Attention Selectively Boosts Learning of Statistical Structure

Tess Allegra Forest (tess.forest@mail.utoronto.ca) and Amy Sue Finn (finn@psych.utoronto.ca)
Department of Psychology, 100 St. George Street
Toronto, ON, M5S-3G3, Canada

Abstract

While *statistical learning* (*SL)* has long been described as a learning mechanism that operates automatically across ages and modalities, there are a growing number of cases in which statistical regularities are not learned automatically, and in which attention seems to impact learning. We examined the role of attentional instruction on adults' ability to learn two statistical patterns simultaneously. Results suggest that even without explicit instruction to attend to either pattern, participants automatically learn both patterns, and that explicit instruction to attend to one or both streams improves learning, but only for the attended stream(s). In addition, when attention is directed at only one stream, the learning benefit for that stream is coupled with a learning cost for the unattended stream. This adds to our understanding of the nuanced relationship between attention and SL, by suggesting that when more than one structure is present attention selectively improves SL of attended information in adults, but at the cost of unattended information.

Keywords: Statistical Learning; Attention; Learning

Introduction

SL has been described as an implicit learning process that allows us to *automatically* extract structure from our world (Aslin, Saffran, & Newport, 1998); and indeed, SL often occurs without subjects' awareness (c.f. Batterink, Reber, Neville, & Paller, 2015). It is also broadly available to a variety of learners: functioning in human adults and infants, monkeys, and even rats (Hauser, Newport, & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002; Toro & Trobalón, 2005) and is thus a very useful learning system for a wide range of inputs, allowing us to *effortlessly* parse language and form expectations about our experience.

Still, we know surprisingly little about (1) how SL interacts with other cognitive processes, especially attention or effort, and (2) how well it can account for the learning of more complex structure, especially when two structures are presented simultaneously. Some answers can be gained by taking a careful look at the—albeit rare—instances of when SL fails.

When Statistical Learning Fails

Following thinking that SL is implicit, is the assumption that is operates *continuously* and in all learning contexts, including ones that are noisy and more complex than what typically occurs in the lab. However, it seems that SL does not always operate continuously; when adult participants were exposed to two different auditory patterns in succession, they failed to show learning of the second pattern unless they were presented with a clear cue to start tracking a new pattern, like a pause or voice change (Bulgarelli & Weiss, 2016; Gebhart, Aslin, & Newport, 2009).

SL also does not seem to operate equally well when the stimuli are more complex. For example, infants were unable to successfully track the statistical patterns in a stream of artificial speech if there were two voices present during their exposure (Graf Estes & Lew-Williams, 2015). And infants who were exposed to a stream of artificial speech made up of both two- and three-syllable nonce words were unable to successfully segment that stream of speech, presumably because of this increased variability (Johnson & Tyler, 2010). Furthermore, cross-modality studies indicate that when information is present across modalities, adults' learning is more successful when the information in the two modalities is correlated (Glicksohn & Cohen, 2013; Mitchel & Weiss, 2011)

Collectively, these instances all raise questions about how automatic and continuous SL is, especially when the stimuli are more complex. One reason why we observe these failures in learning could have to do with the fact that other cognitive processes, like attention, might be more important for SL than originally thought.

Attention

Indeed, attention appears to boost SL and can sometimes even be necessary for learning. Toro et. al. (2005) showed auditory SL was negatively impacted when there were high attentional demands from a simultaneous task in the same auditory stream, a different auditory stream, or an accompanying visual stream. It was also shown that instructions to attend to one pattern (i.e., words) improved the learning of that pattern, possibly at the expense of others (i.e., the grammar governing the relationship between those words; Finn, Lee, Kraus, & Hudson Kam, 2014).

It has also been shown that when attention is directed toward only one of two structures, learning of only the attended structure occurs; when participants were asked to perform an attentionally demanding task (an n-back) over the green elements of a green sequence that was interleaved with an independent sequence (that appeared in red), participants only learned the green sequence (Turk-Browne, Jungé, & Scholl, 2005). However, a modified replication of this design showed no impairment in learning the unattended stream, although the task differed: participants were instructed to press for stimulus X in a given color, but not for stimulus X in the other color. (Musz, Weber, & Thompson-Schill, 2015).

Given this work, it is not entirely clear whether and how attention boosts SL, nor in which learning contexts. It remains especially elusive as to *how* attention facilitates SL in learning more complex stimuli. For example, when there are multiple patterns to learn from, how does increasing attention to one impact learning of the other? As noted, one study demonstrated learning of only one stream, but was this preserved or boosted relative to if no attention was directed to either stream? Another study showed learning of both even though only one was attended. *Why?* As these authors note, it could be because the target stimulus appears in the other stream, thus boosting attention to it. This highlights a gap in our understanding: mechanistically, we do not yet know *how* attention impacts learning. Does attention prioritize some information for learning at the expense of other information? Or, might attention improve the learning of all information (relative to no-attention)? As yet, no studies have provided a clear comparison between learning when attention is not manipulated (learning is passive), to when attention is directed to a single aspect of a learner's input (one stream), to when attention is directed globally at all aspects of a learner's input (both streams).

Another important note about the role of attention in SL has to do with how attention is manipulated. In previous work probing the learning of two streams (red and green) mentioned above, attention was manipulated by asking participants to complete a task that was specifically designed to be unrelated to the statistical structure participants saw. The logic was that, in order to preserve the implicitness of SL, the attentional manipulation should not draw attention to the sequential structure of interest for SL outcomes. However, this circumvents the critical reason for studying SL and attention: if, as eluded to above, attention interacts with SL, then SL may not be a simply implicit learning mechanism as it was originally conceived (Batterink et al., 2015). To the best of our knowledge, only one previous study has manipulated attention in a structure-focused (rather than stimulus-focused) way, in which the authors operationalized attention as learners exerting effort toward learning certain patterns present in their input (Finn et al., 2014). Yet, this approach is potentially more akin to some real-world learning (for example, parents may instruct their children to pay attention to aspects of their environment they are trying to learn from, or adult language learners may be instructed to try and learn particular patterns in a foreign language class). To ask whether attention to structure itself impacts SL, we manipulate attention in a structure-focused way, rather than a stimulus-focused way.

The Current Research

We created a visual statistical learning (VSL) paradigm in which participants were exposed to two overlaid visual patterns. This differs from previous multi-stream VSL experiments by presenting the two patterns at exactly the same time (rather than in an interleaved manner). We then conducted a series of studies to ask: 1) Can people learn two statistical patterns simultaneously? 2) How does explicit instruction to attend to structure in one or both streams impact learning for either stream? Answering these questions will help make sense of the differing findings regarding the role of attention in SL and help understand how it interacts with other cognitive mechanisms to enable learning.

Experiment 1

Method

Participants 32 students from the University of Toronto participated in exchange for course credit (Mean Age = 18.03 years, 81% female).

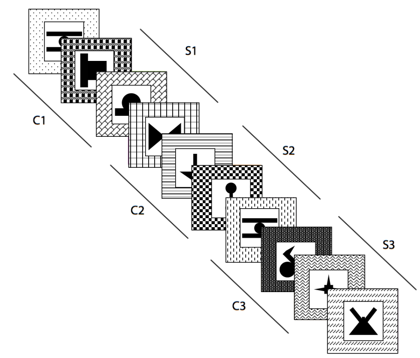

Figure 1. Exposure Stimuli for Exp. 1 and Exp. 2 (Nine distinct colors are represented here as textures to indicate uniqueness in greyscale)

Stimuli and Familiarization All stimuli appeared on an Apple desktop computer screen, and were presented using PsychoPy (Peirce, 2009). Stimuli displays consisted of two streams of visual objects. The first stream was made up of 9 distinctly colored squares, which appeared in 3 predictable triplets (Figure 1, C1, C2, C3). The second stream consisted of 9 distinct shapes. These shapes were also divided into three triplets. In both streams, the transitional probability (TP) within a triplet was 1.0, and the TP between triplets was 0.5. One shape appeared in the center of each colored square. The streams were correlated such that each object in one stream could appear in tandem with three colors from the second stream, and vice versa. Consequently, while there was a correlation between the two streams, one color was not uniquely associated with only one shape. Participants watched a 7-minute sequence of overlaid images, in which triplets from each stream were offset so that the dip in TP which signaled a triplet boundary in one pattern did not correspond with the TP dip in the other pattern (the first item in a shape-triplet was overlaid on the second item in a color-triplet). In each stream, the order of the triplets was randomly generated for each participant with the constraint that one triplet could not immediately follow an instance of the same triplet. All triplets appeared an equal number of times. Each image was presented for 600ms with an inter-stimulus interval of 200ms. A pilot study (n=32) indicted that both streams were learnable when presented individually (Color Mean = 77%, Shape Average = 89%).

Tests Participants completed two types of tests to assess their knowledge of the structure in each stream. First, participants completed 27 two-alternative force choice (AFC) questions (made up of three different test types), where they saw two sequences and were asked to choose which seemed more familiar. In all cases, participants chose between a triplet that was present during exposure and sequences of three items that were not a triplet during exposure. In all foils, each item maintained the same position it had appeared in during exposure, to create 'position-matched' foils (For example, a color that had been the first color in a triplet maintained its position as the first color in a foil, but was paired with the incorrect second and third position items). For all tests, both shape and color were present. One test type (color test) tested learning of the color stream, by comparing triplets from the color stream to position-matched color foils. The same shapes appeared with each choice, and were position-matched shape foils. Participants completed nine color tests.

The second type of test (shape test) compared triplets from the shape stream to position-matched shape foils. Here, the same position-matched color foil was presented with each choice. Participants completed 9 shape tests. The third type of test (preference test) asked participants to choose between a correct color triplet and a correct shape triplet. Participants completed nine preference tests, which allowed us to determine if participants had a preference for the structure of one stream over the other. Participants completed all tests in a blocked order. Half of the participants completed the color test first, and half completed the shape test appeared first. The preference test appeared last for all participants. For each of these tests, each item from a triplet (or foil) was presented in succession on one half of the screen and then disappeared, followed by the second set of items (either triplet or foil), with presentation location of the correct item (right or left side of the screen) randomized between trials.

After completing the AFC tests, participants completed a forced recall task, in which they wrote down the pattern that the colors and, separately, the shapes had appeared in. They were provided a printed bank of all the shapes and the names of the colors they had seen during exposure.

Scoring AFC data were analyzed by scoring participants' responses on whether they correctly chose the previously presented triplet, and calculating an average response accuracy for each participant, separately for each test type. Forced recall scores were coded as the number of correct within-triplet transitions a participant recalled (maximum correct = 6). This was calculated separately for color and shape streams.

Results

Alternative Force Choice Participants showed learning of the color stream greater than what would be expected by chance (M = 65% correct, one-sample t-test: ($t(31) = 4.10$, $p < .001$, $d = 0.73$; Fig. 1A), and marginal learning of the shape stream (M = 58% correct, one-sample t-test: $t(31) = 1.98$, $p = .056$, $d = 0.35$; Fig. 1A). A within-subjects t-test showed that performance on the color and shape streams did not differ ($t(61.25) = 1.26$, $p = .21$, $d = 0.32$; Fig. 1A). Participants did not show a preference for shape or color triplets ($t(31) = -1.45$, $p = .16$, $d = 0.26$).

Forced Recall Participants recalled more than zero shape transitions ($M = 2.66$; one-sample t-test ($t(31) = 8.93$, $p < .001$, $d = 1.58$; Fig. 1B). Color pattern recall ($M = 2.06$, $SD = 0.95$), was also greater than 0 (one-sample t-test: $t(31) = 12.30$, $p < .001$, $d = 2.18$; Fig. 1B) and there was no difference in recall of the color and shape streams ($t(49.58) = 1.58$, $p = .120$, $d = 0.40$).

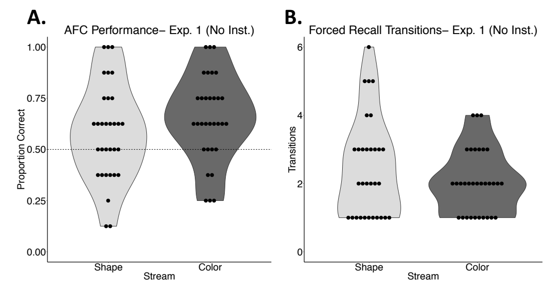

Figure 1. Experiment 1 (No Instruction)

Discussion

Experiment 1 shows that participants were able to learn two statistical streams simultaneously. This demonstrates that two simultaneously occurring patterns can be learned without explicit instruction. The results from Experiment 1 lay the groundwork for Experiment 2, which asks how explicit instruction to one or both stream(s) impacts learning for both streams.

Experiment 2

Method

Participants 96 students from the University of Toronto participated in exchange for course credit (Mean Age = 18.84 years, 69% female).

Stimuli and Familiarization All stimuli and devices were identical to Exp. 1, except for the instructions. In Exp. 2, all participants were informed, prior to familiarization, that an order governed the items that they would see, and that they should try to learn it. 32 participants were told that the shapes occurred in a particular order (Shape Instruction), 32 were told that the colors occurred in a particular order (Color Instruction) and 32 were told that both the shapes and colors occurred in particular orders (Dual Instruction).

Results

Shape Instructions

Alternative Force Choice A one-sample t-test revealed that participants performed greater than chance on the shape stream (M = 72% correct, $t(31) = 5.26$, $p < .001$, $d = 0.93$; Fig. 2A), and performance on this test was better than Exp. 1 ($t(62.0) = 2.33$, $p = .02$, d = 0.58; Table 1). An additional one-sample t-test revealed that there was a trend toward better-than-chance performance on the color stream (M = 56% correct, $t(31) = 1.88$, $p = .07$, $d = 0.33$; Fig. 2A), and a trend toward performance on the color test items being lower than Exp. 1 (between subjects t-test: $t(61.25) = -1.80$, $p = 0.08$, $d = 0.45$; Table 1). Finally, a within-subjects t-test revealed that participants chose shape triplets more often than color triplets ($t(31) = -3.26$, $p = .003$, $d = 0.58$) when forced to choose on preference tests.

Forced Recall Participants recalled more than zero shape transitions ($M = 3.84$, $SD = 2.13$; one-sample t-test: $t(31) = 10.22$, $p < .001$, $d = 1.81$; Fig. 2B). Color pattern recall ($M = 1.75$, $SD = 0.76$) was also greater than zero (one-sample t-test: $t(31) = 12.30$, $p < .001$, $d = 2.18$; Fig. 2B). A within-subjects t-test showed that participants recalled more shape transitions than color transitions ($t(38.83) = 5.24$, $p < .001$, $d = 1.31$). In addition, more shape transitions were recalled as compared to Exp. 1 (between subjects t-test: $t(58.29) = 2.63$, $p = 0.01$, $d = 0.66$). The number of color transitions recalled did not differ from Exp. 1 ($t(59.25) = -1.45$, $p = 0.15$, $d = 0.36$).

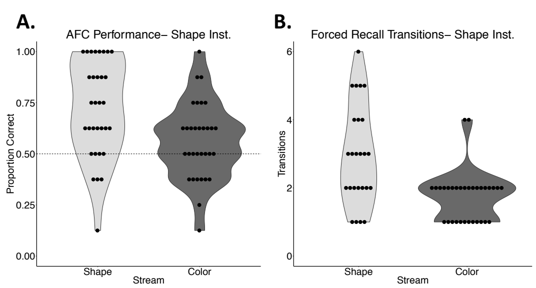

Figure 2. Shape Instruction Performance

Color Instructions

Alternative Force Choice A one-sample t-test indicated that participants' performance was greater than chance on the color stream (M = 85% correct; one sample t-test: $t(31) = 11.17$, $p < 0.001$, $d = 1.97$; Fig. 3A), which was greater than in Exp. 1 (between subjects t-test: $t(60.38) = 4.09$, $p < .001$, $d = 1.02$; Table 1). Their performance was greater than chance on the shape stream (M = 57% correct, $t(31) = 2.12$, $p = 0.04$, $d = 0.37$, Fig. 3A), but performance was not different than Exp. 1 (between subjects t-test: $t(59.17) = -0.22$, $p = 0.83$, $d = 0.05$; Table 1). Finally, a within-subjects t-test revealed that participants chose color more often than shape on preference tests ($t(31) = 3.45$, $p = .002$, $d = 0.61$).

Forced Recall Participants recalled the more than zero color transitions ($M = 3.59$, $SD = 1.95$; one-sample t-test: $t = t(31) = 10.43$, $p < .001$, $d = 1.84$; Fig. 3B). Shape pattern recall ($M = 2.66$, $SD = 1.73$) was also greater than zero (one-sample t-test: $t(31) = 8.67$, $p < .001$, $d = 1.53$; Fig. 3B). A within-subjects t-test showed that participants recalled slightly more color transitions than shape transitions ($t(61.18) = 2.03$, $p = .046$, $d = 0.51$). Color pattern recall was higher than in Exp. 1 (between-subjects t-test: $t(44.90) = 4.00$, $p < .001$, $d = 1.00$). However, shape recall was not different than Exp. 1 (between-subjects t-test: $t(61.82) = 0.15$, $p = 0.88$, $d = 0.00$).

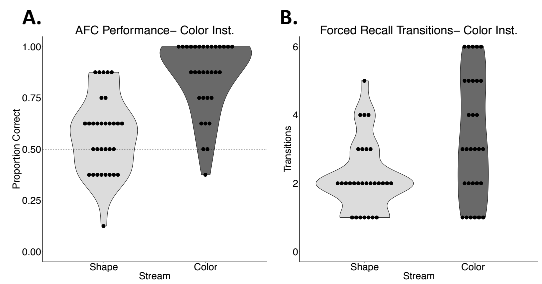

Figure 3. Color Instruction Performance

Dual Instructions

Alternative Force Choice Participants showed shape stream learning (M = 60% correct, $t(31) = 2.29$, $p = 0.02$, $d = 0.40$; Fig. 4A), which was not different than shape stream performance in Exp. 1 ($t(61.95) = 0.26$, $p = 0.79$, $d = 0.07$; Table 1). They also showed learning of the color stream (M = 77% correct) that was different from chance ($t(31) = 8.30$, $p < .001$, $d = 1.47$; Fig. 4A), and higher than chance performance on the color stream from Exp. 1 ($t(61.14) = 2.44$, $p = 0.02$, $d = 0.61$; Table 1). Participants did not show a preference for shape or color triplets ($t(31) = -0.17$, $p = 0.87$, $d = 0.03$). Additional between-subjects t-tests also showed that shape stream performance was lower than with only shape instructions ($t(61.96) = -2.03$, $p = .046$, $d = 0.51$), and that there was a non-significant trend for color performance to be lower than with only color instructions ($t(61.87) = -1.72$, $p = 0.09$, $d = 0.43$).

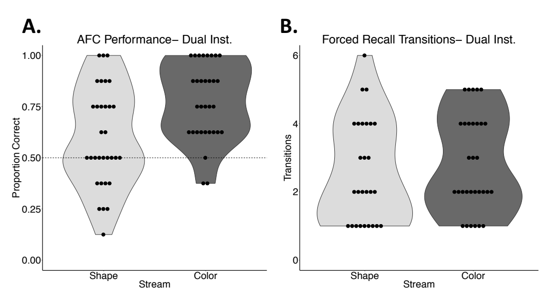

Figure 4. Dual Instruction Performance

Forced Recall Participants recalled the more than zero shape transitions (M = 3.28, SD = 2.14; one-sample t-test: $t(31)$ = 8.66, p < .001, d = 1.53; Fig. 4B). Color pattern recall (M = 2.99, SD = 1.58), was also greater than zero (one-sample t-test: $t(31)$ = 10.43, p < .001, d = 1.84; Fig. 4B). There was no difference in recall of the color and the shape streams ($t(56.93)$ = -0.66, p = 0.51, d = 0.17). A between-subjects t-test also showed recall of the shape stream was not different than in Exp. 1 ($t(58.08)$ = 1.44, p = 0.16, d = 0.36) while color recall was higher than Exp. 1 (between subjects t-test: $t(50.85)$ = 2.79, p = .007, d = 0.70).

Table 1: Shape and Color Performance

Instructions	Shape AFC	Color AFC	Preference?
None	58%	65%	No
Shape	72%	56%	Yes, Shape
Color	57%	85%	Yes, Color
Dual	60%	77%	No

Discussion

Results from Experiment 2 suggest that participants who attended to one structure show increased knowledge for that structure. Results also suggest that unattended information is learned less well than when the same information is learned without any instructions. While this supports the idea that attention benefits SL, it adds the caveat that the scope of this benefit is limited to the attended information. Moreover, when multiple sources of structure are attended, learning is not improved as much as attending to only one source (be it color or shape).

General Discussion

Our results show that adults are able to learn two simultaneously presented, but independent, statistical patterns. Additionally, they show that attention to one pattern of information improves learning of the attended information, but that attention to multiple patterns does not improve performance as much as attending to one pattern alone. It was previously unknown *how* attention to structure would impact the learning of multiple structures simultaneously. This is a critical piece of information given that 1) humans are naturally exposed to multiple structures simultaneously, rather than in isolation and 2) that attentional capacity has limits (Cowan et al., 2009), with learners constantly having to prioritize different information. In the following paragraphs, we will unpack two important nuances in these data that are central to guiding our conclusions and then review additional relevant work.

First, the benefit of attention appears to be limited: participants who attended to one stream showed greater performance on the attended stream than participants who attended to both streams. This suggests that the extent to which attention can improve SL is limited by how much attention can be allocated to any source of information.

Secondly, although there was a cost for the learning of shapes when attending to color, it's important to note that performance was still greater than chance for the unattended information in both color and shape instruction conditions. This finding stands in contrast to previous work which reported that participants showed no knowledge of the structure of the unattended stream (Turk-Browne et al., 2005). A possible reason for this difference is that—in the present study—both streams of information were present simultaneously; this could have made it harder for participants to ignore the unattended stream (see Musz et al., 2015 for a similar argument about the difficulty of not attending to the unattended information). This explanation is likely. Indeed, a failure to successfully ignore the unattended information leading to learning mirrors results from older adults who show can show superior learning precisely because they have reduced cognitive control (Amer, Campbell, & Hasher, 2016). Moreover, the attentional manipulation used here directed participants to attend to the *structure* present in one stream. This way of manipulating attention (in a top-down, structure-focused way) potentially has very different interactions with SL than when participants' attention is manipulated by an unrelated task in a stimulus-focused way.

Taken together, this pattern of data clearly show that attention can boost SL when attention is directed toward the structure itself. How are we to reconcile this with the observation that SL is available in populations with less advanced attentional abilities, like preverbal infants and rats?

One important note is that many of the studies that have examined the role of attention in SL with infants and non-human animals have used simple patterns. Introducing two statistical patterns suggests that environmental complexity could impact how attention mediates SL. Furthermore, the argument that attention is not required for SL because infants can make use of this learning mechanism presupposes that the attentional abilities of young infants are not sophisticated. Yet, there is increasing evidence that the allocation of attention in young infants is more advanced than we might presume. For example, infants allocate their attention to elements of their input that are neither the most nor the least predictable; instead, they look longer at elements that are mildly novel (Kidd, Piantadosi, & Aslin, 2014). While these authors did not examine learning directly, their findings suggest that the attentional abilities of infants as young as 7 months are more advanced than initially believed. Furthermore, it suggests that attention could facilitate learning in infants and adults alike.

Furthering the link between attention and SL may also shed light on differences in learning that are observed between children and adults, given that attention matures throughout childhood. Since children have less focused attention than adults, it is possible that the highly specific benefit of attention observed in adults may be less constrained in children. It is also possible that the highly selective increase in performance for attended information will be more widespread in children, and that they will improve on unattended information as well. This could eventually help us understand why it is that children often out-perform adults on feats like language learning—if the mechanisms that help children learn language so successfully interact with cognitive mechanisms that are more developed in adults, childhood could be a unique window of opportunity for learning without the influence of cognitive control mechanisms that emerge later in life.

Understanding how SL interacts with the development of other cognitive processes, like attention, will be a major direction for future research. The present findings indicate that, in adults, attention to structure has the power to improve knowledge of statistical information, but that this benefit is both limited and selective. These findings move us closer towards understanding how SL is able to operate successfully across learning environments that vary in complexity, and elaborate on our understanding of how attention mediates this fundamental learning mechanism.

Acknowledgments

Special thanks to Z. Abolgashem for help with data collection; to A. Kiyomarsi with programming; to M. Dubois and S. Sutherland for their helpful comments on earlier drafts of this manuscript; and to the rest of the Learning and Neural Development Lab for their ongoing support during this project.

References

Amer, T., Campbell, K. L., & Hasher, L. (2016). Cognitive Control As a Double-Edged Sword. *Trends in Cognitive Sciences*, *20*(12), 905–915. http://doi.org/10.1016/j.tics.2016.10.002

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324. http://doi.org/10.1111/1467-9280.00063

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. http://doi.org/10.1016/j.jml.2015.04.004

Bulgarelli, F., & Weiss, D. J. (2016). Anchors Aweigh : The Impact of Overlearning on Entrenchment Effects in Statistical Learning, *42*(10), 1621–1631.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., & Conway, A. R. A. (2009). On the Capacity of Attention: Its Estimation and Its Role in Working Memory and Cognitive Aptitudes. *Cognitive Psychology*, *51*(1), 42–100. http://doi.org/10.1016/j.cogpsych.2004.12.001.On

Finn, A. S., Lee, T., Kraus, A., & Hudson Kam, C. L. (2014). When it hurts (and helps) to try: The role of effort in language learning. *PLoS ONE*, *9*(7). http://doi.org/10.1371/journal.pone.0101806

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, *33*(6), 1087–1116. http://doi.org/10.1111/j.1551-6709.2009.01041.x

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review*, *20*, 1161–1169. http://doi.org/10.3758/s13423-013-0458-4

Graf Estes, K., & Lew-Williams, C. (2015). Listening Through Voices: Infant Statistical Word Segmentation Across Multiple Speakers. *Developmental Psychology*, *51*(11), 1–12. http://doi.org/10.1037/a0039725

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), 53–64. http://doi.org/10.1016/S0010-0277(00)00132-3

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. http://doi.org/10.1111/j.1467-7687.2009.00886.x

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. http://doi.org/10.1111/cdev.12263

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain-general learning mechanism. *Cognition*, *83*, 35–42. http://doi.org/10.1016/S0010-0277(02)00004-5

Mitchel, A. D., & Weiss, D. J. (2011). Learning Across Senses: Cross-Modal Effects in Multisensory Statistical Learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(5), 1081–1091. http://doi.org/10.1037/a0023700.Learning

Musz, E., Weber, M. J., & Thompson-schill, S. L. (2015). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Attention, Perception & Psychophysics*, *77*, 78–96. http://doi.org/10.3758/s13414-014-0757-5

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy, *2*(January), 1–8. http://doi.org/10.3389/neuro.11.010.2008

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), 25–34. http://doi.org/10.1016/j.cognition.2005.01.006

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–875. http://doi.org/10.3758/BF03193539

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology. General*, *134*(4), 552–64. http://doi.org/10.1037/0096-3445.134.4.552