann offers a possible account of how semantic types acquired through SL underpin the ability to process and generate novel episodic tokens. By pointing to the reciprocal relationship between comprehension and generation of sentence meaning, we gain novel insight regarding the tight and intertwined relationship between SL, semantic memory and the comprehension of novel episodes.

The volume closes with an evolutionary perspective on the interaction between SL, language learning, and the evolution of linguistic variation. Smith and colleagues (**this issue**) put forward the hypothesis that the relatively low prevalence of unpredictable variation in natural languages could be attributed to children’s SL biases against such variations, along with processes related to language transmission over multiple generations. To substantiate this idea, they develop a Bayesian model of language learning and language transmission, and compare its performance against that of humans in an artificial language learning task. The data generated by this approach cast light on the rich and complex relationships between the constraints imposed by SL and the evolution of linguistic structure. The emergent perspective considers SL not simply in terms of individuals extracting the regularities of the environment. Rather, there is a two way street between human created “environments” such as language and SL learning mechanisms.

Collectively the series of papers reveal that the tide is beginning to turn in the SL community, where the accumulated evidence regarding processing regularities in the environment is now taken to shape and constrain theories of cognitive systems. The outcome of SL is not simply a veridical internal representation of the regularities of the environment. Rather it is a product of the interaction between environmental statistics, the computational principles of the cognitive systems in which learning takes place, and pre-existing biases, either from prior exposure to other input patterns or architectural constraints. The discussions going forward will consequently inevitably shift from dialog within community to cross-disciplinary interactions between communities. This would gradually narrow the gulf between the original promise of SL as a theoretical construct, and its actual implementation and impact on theories of language, vision, audition, memory, social behavior, etc.

Such a change of perspective, however, brings a new set of challenges and questions to center stage. For example, how does encoding uncertainty in low-level biology (**Hasson, this issue**) relate to uncertainty in high-level domains such as visual word recognition, or sentence comprehension? How would the hippocampal system capable of encoding both statistical regularities and distinct episodes (**Schapiro et al., this issue**) relate to the representation of semantic types and episodic tokens (**Altmann this issue**)? Would the basic computational mechanisms tested in small artificial language experiments (**Thiessen this issue; Mareschal and French, this issue**) scale up to dealing with the real-world input, such as natural language (**Clerkin and colleagues, this issue**)? This small sample of questions highlights the new frontiers of SL research for the road ahead.

Acknowledgment: This paper was supported by the Israel Science Foundation (Grant 217/14 awarded to Ram Frost), by the National Institute of Child Health and Human Development (RO1 HD 067364 awarded to Ken Pugh and Ram Frost, PO1-HD 01994 awarded to Haskins Laboratories), and by the European Research Council (project ERC-ADG-692502).

References

1. Reber, A. S. 1967 Implicit learning of artificial grammars. *J. Verbal Learning Verbal Behav.* **6**, 855–863. (doi:10.1016/S0022-5371(67)80149-X)
2. Saffran, J. R., Aslin, R. N. & Newport, E. L. 1996 Statistical Learning by 8-Month-Old Infants. *Science (80-.).* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)
3. Bulf, H., Johnson, S. P. & Valenza, E. 2011 Visual statistical learning in the newborn infant. *Cognition* **121**, 127–132. (doi:10.1016/j.cognition.2011.06.010)
4. Fiser, J. & Aslin, R. N. 2001 Unsupervised statistical learning of higher-order spatial structures

- from visual scenes. *Psychol. Sci.* **12**, 499–504. (doi:10.1111/1467-9280.00392)
5. Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. 2002 Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* **83**, B35–B42. (doi:10.1016/S0010-0277(02)00004-5)
 6. Turk-Browne, N. B., Junge, J. A. & Scholl, B. J. 2005 The automaticity of visual statistical learning. *J. Exp. Psychol.* **134**, 552–564. (doi:10.1037/0096-3445.134.4.552)
 7. 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. *Psychol. Sci.* **8**, 101–105. (doi:10.1111/j.1467-9280.1997.tb00690.x)
 8. Conway, C. M. & Christiansen, M. H. 2005 Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 24–39. (doi:10.1037/0278-7393.31.1.24)
 9. Gebhart, A. L., Newport, E. L. & Aslin, R. N. 2009 Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* **16**, 486–490. (doi:10.3758/PBR.16.3.486)
 10. Endress, A. D. & Mehler, J. 2009 The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *J. Mem. Lang.* **60**, 351–367. (doi:10.1016/j.jml.2008.10.003)
 11. Gómez, R. L. 2002 Variability and detection of invariant structure. *Psychol. Sci.* **13**, 431–436. (doi:10.1111/1467-9280.00476)
 12. Newport, E. L. & Aslin, R. N. 2004 Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* **48**, 127–162. (doi:10.1016/S0010-0285(03)00128-2)
 13. Evans, J., Saffran, J. & Robe-Torres, K. 2009 Statistical learning in children with Specific Language Impairment. *J. Speech, Lang. Hear. Res.* **52**, 321–335.
 14. 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. *Psychol. Sci.* **3**, 172–179. (doi:10.1111/j.1467-9280.1992.tb00021.x)