This article is part of the topic "Two approaches, one phenomenon: Aligning implicit learning and statistical learning," Padraic Monaghan and Patrick Rebuschat (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview

Implicit Statistical Learning: A Tale of Two Literatures

Morten H. Christiansen[a,b,c]

[a]*Department of Psychology, Cornell University*
[b]*Interacting Minds Centre and School of Communication and Culture, Aarhus University*
[c]*Haskins Laboratories*

Abstract

Implicit learning and statistical learning are two contemporary approaches to the long-standing question in psychology and cognitive science of how organisms pick up on patterned regularities in their environment. Although both approaches focus on the learner's ability to use distributional properties to discover patterns in the input, the relevant research has largely been published in separate literatures and with surprisingly little cross-pollination between them. This has resulted in apparently opposing perspectives on the computations involved in learning, pitting chunk-based learning against probabilistic learning. In this paper, I trace the nearly century-long historical pedigree of the two approaches to learning and argue for their integration under the heading of "implicit statistical learning." Building on basic insights from the memory literature, I sketch a framework for statistically based chunking that aims to provide a unified basis for understanding implicit statistical learning.

Keywords: Statistical learning; Implicit learning; Chunking; Serial recall; Memory; Nonword repetition

1. Introduction

The popularity of statistical learning research has grown rapidly across the past two decades (see e.g., Frost, Armstrong, Siegelman, & Christiansen, 2015, for a review)—yet

---

Correspondence should be sent to Morten H. Christiansen, Department of Psychology, 228 Uris Hall, Cornell University, Ithaca, NY 14853. E-mail: christiansen@cornell.edu

surprisingly few researchers are aware that the study of how learners can implicitly use distributional properties of the input to discover patterned regularities in their environment goes back nearly a century. Consequently, many of the insights into such learning discovered prior to the seminal study by Saffran, Aslin, and Newport (1996) have gone unnoticed in the statistical learning literature. Indeed, a substantial proportion of the continuation of such early work tends to be published in a separate literature on *implicit learning* (see Perruchet, 2008; Rebuschat & Monaghan, this issue, for reviews). However, if statistical learning is to fulfill its promise and explain key aspects of the development of cognitive abilities—from language to social cognition—we need to integrate lessons from its long historical pedigree into our current thinking. This paper, therefore, aims to situate what I will refer to as "implicit statistical learning" (ISL) within its broader historical context, pointing to important parallels with the learning and memory literature, and proposing a framework for understanding its role in language.

1.1. A century of implicit statistical learning

Erwin A. Esper may be one of the most visionary pioneers within the study of ISL that few people have ever heard of. Inspired by an interest in language change, Esper (1925) published one of the very first artificial language learning studies,[1] in which he, as a behaviorist, was interested in how statistical patterns may give rise to grammatical categories (something of keen interest to current ISL researchers, for example, Frost, Monaghan, & Christiansen, 2016). As another first, Esper (1933) documented individual differences in ISL, which has only recently come to the forefront of contemporary ISL research (e.g., Siegelman, Bogaerts, Christiansen, & Frost, 2017). Finally, Esper (1966) conducted the first study of how language change can be explored via cultural transmission of an artificial language across multiple "generations" of learners (a key paradigm in recent language evolution research; Kirby, Cornish, & Smith, 2008). Esper's work was sidelined following the cognitive revolution, but given the current interest in distributional learning, and a reemphasis on the importance of linguistic experience in language acquisition and use, it seems timely to acknowledge his important contributions to ISL.

Importantly, though the idea of using artificial language learning to study language acquisition did not disappear with behaviorism, but became a part of the methodological toolkit of cognitive psychology—however, explanations of learning shifted away from statistical patterns to rules. One of the first examples of such work was George Miller's Project Grammarama, which he began in 1957 (described in Miller, 1967). Whereas, Miller worked within the general framework of formal language theory, Arthur Reber did similar work with artificial grammars but with the aim of understanding implicit learning (Reber, 1967). As subsequent research tended to follow one of these two different conceptions of ISL at the expense of the other, two separate literatures emerged over time. Research on implicit learning tended to involve the artificial grammar learning (AGL) methodology developed by Reber, but also incorporated novel tasks such as probability learning (e.g., Berry & Broadbent, 1984), in which participants learn to control complex systems (such as a manufacturing plant), and serial-reaction time (SRT) studies (Nissen

& Bullemer, 1987), where participants' implicit knowledge of repeated visual sequences is evidenced by their response time patterns. This work focused on the mechanisms involved in implicit learning (including the role of consciousness and the explicit vs. implicit nature of exposure) and tended to fall within the purview of cognitive psychology, appearing in journals such as *Journal of Experimental Psychology: Learning, Memory and Cognition, Journal of Experimental Psychology: General,* and *Quarterly Journal of Experimental Psychology*.

By contrast, the work following Miller's language-oriented approach focused primarily on the kinds of structure that could be learned, given linguistic and psycholinguistic considerations. This research is represented, for example, by studies investigating artificial language learning in children (Braine, 1966) and in the context of integrating multiple cues (e.g., from prosody: Morgan, Meier, & Newport, 1987). This work tended to be published in journals like *Cognitive Psychology, Journal of Memory and Language,* and *Cognition.* Consequently, the two approaches to ISL remained largely isolated from one another until quite recently, even after the introduction of the term "statistical learning" by Saffran, Aslin, and Newport (1996). Perruchet and Pacton (2006) noted the similarities between the two approaches, and Conway and Christiansen (2006) offered the term "implicit statistical learning" as a unifying term for the parallel work conducted within the two traditions (see also Perruchet, this issue). Similarly, the work presented in this special topic and in a recent special issue of *Philosophical Transactions of the Royal Society B: Biological Sciences* (Armstrong, Frost, & Christiansen, 2017) constitutes the most recent examples of this move toward a more unified approach to ISL, seeking to integrate the previously separate literatures.

A substantial challenge, however, for much current and past ISL work is the methodological disconnect between the very nature of this type of learning and the way in which its effects are measured. In a typical experiment, participants are provided with repeated exposure to patterned stimuli (such as consonant strings from an artificial grammar or recurrent syllable triplets). As a first approximation, learning may be construed as involving gradual changes to the processing of the input, resulting in sensitivity to its inherent regularities. Although selective attention to the input appears to boost learning (e.g., Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne, Jungé, & Scholl, 2005), what is learned in ISL studies is largely implicit (i.e., mostly outside conscious awareness).[2] Yet, to measure the effects of learning, participants are typically asked to reflect on what they have learned and make explicit decisions about test stimuli (using on their "gut feeling"). In other words, participants are asked to channel the primary *processing-based* effects of learning through what we might call a "consciousness filter" to produce a secondary *reflection-based* overt response. Because this consciousness filter is likely to be differently tuned from person to person, it adds a considerable amount of noise to the secondary response data. This additional noise makes reflection-based tests problematic as measures of individual differences in ISL (Siegelman et al., 2017). Reflection-based measures also pose problems for group-level studies because they are not measuring what ISL researchers typically think they are measuring (i.e., processing-based effects of learning). Indeed, reflection-based tests are likely to underestimate the effects of learning in ISL studies

relative to processing-based measures (e.g., Batterink, Reber, Neville, & Paller, 2015; Vuong, Meyer, & Christiansen, 2016).

Although most ISL studies with adult participants still use reflection-based tests of learning, there are some notable exceptions that employ processing-based measures instead.[3] The majority of these studies relies on reaction times (RTs) to measure the effects of learning on processing in a speeded target detection task. A classic example is the SRT task (Nissen & Bullemer, 1987), in which participants respond serially to one of several alternating targets as quickly as possible, with decreases in RTs resulting from exposure to patterned regularities (see Misyak, Christiansen, & Tomblin, 2010, for a cross-modal extension to AGL). A more recent variation is the speeded detection of a single target within a longer sequence—for example, a particular shape within a string of visual shapes (Turk-Browne et al., 2005) or a specific syllable within a stream of syllables (e.g., Batterink et al., 2015; Franco, Eberlen, Destrebecqz, Cleeremans, & Bertels, 2015). Whereas, the classic SRT task integrates exposure and test, the more recent innovation is meant to replace the standard reflection-based test following the training phase. As another possible replacement, I focus here on a basic measure from the memory literature as a novel way to capture processing-based effects of learning.

In this paper, I propose to reconsider ISL in the context of basic mechanisms for learning and memory. I first briefly review similarities between ISL and immediate memory processes, suggesting we can employ the classic serial recall memory task to measure statistical learning implicitly. I then discuss how this approach further allows us to study ISL "in the wild" as statistically based chunking, demonstrating how the learning of real-world statistics can facilitate language processing.

2. Memory and implicit statistical learning

When thinking about ISL tasks, it is worth considering what it is that we are trying to measure. Typically, when approaching some psychological phenomenon, a researcher employs a task to tap into that aspect of cognitive behavior, often assuming that there is a dedicated neural or cognitive mechanism that corresponds to the task in question. For example, many ISL researchers talk about a "statistical learning mechanism" (e.g., Kirkham, Slemmer, & Johnson, 2002). However, the inference from a given behavioral task to a specific brain mechanism may in many cases be invalid. Consider, for example, the classic lexical decision task, in which participants observe letter strings on a screen and have to decide, as quickly as possible, whether the stimulus is a word or not. Lexical decision has been used in hundreds, if not thousands of studies to gain insights into the structure of the vocabulary (e.g., Plaut, 1997), the organization of semantic memory (e.g., Meyer & Schvaneveldt, 1971), and even social cognition (e.g., Wittenbrink, Judd, & Park, 1997). Although this task has proven to be incredibly useful for psychology, it is also clear that there is not a dedicated mechanism for lexical decision. Instead, this task taps into orthographic, phonological, semantic, and other components of our reading and language processing system to produce a response (Price & Devlin, 2011).

In a similar vein, I suggest that ISL tasks do not tap into a dedicated statistical learning mechanism but, rather, that responses in such tasks depend on the recruitment of more basic systems for learning and memory, with the specific neural network in question varying with differences in task demands. As a first approximation, we can construe ISL as a proxy for the brain's sensitivity to real-world statistics as reflected by the input to the organism. Rather than a dedicated statistical learning mechanism, sensitivity to statistical patterns is mediated by basic learning and memory processes, such as chunking (e.g., Miller, 1956).

Chunking, as a basic cognitive skill, has recently been suggested to provide a way to overcome the *Now-or-Never Bottleneck* that results from the fleeting nature of both the input and our memory for it (Christiansen & Chater, 2016b). For example, our language system faces a formidable three-pronged challenge: (a) the speech signal is highly transient (50–100 ms, Remez et al., 2010); (b) normal speech is fast (about 150 words per minute, Studdert-Kennedy, 1986); and (c) memory for auditory sequences is very limited (between 4 ± 1, Cowan, 2000; and 7 ± 2, Miller, 1956). To get an intuitive feel for how chunking can facilitate processing, compare the difficulty of recalling the following string of random letters from memory, "l h p a e i c p r a," with the same 10-letter string reorganized into "a p p l e c h a i r." Because we automatically chunk the letters in the second string into the two words *apple* and *chair*, their component letters are easy to recall.

Christiansen and Chater (2016b) argue that to cope with the Now-or-Never Bottleneck, we learn through everyday language exposure to rapidly compress and recode the input into chunks, which are immediately passed to a higher level of linguistic representation. The chunks at this higher level are then themselves subject to the same *Chunk-and-Pass* procedure, resulting in progressively larger chunks of increasing linguistic abstraction. Crucially, given that chunks recode increasingly larger stretches of input from lower levels of representation, the chunking process enables input to be maintained over larger and larger temporal windows. It is this repeated chunking of lower level information that makes it possible for the cognitive system to deal with the continuous deluge of input, which, if not recoded, is rapidly lost. When it comes to the production of action sequences (including language), the chunking process roughly goes in the opposite direction, from higher level chunks all the way down to motor commands.

Importantly, the chunking processes appear to be largely statistically based (at least in relation to language; McCauley & Christiansen, 2015), suggesting that ISL may be construed as statistically based chunking. Indeed, a considerable portion of the implicit learning literature has been dedicated to the role of chunk-based (or "fragment"-based) information in such learning (see Perruchet & Pacton, 2006, for a review). For example, studies have found that the so-called chunk strength—the relative frequency of two- and three-element subsequences—affects participants' ability to distinguish legal from illegal test items in an AGL task (e.g., Knowlton & Squire, 1994). Effects of chunking strategies have also been observed in statistical learning studies (e.g., Slone & Johnson, 2015; see Perruchet, this issue, for a review), and a purely chunk-based computational model, PARSER (Perruchet & Vinter, 1998), is able to simulate data from the Saffran, Newport, and Aslin (1996) ISL study. More recently, McCauley and Christiansen (2011) developed a

computational model, the Chunk-Based Learner, which uses statistical information to discover multiword chunks to facilitate language processing. Thus, there are already both behavioral and computational modeling results that underscore the potential role of chunking in ISL.

The similarities between ISL and chunking do not end here: Over the past decade, several studies have revealed important modality differences in ISL (see Frost et al., 2015, for a review) that can also be observed in memory performance (for which chunking is crucial; Miller, 1956). For example, ISL is superior in the auditory modality compared to the visual modality (Conway & Christiansen, 2005, 2009), likely reflecting a more basic auditory superiority effect in memory as measured by serial recall (Drewnowski & Murdock, 1980). Similarly, modality differences in primacy (visual advantage) and recency (auditory advantage) effects have been observed in an ISL task (Conway & Christiansen, 2009), on par with differences observed in basic memory recall (e.g., Beaman, 2002). Modality differences in ISL also emerge for different rates of presentation (Emberson, Conway, & Christiansen, 2011), favoring audition at fast rates and vision at slow rates—again in line with observations seen for memory performance (Collier & Logan, 2000). Finally, evidence from both ISL (Conway & Christiansen, 2006; Siegelman & Frost, 2015) and memory tasks indicates that there are separate mechanisms for auditory and visual input (e.g., Baddeley & Hitch, 1974). Together, these results reveal identical patterns of modality-specific effects across memory and ISL, suggesting that ISL may be supported by basic memory mechanisms, from which statistical learning inherits its modality constraints.

Another point of contact between ISL and memory can be found in terms of individual differences. As noted earlier, Esper (1933) was the first to document individual variation in ISL (see Siegelman et al., 2017, for a review of recent work). There are significant differences in ISL that provide key insights into the nature of ISL and its relationship to other aspects of cognition (Frost et al., 2015). Individual differences in ISL correlate with both reading and language (e.g., Frost, Siegelman, Narkiss, & Afek, 2013; Misyak et al., 2010). Similarly, individual differences in memory abilities correlate with reading and language (e.g., Daneman & Carpenter, 1980; Huettig & Janse, 2016). Indeed, it has been suggested that what is measured as individual differences in working memory and language may reflect differences in statistical learning from language input (Misyak et al., 2010; Wells, Christiansen, Race, Acheson, & MacDonald, 2009), likely mediated by underlying chunking skills (McCauley & Christiansen, 2015; McCauley, Isbilen, & Christiansen, 2017).

To summarize, this brief review has pointed to an already existing literature on the role of chunking in ISL, while also noting the similar patterns of modality differences in memory and ISL, as well as how individual differences in both predict language and reading skills. A possible objection, though, is that serial recall is normally thought to involve short-term memory, whereas ISL is about longer term learning. However, current memory research suggests that there may be no sharp distinction between short- and long-term memory but rather that the two are intricately intertwined (e.g., Hasson, Chen, & Honey, 2015). In fact, performance on the standard digit recall task has been shown to

reflect sensitivity to statistical patterns of recurring digit chunks in a natural-language corpus (Jones & Macken, 2015). Similarly, performance in ISL tasks has been shown to be sensitive to natural-language statistics, such as those governing word-initial sounds (e.g., Onnis, Monaghan, Richmond, & Chater, 2005). Thus, we can construe both ISL and serial recall tasks as tapping into long-term distributional learning affected by real-world statistics.

3. Implicit statistical learning as chunking

If ISL relies on basic learning and memory processes to acquire statistical regularities, we should be able to show evidence of such distributional learning using standard memory tasks, such as serial recall. The rationale is that if participants are sensitive to the statistical regularities in the input, they should chunk coherent statistical patterns into larger units, which should facilitate recall. Indeed, in the initial work on ISL within the cognitive tradition, both Miller (1958) and Reber (1967) used serial recall to demonstrate that repeated exposure to structured strings (from an artificial grammar) facilitates memory recall relative to random strings. More recently, immediate recall has been applied to study both visual (Karpicke & Pisoni, 2004) and auditory (Conway, Bauernschmidt, Huang, & Pisoni, 2010) ISL, again indicating that distributional learning facilitates short-term memory performance. Importantly, such statistically-induced facilitation of recall should be observable not only in the context of experiments with artificial language stimuli but also in studies involving real-world natural language statistics. Next, I discuss experimental results that confirm these predictions for both artificial and natural language stimuli.

3.1. A processing-based measure of ISL

Isbilen, McCauley, Kidd, and Christiansen (2017) developed a novel experimental paradigm—the statistically induced chunking recall (SICR) task—which leverages the general capacity for chunking as a processing-based measure of ISL. Participants first listened to a continuous stream of syllables, created by concatenating random combinations of six trisyllabic nonsense words into an 11-minute long input sequence. Following the original study by Saffran, Newport, and Aslin (1996), differences in the transitional probabilities between pairs of syllables that occur within versus between the trisyllabic words can be used to discover word boundaries. After exposure to this artificial language, participants were asked to verbally recall strings consisting of six auditorily presented syllables from the input. Crucially, half of these strings, the experimental items, comprised concatenations of two words from the input language (e.g., *kibudu* + *latibi* → *kibudulatibi*). The other half of the strings, the control items, contained the same syllables as experimental items but presented in a pseudo-random order to remove transitional-probability information (e.g., *kibudulatibi* → *tidubibulaki*). Isbilen et al. predicted that if participants had statistically chunked the syllables from the input into word-like units, then they

should be better able to recall strings consisting of two such words than the same set of syllables in random order.

A first experiment confirmed this prediction, with participants able to correctly recall 37% more syllables from the experimental items compared to the control items. When measuring the recall of syllable trigrams, participants accurately recalled nearly three times as many words compared to random trigrams from the control items. Participants' knowledge of the statistical regularities in the language was also measured using the 2AFC task that is typically used in ISL studies (e.g., Reber, 1967; Saffran, Newport, & Aslin, 1996). Performance was above chance, but subject to an order effect (the order of the SICR and 2AFC tasks was counter-balanced across participants). A 7%-point boost in performance was observed when the 2AFC task was administered after the SICR task. In contrast, the SICR task was not affected by order, suggesting that it may be more robust than the 2AFC task to within-experimental confounding factors. Moreover, there was no correlation between SICR and 2AFC performance, which is consistent with previous studies comparing processing-based measures of ISL to reflection-based ones (e.g., Batterink et al., 2015; Franco et al., 2015). Isbilen and colleagues suggested that the reflection-based 2AFC task may capture more explicit decision-based processing during test, whereas processing-based tasks like SICR better reflect implicitly acquired statistical regularities about which the participants lack awareness.

In a second experiment, Isbilen et al. (2017) asked participants to do the SICR task twice, separated by three weeks, to determine whether performance is reliable across time. Whereas, the 2AFC task for auditory ISL has a mixed record in terms of such test-retest reliability (Siegelman & Frost, 2015), the SICR task was found to be highly reliable for experimental items. These results suggest that SICR has great promise as a measure of individual differences in ISL. Indeed, preliminary results from an ongoing study with 5–6-year-old children indicate that performance on the experimental items in the SICR task correlates significantly with language skill—even when partialling out control item recall, verbal working memory, nonverbal IQ, vocabulary, and age. In contrast, 2AFC performance does not correlate with language comprehension in this sample of children. Thus, the ability to statistically chunk an artificial language appears to be a useful predictor of natural language processing (which likely involves similar chunking processes; Christiansen & Chater, 2016b).

3.2. Real-world statistics facilitate chunking

Just as the SICR task measures the learning of statistical patterns in an artificial language, we might expect that a basic recall task may also reveal sensitivity to real-world statistics. McCauley and Christiansen (2015) tested this prediction by asking participants to recall visually presented strings of consonants, eight or nine letters long. Half of the strings, the experimental items, were created by concatenating consonant bigram or trigram chunks that had a mid to high frequency of occurrence in a large corpus of American English (e.g., *x p l n c r n g l*). The other half of the strings consisted of

pseudo-randomized versions of each of the experimental items, such that the resulting control item has the lowest possible *n*-gram frequency for the component substrings (e.g., *l g l c n p x n r*). When comparing recall of experimental versus control items, McCauley and Christiansen found that readers are indeed sensitive to the statistical patterns of consonant chunks in natural language—even though they do not form syllables—and can draw on this knowledge when they come across those chunks in a novel context. The results also revealed considerable individual differences in sensitivity to real-world statistics over consonant chunks. These differences in "chunk sensitivity" were found to predict the amount of difficulty encountered during online processing of sentences with embedded subject or object relative clauses underlined in (1) and (2), respectively. Participants with better chunking ability processed the sentences faster overall, and experienced less difficulty with more complex object relative clauses.

1. *The reporter that attacked the senator admitted the error.*
2. *The reporter that the senator attacked admitted the error.*

Building on this initial study, McCauley et al. (2017) investigated whether statistical chunking ability at different levels of linguistic abstraction may predict different aspects of language processing. They devised two distinct recall tasks directed at different hypothesized levels of Chunk-and-Pass language processing (Christiansen & Chater, 2016b): phonological and multiword chunking. At the phonological level, they developed a statistically-based version of the classic nonword repetition task (e.g., Dollaghan & Campbell, 1998). Nonwords comprising strings of 4, 5, or 6 syllables were created, and phoneme trigram statistics from a large corpus of American English were used to identify sequences with high likelihood of occurrence (e.g., *krew-ih-tie-zuh*). These "chunk-like" experimental items were contrasted with control items that consisted of the same syllables in a low-frequency combination given triphone statistics (e.g., *tie-zuh-ih-krew*). At the multiword level, high chunkability sequences consisting of four multiword trigrams were constructed based on corpus statistics (e.g., *have to eat good to know don't like them is really nice*).[4] These sequences were matched with low chunkability control items that featured the same functors as the experimental sequences, along with frequency-matched content words, presented in random order (e.g., *years got don't to game have she mean to them far is*). Significant differences were found between the recall of experimental versus control items at both the phonological and multiword levels.

Interestingly, although there were significant differences in performance across individuals for both tasks, the two measures of chunking ability did not correlate with one another. Indeed, the two types of chunking appear to contribute separately to distinct aspects of on-line language processing. When processing sentences with phonological overlap between key words as in (3), participants who were good at phonological chunking experienced little interference relative to sentences such as (4). Within the Chunk-and-Pass framework, this suggests that individuals who are more proficient at phonological chunking may be able to pass up these chunks faster to a higher level of abstraction, and thus overcome the phonological overlap across nouns

and verbs. In contrast, when processing sentences with locally distracting number information as in (5), participants who were good at multiword chunking had fewer problems compared to neutral sentences such as (6). These results suggest that individuals proficient in multiword chunking may be able to chunk the noun-phrase *the key to the cabinets* more efficiently and thereby experience less interference from the number mismatch between *cabinets* and *was* (in 5). Importantly, there was no correlation between phonological chunking ability and the difficulty associated with processing locally distracting number information (in 5); nor was there any correlation between multiword chunking ability and processing difficulty in sentences with phonological interference (in 3). Together, the patterns of correlations between the two chunking sensitivity measures and the sentence processing data suggest that chunking abilities at different levels of linguistic abstraction may affect language processing separately, consistent with predictions from the Chunk-and-Pass framework.

3. *The* <u>cook</u> *that the* <u>crook</u> **consoles controls** *the politician.*
4. *The prince that the crook comforts controls the politician.*
5. *The* **key** *to the* <u>cabinets</u> <u>**was**</u> *rusty from many years of disuse.*
6. *The* **key** *to the* <u>cabinet</u> <u>**was**</u> *rusty from many years of disuse*

4. Conclusion

In this paper, I have proposed to reconcile the separate literatures on statistical learning and implicit learning under the banner of implicit statistical learning. I argued that a unifying perspective might be found by grounding ISL in basic mechanisms of learning and memory—with an emphasis on the uncontroversial process of chunking. This approach offers a more robust measure of sensitivity to statistics in the input by using processing-based test paradigms such as the SICR task instead of the standard 2AFC task (which relies on reflection-based decision processes). By aligning ISL with more basic learning and memory processes, it may also be possible to extend ISL to real-world statistics, for which artificial language learning may be construed as being a proxy. From this perspective, the serial recall task (including its SICR variation) taps into memory for distributional regularities in the input, gleaned through chunk-based processing. This opens new avenues for thinking about language processing and how individual differences in chunking ability might be reflected in skill variation among language learners and users. Of course, this leaves the question of where the statistics come from in the first place. One recent suggestion from the field of language evolution is that constraints on chunking, amplified by cultural evolution, may have shaped linguistic structure (Cornish, Dale, Kirby, & Christiansen, 2017). The hope is that by integrating ISL with basic work on learning and memory, we may come to a more comprehensive, unified perspective on how ISL may contribute to our understanding of the processing, acquisition, and evolution of language (Christiansen & Chater, 2016a).

Acknowledgments

Notes

1. Levelt (2013) notes that Fischer (1922) was the first to conduct an artificial language learning experiment. However, Esper (1925) was the first artificial language study published in English.
2. This does not exclude the possibility that learners might become partially aware of changes in how they process the input. For example, after prolonged exposure to repeated syllable triplets, some of these may appear to "stand out" as rhythmic groupings.
3. In the implicit learning literature, "direct" and "indirect" measures are often used to refer to what I have called reflection-based and processing-based tests, respectively (e.g., Batterink et al., 2015). However, as I take processing-based learning to be the primary target of ISL research, it seems incoherent to refer to measures that more closely tap into such learning as "indirect" and measures that only secondarily relate to such learning as "direct." I therefore have opted not to use these terms in this paper.
4. There is a growing body of work highlighting the use of multiword chunks as building blocks in language acquisition and use (see Arnon & Christiansen, 2017, for review, and the contributions in Christiansen & Arnon, 2017), including how they may be statistically derived (e.g., McCauley & Christiansen, 2011, 2014).

References

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: Past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 20160047.

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, *9*, 621–636.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.

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

Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology*, *55A*, 371–389.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, *36A*, 209–231.

Braine, M. D. (1966). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology*, *72*, 532–540.

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, *9*, 542–551.

Christiansen, M. H., & Chater, N. (2016a). *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.

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

Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory and Cognition*, *28*, 529–538.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356–371.

Conway, C. M., & Christiansen, M. H. (2005). Modality constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *31*, 24–39.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus specific representations. *Psychological Science*, *17*, 905–912.

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*, 561–580.

Cornish, H., Dale, R., Kirby, S., & Christiansen, M. H. (2017). Sequence memory constraints give rise to language-like structure through iterated learning. *PLoS ONE*, *12* (1), e0168532.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, *24*, 87–185.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450–466.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1136–1146.

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 319–332.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation timing have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, *64*, 1021–1040.

Esper, E. A. (1925). *A technique for the experimental investigation of associative interference in artificial linguistic material*. Philadelphia, PA: Linguistic Society of America.

Esper, E. A. (1933). Studies in linguistic behavior organization: I. Characteristics of unstable verbal reactions. *The Journal of General Psychology*, *8*, 346–381.

Esper, E. A. (1966). Social transmission of an artificial language. *Language*, *42*, 575–580.

Fischer, S. (1922). Über das Entstehen und Verstehen von Namen, mit einem Beitrage zur Lehre von den transkortikalen Aphasien. *Archiv für die Gesamte Psychologie*, *43*, 32–63.

Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid serial auditory presentation. *Experimental Psychology*, *62*, 346–351.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality vs. modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*, 117–125.

Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2016). Using statistics to learn words and grammatical categories: How high frequency words assist language acquisition. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 81–86). Austin, TX: Cognitive Science Society.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*, 1243–1252.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*, 304–313.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*, 80–93.

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

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

Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory and Cognition*, *32*, 956–964.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10685.

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.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 79–91.

Levelt, W. (2013). *A history of psycholinguistics: The pre-Chomskyan era*. Oxford, UK: Oxford University Press.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1619–1624). Austin, TX: Cognitive Science Society.

McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *Mental Lexicon*, *9*, 419–436.

McCauley, S. M., & Christiansen, M. H. (2015). Individual differences in chunking ability predict on-line sentence processing. In D. C. Noelle & R. Dale (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

McCauley, S. M., Isbilen, E. S., & Christiansen, M. H. (2017). Chunking ability shapes sentence processing at multiple levels of abstraction. G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2681–2686). Austin, TX: Cognitive Science Society.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology: General*, *90*, 227–234.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

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

Miller, G. A. (1967). Project Grammarama. In *The psychology of communication: Seven essays* (pp. 125–187). New York: Basic Books.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, *2*, 138–153.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*, 498–550.

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

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*, 225–237.

Perruchet, P. (2008). Implicit learning. In J. Byrne (Ed.), *Learning and memory: A comprehensive reference.* Vol. *2* (pp. 597–621). Oxford, UK: Elsevier.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233–238.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*, 246–263.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, *12*, 765–806.

Price, C. J., & Devlin, J. T. (2011). The interactive activation account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Science*, *15*, 246–253.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863.

Remez, R. E., Ferro, D. F., Dubowski, K. R., Meer, J., Broder, R. S., & Davids, M. L. (2010). Is desynchrony tolerance adaptable in the perceptual organization of speech? *Attention, Perception, & Psychophysics*, *72*, 2054–2058.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of memory and language*, *35*, 606–621.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, pii-20160059.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of memory and language*, *81*, 105–120.

Slone, L., & Johnson, S. P. (2015). Statistical and chunking processes in adults' visual sequence learning. In D. C. Noelle & R. Dale (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2218–2223). Austin, TX: Cognitive Science Society.

Studdert-Kennedy, M. (1986). Some developments in research on language behavior. In N. J. Smelser & D. R. Gerstein (Eds.), *Behavioral and social science: Fifty years of discovery* (pp. 208–248). Washington, DC: National Academy Press.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*, B25–B34.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*, 552–564.

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent learning of adjacent and nonadjacent dependencies. *Language Learning*, *66*, 8–30.

Wells, J., Christiansen, M. H., Race, D. S., Acheson, D., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250–271.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationships with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262–274.