Cognition 177 (2018) 198–213

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Linguistic entrenchment: Prior knowledge impacts statistical learning performance

Noam Siegelman[a,*], Louisa Bogaerts[a], Amit Elazar[a], Joanne Arciuli[b], Ram Frost[a,c,d]

[a] The Hebrew University of Jerusalem, Israel
[b] University of Sydney, Australia
[c] Haskins Laboratories, New Haven, CT, USA
[d] BCBL, Basque Center of Cognition, Brain and Language, San Sebastian, Spain

ARTICLE INFO

Keywords:
Statistical learning
Prior knowledge
Entrenchment
Domain generality vs. domain specificity

ABSTRACT

Statistical Learning (SL) is typically considered to be a domain-general mechanism by which cognitive systems discover the underlying statistical regularities in the input. Recent findings, however, show clear differences in processing regularities across modalities and stimuli as well as low correlations between performance on visual and auditory tasks. Why does a presumably domain-general mechanism show distinct patterns of modality and stimulus specificity? Here we claim that the key to this puzzle lies in the prior knowledge brought upon by learners to the learning task. Specifically, we argue that learners' already entrenched expectations about speech co-occurrences from their native language impacts what they learn from novel auditory verbal input. In contrast, learners are free of such entrenchment when processing sequences of visual material such as abstract shapes. We present evidence from three experiments supporting this hypothesis by showing that auditory-verbal tasks display distinct item-specific effects resulting in low correlations between test items. In contrast, non-verbal tasks – visual and auditory – show high correlations between items. Importantly, we also show that individual performance in visual and auditory SL tasks that do not implicate prior knowledge regarding co-occurrence of elements, is highly correlated. In a fourth experiment, we present further support for the entrenchment hypothesis by showing that the variance in performance between different stimuli in auditory-verbal statistical learning tasks can be traced back to their resemblance to participants' native language. We discuss the methodological and theoretical implications of these findings, focusing on models of domain generality/specificity of SL.

1. Introduction

The demonstration that infants can extract statistical properties from continuous speech (Saffran, Aslin, & Newport, 1996) has set the foundations for modern research on Statistical Learning (SL). The study by Saffran et al. (1996) offered a new perspective on how language is acquired by highlighting experience-based principles for detecting regularities in the environment, mainly, the tracking of transitional probabilities (TPs) between adjacent elements in sequentially presented input. In the many studies that followed, this initial demonstration was extended to different modalities (e.g., Fiser & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002), stimuli (e.g., Brady & Oliva, 2008; Gebhart, Newport, & Aslin, 2009), and ages (e.g., Arciuli & Simpson, 2011; Bulf, Johnson, & Valenza, 2011; Campbell, Zimerman, Healey, Lee, & Hasher, 2012), leading to the widespread perception that SL reflects domain-general cognitive computations for extracting and recovering the statistical regularities embedded in any sensory input (see Frost, Armstrong, Siegelman, & Christiansen, 2015, for a review).

At the core of this widely accepted view of SL is the assumption that there is something "common" underlying the learning of regularities across domains. Yet, a range of recent findings seem to challenge this assumption. First, domain-generality, as a theoretical construct, requires that at least some commonalities should exist in computing TPs across sets of visual and auditory stimuli, even if there are some inherent differences in perceiving regularities in different modalities. However, when this was tested by looking at correlations between individual performance across different SL tasks, the results consistently did not support domain-generality. For example, Siegelman and Frost (2015) reported that while the ability to extract TPs in the visual and auditory modality is a stable characteristic of the individual (with a test-retest reliability of around 0.6), correlation between performance in the auditory SL task (modeled on Saffran, Newport, & Aslin, 1996),

* Corresponding author at: Department of Psychology, The Hebrew University, Jerusalem 9190501, Israel.
E-mail address: noam.siegelman@gmail.com (N. Siegelman).

https://doi.org/10.1016/j.cognition.2018.04.011
Received 22 August 2017; Received in revised form 8 April 2018; Accepted 11 April 2018

and a parallel task in the visual modality (modeled on Turk-Browne, Junge, & Scholl, 2005), is virtually zero.[1] Why is it that there is no trace of shared computations across modalities? Even more puzzling, Erickson and her colleagues have recently examined individual performance in two similar auditory SL tasks that varied only in their syllabic components (Erickson, Kaschak, Thiessen, & Berry, 2016). Similar to Siegelman and Frost (2015), they reported that performance for a given set of syllables was highly reliable, with a test-retest reliability spanning between 0.59 and 0.66. However, individual-level correlation in performing the two auditory SL tasks was strikingly low and not significant ($r = 0.17$).[2] Why is it that the seemingly random choice of "words" (i.e. the syllables that co-occur within a familiarization stream) leads to very different learning outcomes, when the same mechanism presumably computes the statistical properties of any speech stream?

A recent developmental study tracking visual and auditory SL performance at different ages (Raviv & Arnon, 2017) showed another puzzling outcome. Whereas visual SL performance improved linearly with age (7–12 years, and see Arciuli & Simpson, 2011, for similar findings), auditory SL performance, albeit lower on the average, did not show any improvement with age. If there is something like a domain-general mechanism for extracting patterns across modalities, why do we observe different developmental trajectories in the visual and auditory modalities?

Another puzzle concerns the very different results obtained with identical auditory SL tasks across speakers of different languages. Two recent studies, one with Italian speakers and one with French speakers, employed an identical experimental design to compare performance on "words" and "phantom words" (sequences of syllables that have the same TP structure as "words" but that never occur in the familiarization steam as a chunk). Surprisingly, these two studies found a virtually opposite pattern of results: In the study with Italian speakers, Endress and Mehler (2009) found that participants were equally familiar with "words" and "phantom words", and concluded that "phantom words" are treated as words. In contrast, in the study with French speakers (Perruchet & Poulin-Charronnat, 2012) consistent preference for "words" over "phantom words" was observed, which suggests that phantoms are not treated as words but rather as non-words. Since the experience-based principles for detecting regularities in continuous speech are supposedly universal, and certainly not privileged to the speakers of only a subset of natural languages, why is it that the language background of the participants appears to determine the outcome of the study?

What is going on, then, in the auditory SL task? Why is it that a task that is taken to reflect a domain-general capacity for registering distributional properties, either through TP computations (e.g., Endress & Langus, 2017; Endress & Mehler, 2009), or through chunk extraction (e.g., Perruchet, Poulin-Charronnat, Tillmann, & Peereman, 2014; Perruchet & Vinter, 1998), shows such peculiar patterns of modality, language, and stimulus specificity? The aim of the present study is to offer some novel insights regarding this important question.

1.1. The tabula rasa assumption

SL research often assumes the learner to be a tabula rasa, thereby viewing learning as the process of assimilating novel regularities. Following this assumption, the learning outcomes of an experiment are typically understood by considering the input structure alone. For example, if participants are presented during familiarization with an input containing 6 "words", with TPs of 1.0 between elements within words, their relative success in 2-AFC trials during the subsequent test phase is discussed by considering (1) the number of words in the stream, (2) the extent of the TPs between elements, and (3) the difference in TPs between "words" and foils in the test phase. The tabula rasa assumption is that the "words" (as well as the foils) were unknown to the participants at the start, so whatever is acquired (or not) during the familiarization session reflects the net efficiency of SL computations.

The tabula rasa assumption may indeed be true in many experimental designs when there is *no prior knowledge regarding co-occurrences of elements in the stream* (e.g., when learning abstract shapes, e.g., Turk-Browne et al., 2005; fractal visual stimuli, Schapiro, Gregory, & Landau, 2014, or novel cartoon figures, Arciuli & Simpson, 2011). However, in the domain of language, the tabula rasa assumption is unlikely. Humans hear speech from birth and start accumulating knowledge about the statistical properties of speech sounds in their native language by the hour. Here we claim that when participants perform an auditory SL task that utilizes verbal material, their existing representations regarding probabilistic co-occurrences of speech sounds in their native language impacts their performance on the task to a large extent. In a nutshell, we argue that one cannot predict the learning outcomes of an auditory SL task that contains linguistic elements, without weighing how the statistical properties of the input steam interact with participants' established expectations regarding the co-occurrences of speech sounds in their native language.

The suggestion that prior linguistic knowledge can modulate performance on auditory SL tasks is not entirely novel: It was raised as a possible explanation when accounting for discrepant results in the auditory SL task (and see Christiansen, Conway, & Curtin, 2000; Christiansen & Curtin, 1999, for an earlier version of this criticism). For example, whereas Perruchet and Poulin-Charronnat (2012) suggested that some peripheral factors of intelligibility of the speech stream could account for Endress and Mehler (2009) reporting no preference for words over phantom words in Italian speakers, Endress and Langus (2017) have raised the possibility that perhaps participants' prior experience in their native language (Italian vs. French) led to the discrepant findings (Footnote 3, p. 41). This issue, however, has critical importance, and cannot be left as a possible post hoc and open explanation for discrepant findings between laboratories. For if Endress and Langus (2017) are right, then the outcome of any study involving the learning of syllables during an auditory SL task, will be contingent on the sampled population. In other words, performance in the task does not simply reflect efficiency of SL computations as it was originally assumed, but reflects patterns of entrenchment of participants in their already established statistics.

The present paper focuses on this possibility by examining whether performance in the auditory SL task may be influenced by entrenchment. We define entrenchment as the influence of previously assimilated knowledge on the learning of the statistical properties from a new input. We examine this hypothesis by monitoring performance in SL tasks that implicate (or not) prior knowledge about the co-occurrences of patterns in the sensory stream. To preview our results, we show that the classical auditory SL task displays clear patterns of entrenchment. In contrast, SL tasks that do not involve prior knowledge regarding co-occurrence of elements are shown to be free of such entrenchment.

The hypothesis that SL performance is affected by entrenchment is compatible with two lines of existing work. First, there is a relatively large set of studies showing that the expectations that participants bring to SL tasks can be easily manipulated, affecting task performance. For example, pre-exposing participants to isolated words or part-words before the beginning of the familiarization stream has a dramatic effect on SL performance, which can either facilitate (Cunillera, Laine, Camara, & Rodriguez-Fornells, 2010; Lew-Williams, Pelucchi, & Saffran, 2011), or hinder (Perruchet et al., 2014; Poulin-Charronnat,

[1] Note that throughout the paper, unless noted otherwise, by auditory SL tasks we refer to tasks using auditory *verbal* material (e.g., Saffran et al., 1996), and by visual SL tasks we refer to tasks using visual non-verbal material (e.g., Kirkham et al., 2002).

[2] We refer here to the results of Experiment 2 from Erickson et al. (2016). In Experiment 1, zero correlations between different auditory SL tasks were also found, but these may be due to a small number of trials in each task, resulting in high measurement error (see Erickson et al., 2016, for discussion; see also Siegelman, Bogaerts, & Frost, 2016).