



## Introduction

**Cite this article:** Armstrong BC, Frost R, Christiansen MH. 2017 The long road of statistical learning research: past, present and future. *Phil. Trans. R. Soc. B* **372**: 20160047.

<http://dx.doi.org/10.1098/rstb.2016.0047>

Accepted: 12 October 2016

One contribution of 13 to a theme issue 'New frontiers for statistical learning in the cognitive sciences'.

### Subject Areas:

behaviour, cognition, neuroscience

### Author for correspondence:

Blair C. Armstrong

e-mail: [blair.armstrong@utoronto.ca](mailto:blair.armstrong@utoronto.ca)

# The long road of statistical learning research: past, present and future

Blair C. Armstrong<sup>1,2,3</sup>, Ram Frost<sup>3,4,5</sup> and Morten H. Christiansen<sup>5,6,7</sup>

<sup>1</sup>Department of Psychology, and <sup>2</sup>Centre for French and Linguistics, University of Toronto Scarborough, Toronto, Canada

<sup>3</sup>BCBL, Basque Center on Cognition, Brain, and Language, San Sebastian, Spain

<sup>4</sup>Department of Psychology, The Hebrew University, Jerusalem, Israel

<sup>5</sup>Haskins Laboratories, New Haven, CT, USA

<sup>6</sup>Department of Psychology, Cornell University, Ithaca, NY, USA

<sup>7</sup>Interacting Minds Centre and School of Communication and Culture, Aarhus University, Aarhus, Denmark

BCA, 0000-0002-6345-0389

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, multimodal 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. Following the foundational work of Reber [1], and Saffran *et al.* [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 (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 and 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 lifespan 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 SL 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 programme 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 [15] 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 *et al.* [16] 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 modelling, 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 [17] 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 [18] 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 multifaceted 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 *et al.* [19] 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 *et al.* [20] 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 behavioural responses in an adaptive way, constrained by preexisting biases in that system. The study by Feher *et al.* [21] 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 *et al.* [22] extends SL research on several important fronts. First, it focuses on visuomotor 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 *et al.* 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 modelling at the cognitive level can provide additional insights into the possible mechanisms underlying SL. Thiessen [23] discusses recent modelling 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 the effects of exposure to statistical patterns are reflected implicitly in the system's memory traces. The paper thus provides a parsimonious way in which to understand SL in the context of exemplar memory.

Mareschal & French [24] 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 [25], 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 volumes close with an evolutionary perspective on the interaction between SL, language learning and the evolution of linguistic variation. Smith *et al.* [26] 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 preexisting 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 behaviour and so on.

Such a change of perspective, however, brings a new set of challenges and questions to centre stage. For example, how does encoding uncertainty in low-level biology [15] 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 [16] relate to the representation of semantic types and episodic tokens [25]? Would the basic computational mechanisms tested in small artificial language experiments [23,24] scale up to dealing with the real-world input, such as natural language [20]? This small sample of questions highlights the new frontiers of SL research for the road ahead.

**Competing interests.** We declare we have no competing interests.

**Funding.** This paper was supported by the Israel Science Foundation (grant no. 217/14 awarded to R.F.), by the National Institute of Child Health and Human Development (RO1 HD 067364 awarded to Ken Pugh and R.F., PO1-HD 01994 awarded to Haskins Laboratories) and by the European Research Council (project ERC-ADG-692502 awarded to R.F.).

## References

1. Reber AS. 1967 Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* **6**, 855–863. (doi:10.1016/S0022-5371(67)80149-X)
2. Saffran JR, Aslin RN, Newport EL. 1996 Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)
3. Bulf H, Johnson SP, 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 RN. 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 NZ, Slemmer JA, Johnson SP. 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 NB, Junge JA, Scholl BJ. 2005 The automaticity of visual statistical learning. *J. Exp. Psychol.* **134**, 552–564. (doi:10.1037/0096-3445.134.4.552)
7. Saffran JR, Newport EL, Aslin RN, Tunick RA, 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 CM, Christiansen MH. 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 AL, Newport EL, Aslin RN. 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 AD, 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 RL. 2002 Variability and detection of invariant structure. *Psychol. Sci.* **13**, 431–436. (doi:10.1111/1467-9280.00476)
12. Newport EL, Aslin RN. 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. (doi:10.1044/1092-4388(2009)07-0189)
14. Knowlton BJ, Ramus SJ, Squire LR. 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)
15. Hasson U. 2017 The neurobiology of uncertainty: implications for statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160048. (doi:10.1098/rstb.2016.0048)
16. Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA. 2017 Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160049. (doi:10.1098/rstb.2016.0049)
17. Gómez RL. 2017 Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Phil. Trans. R. Soc. B* **372**, 20160054. (doi:10.1098/rstb.2016.0054)
18. Arciuli J. 2017 The multi-component nature of statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160058. (doi:10.1098/rstb.2016.0058)
19. Siegelman N, Bogaerts L, Christiansen MH, Frost R. 2017 Towards a theory of individual differences in statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160059. (doi:10.1098/rstb.2016.0059)
20. Clerkin EM, Hart E, Rehag JM, Yu C, Smith LB. 2017 Real-world visual statistics and infants' first-learned object names. *Phil. Trans. R. Soc. B* **372**, 20160055. (doi:10.1098/rstb.2016.0055)
21. Fehér O, Ljubičić I, Suzuki K, Okanoya K, Tchernichovski O. 2017 Statistical learning in songbirds: from self-tutoring to song culture. *Phil. Trans. R. Soc. B* **372**, 20160053. (doi:10.1098/rstb.2016.0053)
22. Shimizu RE, Wu AD, Samra JK, Knowlton BJ. 2017 The impact of cerebellar transcranial direct current stimulation (tDCS) on learning fine-motor sequences. *Phil. Trans. R. Soc. B* **372**, 20160050. (doi:10.1098/rstb.2016.0050)
23. Thiessen ED. 2017 What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Phil. Trans. R. Soc. B* **372**, 20160056. (doi:10.1098/rstb.2016.0056)
24. Mareschal D, French RM. 2017 TRACK2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160057. (doi:10.1098/rstb.2016.0057)
25. Altmann GTM. 2017 Abstraction and generalization in statistical learning: implications for the relationship between semantic types and episodic tokens. *Phil. Trans. R. Soc. B* **372**, 20160060. (doi:10.1098/rstb.2016.0060)
26. Smith K, Perfors A, Fehér O, Samara A, Swoboda K, Wonnacott E. 2017 Language learning, language use and the evolution of linguistic variation. *Phil. Trans. R. Soc. B* **372**, 20160051. (doi:10.1098/rstb.2016.0051)