

Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism(s)

Lauren L. Emberson¹ | Jennifer B. Misyak² | Jennifer A. Schwade³ |
Morten H. Christiansen³ | Michael H. Goldstein³

¹Psychology Department, Princeton University, Princeton, New Jersey

²Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, UK

³Psychology Department, Cornell University, Ithaca, New York

Correspondence

Lauren L. Emberson, Peretsman-Scully Hall, Psychology Department, Princeton University, Princeton, NJ 08544, USA. Email: lauren.emberson@princeton.edu

Funding information

National Institute of Child Health and Human Development, Grant/Award Number: 4R00HD076166-02 and HD076166-01A1; Canadian Institutes of Health Research, Grant/Award Number: 201210MFE-290131-231192; James S. McDonnell Foundation, Grant/Award Number: 22002050

Abstract

Statistical learning (SL), sensitivity to probabilistic regularities in sensory input, has been widely implicated in cognitive and perceptual development. Little is known, however, about the underlying mechanisms of SL and whether they undergo developmental change. One way to approach these questions is to compare SL across perceptual modalities. While a decade of research has compared auditory and visual SL in adults, we present the first direct comparison of visual and auditory SL in infants (8–10 months). Learning was evidenced in both perceptual modalities but with opposite directions of preference: Infants in the auditory condition displayed a novelty preference, while infants in the visual condition showed a familiarity preference. Interpreting these results within the Hunter and Ames model (1988), where familiarity preferences reflect a weaker stage of encoding than novelty preferences, we conclude that there is weaker learning in the visual modality than the auditory modality for this age. In addition, we found evidence of different developmental trajectories across modalities: Auditory SL increased while visual SL did not change for this age range. The results suggest that SL is not an abstract, amodal ability; for the types of stimuli and statistics tested, we find that auditory SL precedes the development of visual SL and is consistent with recent work comparing SL across modalities in older children.

KEYWORDS

abstract, auditory, domain-general, infant, statistical learning, visual

1 | INTRODUCTION

Young infants have the remarkable ability to shape their perceptual and cognitive systems based on their experience. One way that an infant can adapt to their environment is by uncovering statistical regularities in sensory input, a phenomenon known as statistical learning (SL, Saffran, Aslin, & Newport, 1996; Kirkham, Slemmer, & Johnson, 2002). SL has been implicated in the development of language learning (Romberg & Saffran, 2010), object and scene perception (Fiser & Aslin, 2002), and music perception (McMullen & Saffran, 2004). However, despite the importance of SL to understanding perceptual

and cognitive development, very little is known about the nature and development of its underlying mechanisms.

A powerful way to uncover the mechanisms supporting SL is to directly compare learning across perceptual modalities. Comparing SL across modalities entails presenting the same *statistical information* (e.g., the same underlying structure and amount of exposure) while varying *perceptual information* (e.g., whether the individual tokens are auditory or visual). Importantly, perceptual manipulations are well beyond perceptual thresholds so differences in learning

do not arise from an inability to identify individual tokens but from differences in the interaction of perceptual and learning systems in gathering or using statistical information. Identical learning outcomes across different perceptual conditions would indicate that SL is an abstract, amodal learning ability that is insensitive to perceptual information. However, a decade of research in adults has established that SL systematically differs across auditory and visual perceptual modalities (e.g., Conway & Christiansen, 2005, 2009; Saffran, 2002; reviews by Krogh, Vlach & Johnson, 2013; Frost, Armstrong, Siegelman, & Christiansen, 2015). For example, a number of studies have suggested that, in adults, auditory SL is superior to visual SL when statistical information and other perceptual conditions are held constant (Conway & Christiansen, 2005, 2009; Emberson, Conway, & Christiansen, 2011; Robinson & Sloutsky,). Despite the early success of numerous models of SL that focus solely on statistical information (e.g., Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Thiessen & Erickson, 2013), these convergent findings suggest that the mechanisms underlying SL are not amodal and abstract but are importantly affected by perceptual information.

Despite many demonstrations of SL in both auditory and visual modalities in infants (e.g., Fiser & Aslin, 2002; Kirkham et al., 2002; Saffran et al., 1996; Saffran, Johnson, Aslin, & Newport, 1999), no study has directly compared learning across the two modalities. Moreover, it is not possible to compare outcomes from previous studies because of substantial differences in methodology and statistical information. Thus, we present the first direct comparison of SL across perceptual modalities in infancy. There are a number of possible relationships between SL, perceptual modality and development that might be observed. Here, we consider two primary possibilities: It is possible that, early in development, SL is largely unaffected by perceptual information, with modality differences only arising later in development. In contrast, infant SL might be more affected by perceptual information earlier rather than later in development as the developing learning systems are less robust and not able to compensate for biases in perceptual processing. Answers to these questions will inform broader investigations of whether SL is developmentally invariant (Kirkham et al., 2002; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997) or whether SL abilities improve with age (Thiessen, Hill, & Saffran, 2005; Arciuli & Simpson, 2011; see discussion by Misyak, Goldstein, & Christiansen, 2012), and how SL contributes to the development across different domains (e.g., relations between developmental changes in auditory SL and early language development).

As an initial step toward answering these important theoretical questions, the current study presents the first direct comparison of auditory and visual SL in infants, targeting a well-studied age for SL (8- to 10-months-old; Fiser & Aslin, 2002; Kirkham et al., 2002; Saffran et al., 1996). Our goal is to spark investigations into similarities and differences in SL across perceptual modalities. These investigations will bring a deeper understanding of mechanisms supporting SL early in life when this learning ability is believed to support development across numerous domains. This line of research complements efforts by Raviv and Arnon (2018) to compare

Research Highlights

- First direct comparison of statistical learning (SL) abilities across perceptual modalities in young infants (8–10 months) using temporally presented complex, familiar stimuli (speech and faces).
- We find superior auditory SL for speech stimuli in 8–10 month olds.
- Discovery of a developmental shift in auditory (speech) but not visual (faces) SL in this age range.
- Evidence that while SL is domain-general, it is not an abstract ability insensitive to perceptual information early in life.

auditory and visual SL across childhood, and who report modality differences in this age range. The current work extends these investigations to much younger infants and, importantly, to ages where it is believed that SL is an essential skill for breaking into the structure of the environment.

One of the challenges of comparing SL across modalities in infancy is that it is uncommon to compare the *amount* of learning (i.e., using looking times). One way to compare learning is to consider the magnitude of the difference in looking to novel and familiar trials. Kirkham et al. (2002) used this approach and compared looking time to novel and familiar test trials over 3 age groups. An interaction of age and test trial type (mixed ANOVA) would be indicative of changes in learning with age. Another classic way to consider learning is to employ the Hunter and Ames model (1988) where familiarity preferences reflect a weaker stage of encoding than novelty preferences. Numerous studies of SL have evoked this model when considering learning outcomes (Johnson & Jusczyk, 2001; Jusczyk & Aslin, 1995; Saffran & Thiessen, 2003; Thiessen & Saffran, 2003; also see Aslin & Fiser, 2005, for a discussion). For example, Saffran and Thiessen (2003) state that the “direction of preference reflects [...] factors.. such as the speed of the infant’s learning” (p. 485).

We strove to equate learning conditions across modalities. First, while visual SL studies have typically employed infant-controlled habituation (Fiser & Aslin, 2002; Kirkham et al., 2002) and auditory SL studies have employed fixed periods of familiarization to sounds (e.g., Saffran et al., 1996; Graf Estes, Evans, Alibali, & Saffran, 2007), we employed infant-controlled habituation in both visual and auditory conditions. Second, we aimed to better equate the type of stimuli across perceptual modalities: Previous visual SL studies have employed geometric shapes (Fiser & Aslin, 2002; Kirkham et al., 2002, see Sloan, Kim, & Johnson, 2015, for differences in face and shape SL in infants), whereas auditory SL studies have typically used speech sounds (Saffran et al., 1996; however, see Creel, Newport, & Aslin, 2004; Saffran et al., 1999). Infants in the first year of life have had considerable exposure to speech sounds, making speech more familiar than geometric shapes; moreover, speech sounds are more perceptually

complex, and infants are becoming very skilled at processing speech (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984). Faces are a comparable type of stimulus for the visual modality (Nelson, 2001; Pascalis et al., 2005) and thus, using a comparison that has been employed many times in the field of early cognitive/perceptual development (Lewkowicz & Ghazanfar, 2009; Maurer & Werker, 2014), we compared SL using speech and faces. #AuthorQueryReply#

Finally, visual and auditory SL studies with infants always employ different rates of stimulus presentation, with visual stimuli presented at a much slower rate than auditory stimuli (e.g., visual SL: 1 stimulus/s; Kirkham et al., 2002; auditory SL: 4–5 stimuli/s; Saffran et al., 1996; Pelucchi, Hay, & Saffran, 2009). Faster presentation rates decrease visual SL in children (Arciuli & Simpson, 2011) and adults (Conway & Christiansen, 2009; Turk-Browne, Jungé, & Scholl, 2005). Research in adults suggested the opposite effect with auditory SL, with decreased learning at slower rates of presentation (Emberson et al., 2011). Since rate and perceptual modality are two types of perceptual information that have been shown to interact in adult learners (Arciuli & Simpson, 2011; Emberson et al., 2011), we chose presentation rates that balanced the constraints of achieving similar methods across modalities with the rate required by specific perceptual systems (visual rate of presentation: 1 stimulus/s, cf. Kirkham et al., 2002; auditory rate of presentation: 2 stimuli/s, cf. Thiessen et al., 2005). Along with the use of infant-controlled habituation, we can determine and control for any differences in (statistical/perceptual) exposure across perceptual modalities.

2 | METHODS

2.1 | Participants

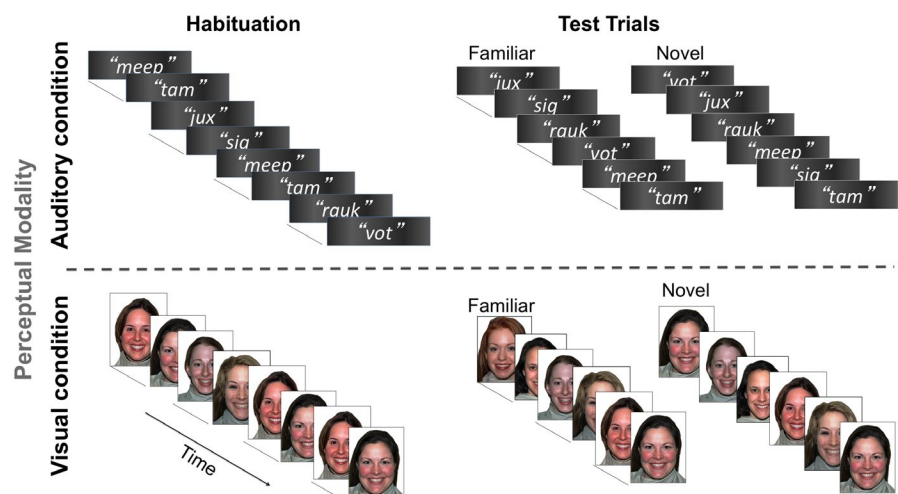
The final sample was 33 infants (auditory:17; visual:16) with a mean age of 9.2 months ($SD = 0.57$, 8.1–10.0 months, 19 female). See Data S1 for more details and exclusionary criteria.

2.2 | Stimuli and statistical sequences

This study employed equivalent sets of visual and auditory stimuli. Six smiling, Caucasian, female faces were selected from the NimStim database (Figure 1; Tottenham et al., 2009). Faces were presented individually at a rate of 1 stimulus/s (250ms SOA). Six monosyllabic nonwords (*vot*, *meep*, *tam*, *jux*, *sig*, *rauk*) were recorded separately to control for effects of coarticulation and produced with equal lexical stress and flat prosody (adult-directed speech) by a female native English speaker. The length of each utterance was edited to have a uniform duration of 375 ms and stimuli were presented at a rate of 2 stimuli/s. Nonwords were presented at 58 dB and accompanied by the projected image of a checkerboard (4 × 4 black-and-white, with gray surround) to direct infant attentional focus. Both face and checkerboard stimuli used for the auditory condition subtended 14.6° of visual angle (Kirkham et al., 2002).

Sequence construction followed Kirkham et al. (2002, Figure 1) such that, for each condition (visual or auditory), the six stimuli (faces or nonwords) were grouped into two mutually exclusive sets of *bigrams*. Each infant was exposed to one bigram set. Habituation sequences were constructed by concatenating bigrams of a given set in random order with the a priori constraint that there could be no more than four consecutive presentations of a single bigram and all presented with equal frequency within the 60 s sequences. The only cue to bigram structure was the statistical information in the stream: Both co-occurrence frequencies and transitional probabilities could support bigram segmentation (Aslin, Saffran, & Newport, 1998). Twelve different habituation sequences were constructed for each bigram set for each condition. There were two types of test trial sequences: Familiar and Novel. Familiar trials were constructed using identical methods as the habituation sequences. Novel trials were constructed using a random order of all stimuli with the constraint that there be no consecutive repeats and all items have equal frequency. Three novel and three familiar test trials were constructed for each bigram set and for each condition. Both habituation and test trial sequences were 60 s long.

FIGURE 1 Depiction of sequences employed for habituation and test trials (familiar and novel) for the visual and the auditory perceptual conditions (bottom and top). Each stimulus was presented individually and centrally to the infants. The order to stimuli presented is depicted along the diagonal in the figure. Lines below the sequences indicate bigram structure. Note that perceptual modality was a between-subjects factor: infants had either visual or auditory exposure



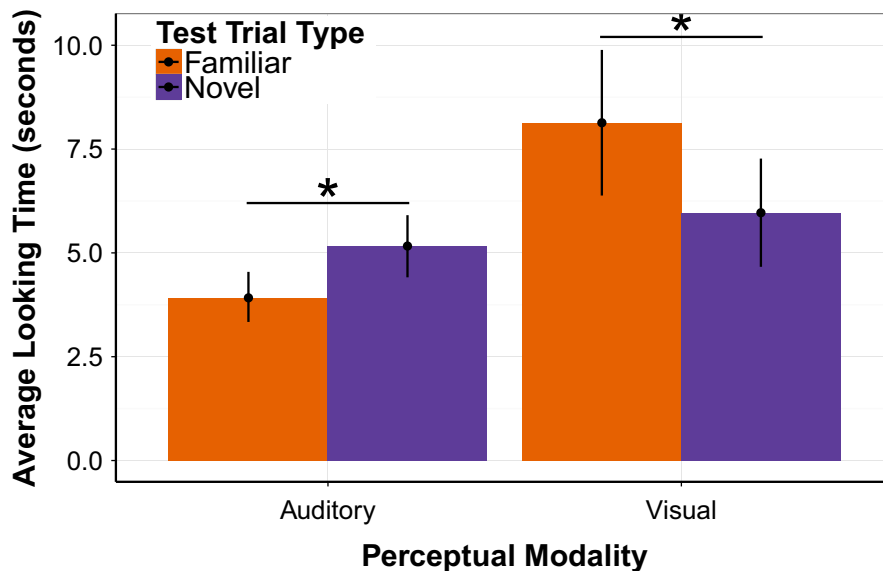


FIGURE 2 Looking to novel and familiar test trials across auditory and visual perceptual modalities

2.3 | Procedure

Infants were seated in a caregiver's lap in a darkened room. Caregivers were instructed to keep their infants on their laps facing forward but not to interfere with infant looking or behavior. Each caregiver listened to music via sound-attenuating headphones and wore a visor that prevented visual access to the stimuli. All visual stimuli (a checkerboard during the auditory condition, faces in the visual condition, and the attention getter, used in both conditions) were projected centrally, and a camera recorded infant eye gaze. Auditory stimuli (speech tokens or the sound for the attention getter) were presented from a speaker placed in front of the infants and below the visual stimuli. Stimulus presentation was controlled by Habit 2000 (Cohen, Atkinson, & Chaput, 2000) operating on a Macintosh computer running OS 9. An observer in a different room, blind to sequences and trial types, recorded looks toward and away from the visual stimuli. See Data S1 for analyses verifying coder reliability.

Infants were presented with an attention-getting animation (rotating, looming disc with sound) between trials until the infants looked centrally at which point a sequence was presented. If the infant did not look at the beginning of the sequence for at least two seconds, the trial was not counted; the attention-getter played again and once the infant looked centrally, the same sequence was repeated (Kirkham et al., 2002). If the infant looked for two seconds or longer, the sequence played until infants looked away for two consecutive seconds or the sequence ended (Saffran et al., 1996).

Habituation sequences were presented in random order until infants either reached the habituation criterion or all habituation sequences had been presented. The habituation criterion was defined as a decline of looking time by more than 50% for four consecutive trials, using a sliding window, compared to the first four habituation trials (Kirkham et al., 2002; similar to Graf-Estes et al., 2007, with four vs. three trials for comparison). Infants were then presented

with six test trials in alternating order by test trial type (familiar and novel) with the order of alternation (i.e., novel first or familiar first) counterbalanced across infants. All statistical analyses were conducted in R (RStudio, 0.98.1028).

3 | RESULTS

3.1 | Comparing learning across perceptual modalities

Mean looking times were submitted to a 2 (test trial: novel, familiar) \times 2 (perceptual modality: visual, auditory) mixed ANOVA (within and between subjects factors, respectively). This analysis revealed an interaction of perceptual modality and test trial type ($F(1, 31) = 24.60$, $p < 0.001$, Figure 2) that was driven by opposite directions of preference at test across perceptual modalities: Infants in the Auditory modality showing a significant novelty preference (12 of 17 infants showed bias toward the Novel trials, Wilcoxon signed-rank test, $V = 129$, $p = 0.01$), and infants in the Visual modality showed a significant familiarity preference (14 of 16 infants looked longer to the familiar trials, Wilcoxon signed-rank test, $V = 11$, $p = 0.002$). Thus, we found evidence of significant learning in each perceptual modality. Based on the Hunter and Ames model (1988), there is evidence of weaker learning in the visual modality compared to the auditory modality, as indicated by different directions of preference (familiarity vs. novelty, respectively).

This analysis also revealed a main effect of Perceptual Modality ($F(1, 31) = 10.09$, $p = 0.003$) driven by longer looking in the visual condition. This finding is surprising because there were no differences in looking across modalities during habituation ($p > 0.3$). However, using proportion of looking to control for the generally longer looking at visual sequences at test, we still found a significant interaction between perceptual modality and test trial, $F(1, 31) = 16.29$, $p < 0.001$ (see Figure S1).

FIGURE 3 Habituated infants in the auditory condition show a significant correlation between age and Difference Score with older infants showing a strong novelty preference. Auditory mean age = 9.4 ($SD = 0.41$); Visual mean age = 9.0 ($SD = 0.66$)

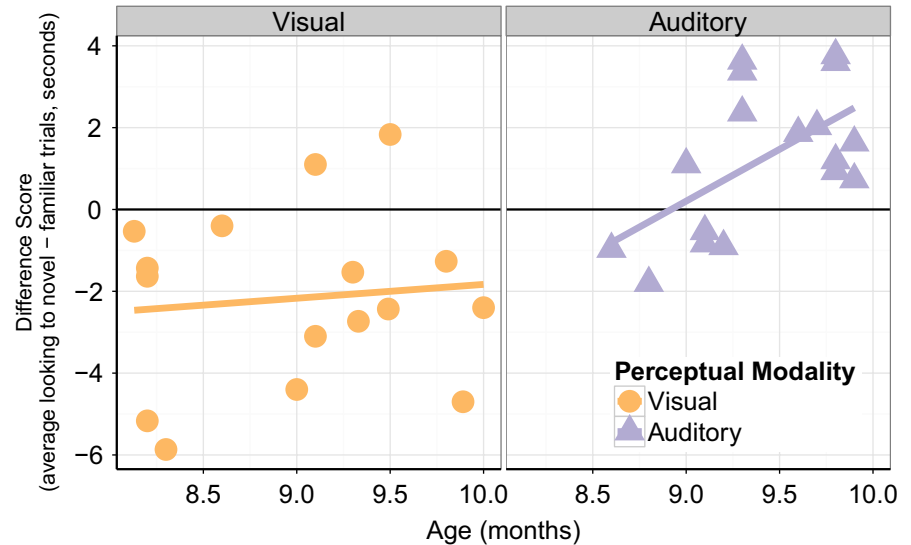


TABLE 1 Comparison of Rate, Age, and Statistical Exposure across Auditory SL studies selected to be most similar to the current paradigm and age range. Note: Exposure was calculated by unit of structure (i.e., each word as in Saffran et al., 1996; each bigram or trigram in Thiessen et al., 2005)

Study	Age (months)	Rate (ms)	Exposure	Outcome
Current:Emberson et al.	9	500	25	Learning
Thiessen et al. (2005)	8	400	24	No learning for AD speech
Saffran et al. (1996)	8	222	45	Learning
Pelucchi et al. (2009)	8.5	167	45	Learning

Abbreviation: SL, statistical learning.

3.2 | Influence of age on SL across perceptual modalities

We also examined whether age (8–10 months) influenced learning outcomes. We found no significant correlation between age and Difference Score (looking to novel–familiar test trials) for infants in the Visual condition ($r = 0.10$, $p > 0.7$) but there was a significant correlation of age with Difference Score for infants in the Auditory condition, $r = 0.58$, $t(14) = 2.75$, $p = 0.015$, with older infants exhibiting a stronger Novelty preference (Figure 3). The x-intercept for the relationship between age and Difference Score is at 9 months of age. This finding suggests that there are age-related differences in auditory but not visual SL in this age range.

3.3 | Learning outcomes in relation to statistical information during habituation

There are two benefits of employing infant-controlled habituation: First, the assumption of this method is that when infants have sufficiently encoded the habituation stimuli, they will have a decline in looking time. Thus, each infant should have received the amount of statistical exposure they needed for learning. Additionally, we can quantify the statistical exposure that they have (overtly) attended.

Given that differences in presentation rate were necessary to elicit SL in both perceptual modalities (see Data S1 for control experiment of auditory SL at 1 s SOA), we can examine looking at test relative to the amount of statistical information (e.g., the number of tokens perceived: 2/s for auditory, 1/s for visual; or approximate repetitions of a bigram by dividing the number of tokens perceived by 6 [since 6 is the number of unique tokens]).

Even though there is no significant difference in total viewing time during Habituation ($p > 0.3$), there is a significant difference in the number of tokens perceived during Habituation across perceptual modalities (auditory: $M = 150$ tokens or ~25 repetitions of each bigram, $SD = 13$ repetitions; visual: $M = 91$ tokens or ~15 repetitions of each bigram, $SD = 8.8$ repetitions; $t(28.62) = 2.63$, $p = 0.014$). This amount of statistical exposure is similar to exposure in comparable SL studies (Table 1). However, we conducted several analyses to confirm that difference in statistical exposure does not account for the differences in learning across perceptual modalities.

Most directly, we examined whether including statistical exposure in our omnibus test would explain a significant portion of the variance in the data. Using linear regression, we first confirmed our results from the ANOVA (Perceptual Modality and Test Trial Type interaction: $\beta = -3.41$, $t = -2.80$, $p < 0.001$; main effect of

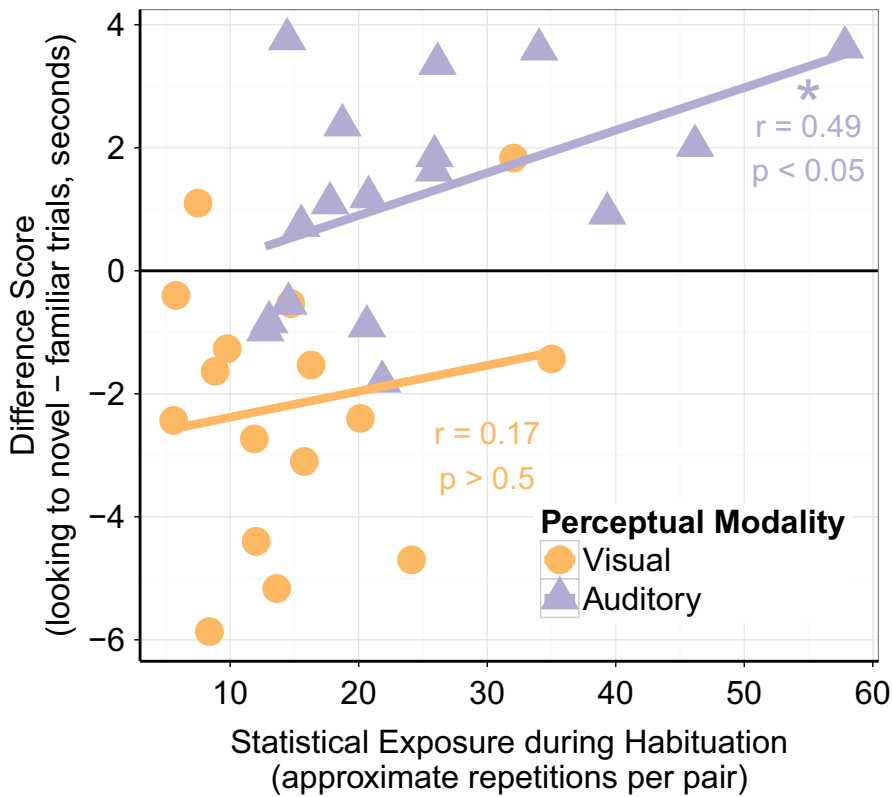


FIGURE 4 Relationship between statistical exposure during habituation and learning outcomes (difference score). Differences in statistical exposure do not explain differences in learning across perceptual modalities

Perceptual Modality: $\beta = 4.22$, $t = 4.89$, $p < 0.0001$) and then compared this base model with a model that includes statistical exposure for each infant. We found that the addition of this factor did not explain any more of the variance ($p = 0.69$), indicating that statistical exposure does not explain a significant portion of the pattern of results and did not affect the significance of the modality by test trial type interaction.

The majority of infants have similar statistical exposure regardless of perceptual modality (Figure 4). Yet, infants exhibit familiarity preferences for the visual modality and a novelty preference for the auditory modality. Moreover, if increasing statistical exposure tends to drive novelty preferences in the visual modality that could suggest that the reduction in statistical exposure might explain the differences in the direction of preference across modalities. However, contrary to this line of reasoning, we found no significant influence of amount of exposure during habituation on difference scores for infants in the Visual condition ($r = 0.17$, $p > 0.5$). There is a significant, positive relationship between amount of exposure and difference scores in the auditory modality ($r = 0.49$, $t(15) = 2.18$, $p = 0.046$). Thus, these two additional analyses confirmed that differences in statistical exposure across perceptual modalities are not driving differences in learning.

Given that viewing time and age both predicted learning outcomes in the Auditory condition, we examined whether age and viewing time were correlated. We found no significant correlation ($p > 0.2$). In addition, we used model comparisons to examine both age and exposure. We report above that exposure does not explain a significant portion of variance above our base model including

modality and test trial type. We found that the same comparison with age shows that age does explain a significant portion of the variance ($\chi^2(1) = 26.21$, $p = 0.033$). Comparing a model with age and a model with age and exposure, we again find that exposure time does not explain any additional variance ($p = 0.99$). Additional analyses revealed no effects on Difference scores of Experimental location, Gender, Bigram set or Test Trial order in either modality condition.

4 | DISCUSSION

This study is the first to directly compare auditory and visual SL in infancy. We choose to compare stimuli that infants frequently experience, that are perceptually complex and become the bases of specialized perceptual processing (i.e., faces and speech). Using these stimuli, we found that auditory SL results in a strong novelty preference while visual SL results in a familiarity preference. We followed the Hunter and Ames model (1988) to interpret these results as weaker learning in the visual compared to the auditory modality. This basis for interpreting differences in the directions of preference is conventional in the infancy literature (e.g., Johnson & Jusczyk, 2001; Jusczyk & Aslin, 1995; Thiessen & Saffran, 2003; though see discussion below). And while uncommon, there is also precedence for finding familiarity preferences in visual learning studies even after infant-controlled habituation (SL: Fiser & Aslin, 2002; visual rule learning: Ferguson, Franconeri, & Waxman, 2018). Finding better auditory SL at this point in infancy dovetails with a decade of research suggesting that,

in adults, auditory SL is stronger than visual SL (e.g., Emberson et al., 2011; Conway & Christiansen, 2005, 2009; Saffran, 2002).

We also found that auditory (speech) SL exhibits a developmental shift at this period of infancy: Infants alter their looking preferences between 8 and 10 months, indicating a change in infants' underlying learning abilities and further suggesting increases in their auditory SL abilities. In particular, our results point to an inflection point around 9 months. No such shift is evident in the visual modality (i.e., there was no change in looking preferences across the age range investigated). Thus, we again find a differential developmental pattern of SL across auditory and visual modalities. Studies of SL in childhood present a convergent picture where visual SL continues to develop into childhood suggesting an earlier development of auditory SL (Raviv & Arnon, 2018).

However, future work is needed before the specifics of these auditory SL changes will be fully understood. For example, a comparison with non-speech auditory stimuli is necessary to determine if this change is specific to speech (or these particular speech stimuli) or is more general. Moreover, given important changes in language and memory development during this time, it would be informative to consider auditory SL in a broader cognitive/developmental context (i.e., Do these changes relate to other changes in language or memory development?). Thiessen and Saffran (2003) also document a change in SL for speech streams between 7 and 9 months with infants shifting their emphasis *away* from statistical information toward stress cues (see Data S1 for further discussion of this topic). Future work is needed to reconcile what appear to be opposite developmental patterns. It could be that when presenting multiple cues in a single stream, the outcome does not reflect learning abilities *per se* but attention to particular cues. To conjecture further, it may be that increases in learning abilities occur alongside decreases in attention because, as more effective learners, attention is less important for encoding those patterns.

Our auditory SL findings in infants are also consistent with previous work suggesting that rate of presentation affected auditory SL in infants as well as adults. Considering the amount of statistical exposure, the use of adult-directed speech, and rate of presentation (factors that can independently modulate learning outcomes), the most comparable study, Thiessen et al. (2005), did not find learning in 8-month-olds. Previous studies that have found auditory SL in younger infants employed both much faster rates of presentation and greater exposure (Table 1). Data S1 present a control study in which slower rates of presentation result in no demonstration of auditory SL for this age group. Thus, we also present initial evidence that auditory SL is related to rate of presentation in infancy with slower rates leading to poorer learning. These results point to a similar relationship between rate of presentation and auditory SL as has been found in adults (Emberson et al., 2011). This relationship between rate and auditory SL suggests that we are finding evidence of better auditory learning in conditions that are not favorable to auditory SL (see Data S1 for considerations of the current results to infants' use of non-statistical speech cues and the role of attention across different types of stimuli).

While this study presents some important, initial findings as to how SL relates across perceptual modalities in infancy, it also highlights the complexity of asking these questions. Here, we address two key issues: First, the dominant method in infancy research (i.e., quantifying looking times to familiar and novel stimuli) is not well-equipped to compare between multiple conditions especially when these conditions vary across stimulus types. While we employed the Hunter and Ames (1988) model to interpret differences in the directionality of looking times, this model has not been broadly validated and may be too simplistic (e.g., see Kidd, Piantadosi, & Aslin, 2012). Moreover, Hunter and Ames has not been used to compare very disparate types of stimuli, as used here (partly because the field has not typically embarked on such comparisons in the first place). Other methods are available but, again, the comparisons between stimulus types or perceptual modalities will be highly complex. For example, functional near-infrared spectroscopy (fNIRS) has been used to investigate learning trajectories (Kersey & Emberson, 2017) and responses to novelty or violations (Emberson, Richards, & Aslin, 2015; Lloyd-Fox et al., 2019; Nakano, Watanabe, Homae, & Taga, 2009). However, comparing between modalities would likely not be straightforward. For example, different modalities likely tap into different neural networks that may vary in availability for measurement, and/or have different spatial or temporal distribution of neural responses that may or may not be related to learning. Indeed, Emberson, Cannon, Palmeri, Richards, and Aslin (2017) used fNIRS to examine repetition suppression (a phenomenon where locally repeated presentation reduces neural responses to particular stimuli) across auditory and visual modalities. That study revealed that the same condition yielded quite different neural responses across modalities even beyond sensory cortices (i.e., differential engagement of the frontal cortex). Here, this paper has erred on the side of a classic interpretation and standard methods, but in order for the field to effectively tackle questions about the mechanisms of learning across perceptual modalities or stimulus types, either a clear way to use the current methods (perhaps in combination) or new methods are needed.

Second, the selection of stimuli is highly complex and importantly constrains the findings. While the selection of stimuli is always important, this is particularly the case when selecting stimuli that are representative of entire perceptual modalities. Given the hypothesized importance of SL to language development, we aimed to provide a direct comparison to speech stimuli. From there, we choose to select a stimulus set from vision that would be similar in terms of an infant's prior experience, the perceptual complexity of the stimulus and emergent specialization of processing for the stimuli. Faces, like speech, are highly familiar to infants, are perceptually complex and are subject to the development of specialized processing. Indeed, faces and speech are analogous stimuli along these dimensions and have been the focus of previous comparisons of development across vision and audition (Maurer & Werker, 2014). However, given that these stimuli are familiar to infants, it is not immediately clear that SL abilities measured here will generalize to all stimuli from the same modality. These complexities of

stimulus selection will be remedied, at least partly, through future work that chooses to compare different types of stimuli (e.g., non-familiar stimuli). However, having principled ways of considering which stimuli to select for comparison would be helpful for future investigations.

Given that we are comparing stimuli that infants have experience with, it is possible their experience before this point is affecting their SL abilities. Indeed, there are now numerous studies that show that infants are tuning themselves to the statistics of their language input in ways that generalize to laboratory tasks (see recent evidence from Orena & Polka, 2017). Given that speech has a strong temporal nature and, here, infants are exposed to temporal statistics, the paradigm may be biased toward auditory SL. This type of finding is consistent with the broader picture that SL is not amodal and has important differences across perceptual modalities and stimulus types. However, it should also be noted that recent work on the visual input of infants has revealed a strong temporal component to early visual input as well (e.g., Sugden & Moulson, 2018, show that young infants see faces in bouts of 1–3 s). Thus, a broader question emerges of how do SL abilities tune themselves to the input that infants receive and are these stimulus- or modality-specific? Comparisons across stimulus types and perceptual modalities will be integral to answering these questions.

Finally, recent work in adults suggests that multisensory SL is an important avenue to be explored (Frost et al., 2015) but very little has been done to this end with developmental populations. Since it is possible that SL in a given modality will be affected by what has been previously learned in another modality, within-subjects designs are a promising way to investigate multisensory SL with infants (see Robinson & Sloutsky,). Relatedly, it should be noted that researchers will need to carefully investigate carry-over effects and multisensory interactions in SL if they wish to use within-subjects designs (see also Charness, Gneezy, & Kuhn, 2012, for a preference of between-subjects designs for experimental questions like these).

In sum, the goal of this work was to provide the first direct comparison of auditory and visual SL in infancy. We found some initial evidence that, similar to adults, auditory SL yields stronger learning than visual SL (in temporal streams with speech and face stimuli) and that auditory SL is developing early. We provide the first evidence that perceptual information significantly modulates SL in infancy (i.e., that it is not equivalent across perceptual modalities). This finding is crucial because, while statistical information itself is an important driver of learning and development, an infant's experience of the world is mediated by sensory input. Thus, an understanding of how exposure to statistical information gives rise to learning and development must consider whether learning is systematically affected by the stimuli and perceptual modality in which the statistics are embedded. Overall, we suggest that comparisons across modalities and different stimulus types are a useful path to investigate mechanistic questions about SL in development, while raising several important issues for researchers to consider in future work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author (Lauren L. Emberson) upon reasonable request.

ENDNOTES

- ¹ Post hoc power analyses revealed this test to have a power of 1.0 (based on an η^2 of 0.79 calculated from the Sum of Squares for the interaction over the residuals or total). Thus, the power reduction in a between-subjects design due to sampling or subject variability is not an issue as our comparisons are very well powered.
- ² This familiarity effect is in contrast to Kirkham et al. (2002) whose paradigm is closely mirrored here. However, the change in complexity of the current visual stimuli (faces) from the abstract, geometric shapes employed by Kirkham et al. (2002) provides an explanation for this change from novelty to familiarity preference that again is well-supported by the Hunter and Ames (1988) model.

REFERENCES

- Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, 14(3), 464 – 473. <https://doi.org/10.1111/j.1467-7687.2009.00937.x>
- Aslin, R. N., & Fiser, J. (2005). Methodological challenges for understanding cognitive development in infants. *Trends in Cognitive Science*, 9(3), 92 – 98. <https://doi.org/10.1016/j.tics.2005.01.003>
- Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321 – 324. <https://doi.org/10.1111/1467-9280.00063>
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject designs. *Journal of Economic Behavior and Organization*, 81, 1 – 8. <https://doi.org/10.1016/j.jebo.2011.08.009>
- Cohen, L., Atkinson, D., & Chaput, H. (2000). *Habit 2000: A new program for testing infant perception and cognition*. [Computer software]. Austin, TX, the University of Texas.
- Conway, C., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(1), 24–38. <https://doi.org/10.1037/0278-7393.31.1.24>
- Conway, C., & 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. <https://doi.org/10.1080/09541440802097951>
- Creel, S., Newport, E., & Aslin, R. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1119–1130. <https://doi.org/10.1037/0278-7393.30.5.1119>
- Emberson, L. L., Cannon, G., Palmeri, H., Richards, J. E., & Aslin, R. N. (2017). Using fNIRS to examine occipital and temporal responses to stimulus repetition in young infants: Evidence of selective frontal cortex involvement. *Developmental Cognitive Neuroscience*, 23, 26–38. <https://doi.org/10.1016/j.dcn.2016.11.002>
- Emberson, L. L., Conway, C., & Christiansen, M. H. (2011). Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, 64(5), 1021–1040. <https://doi.org/10.1080/17470218.2010.538972>
- Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain: Learning-induced expectations rapidly

- affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences*, 112(31), 9585–9590. <https://doi.org/10.1073/pnas.1510343112>
- Ferguson, B., Franconeri, S. L., & Waxman, S. R. (2018). Very young infants abstract rules in the visual modality. *PLoS ONE*, 13(1), e0190185. <https://doi.org/10.1371/journal.pone.0190185>
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99(24), 15822–15826. <https://doi.org/10.1073/pnas.232472899>
- Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, 117, 107–125. <https://doi.org/10.1016/j.cognition.2010.07.005>
- 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. <https://doi.org/10.1016/j.tics.2014.12.010>
- Graf Estes, K., Evans, J., Alibali, M., & Saffran, J. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, 18(3), 254–260. <https://doi.org/10.1111/j.1467-9280.2007.01885.x>
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, 5, 69–95.
- Johnson, E., & Jusczyk, P. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548–567. <https://doi.org/10.1006/jmla.2000.2755>
- Jusczyk, P., & Aslin, R. (1995). Infants detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1), 1–23. <https://doi.org/10.1006/cogp.1995.1010>
- Kersey, A. J., & Emberson, L. L. (2017). Tracing trajectories of audio-visual learning in the infant brain. *Developmental Science*, 20(6), e12480. <https://doi.org/10.1111/desc.12480>
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, 7(5), e36399. <https://doi.org/10.1371/journal.pone.0036399>
- Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35–B42. [https://doi.org/10.1016/S0010-0277\(02\)00004-5](https://doi.org/10.1016/S0010-0277(02)00004-5)
- Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, 3, 1–11.
- Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606–608. <https://doi.org/10.1126/science.1736364>
- Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multi-sensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, 13, 470–8.
- Lloyd-Fox, S., Blasi, A., McCann, S., Rozhko, M., Katus, L., Mason, L., ... Elwell, C. E. (2019). Habituation and novelty detection fNIRS brain responses in 5- and 8-month-old infants: The Gambia and UK. *Developmental Science*, e12817. <https://doi.org/10.1111/desc.12817>
- Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, 56, 154–178.
- McMullen, E., & Saffran, J. (2004). Music and language: A developmental comparison. *Music Perception*, 21(3), 289–311. <https://doi.org/10.1525/mp.2004.21.3.289>
- Misyak, J. B., Goldstein, M. H., & Christiansen, M. H. (2012). Statistical-sequential learning in development. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 13–54). Berlin, Germany: Mouton de Gruyter.
- Nakano, T., Watanabe, H., Homae, F., & Taga, G. (2009). Prefrontal cortical involvement in young infants' analysis of novelty. *Cerebral Cortex*, 19(2), 455–463. <https://doi.org/10.1093/cercor/bhn096>
- Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, 10(1–2), 3–18. <https://doi.org/10.1002/icd.239>
- Orena, A. J., & Polka, L. (2017). Segmenting words from bilingual speech: Evidence from 8- and 10-month-olds. *The Journal of the Acoustical Society of America*, 142(4), 2705. <https://doi.org/10.1121/1.5014873>
- Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences*, 102, 5297–5300.
- Pelucchi, B., Hay, J., & Saffran, J. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80(3), 674–685. <https://doi.org/10.1111/j.1467-8624.2009.01290.x>
- Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, 21(4), e12593. <https://doi.org/10.1111/desc.12593>
- Robinson, C. W., & Sloutsky, V. M. (2007). Visual statistical learning: Getting some help from the auditory modality. In D. S. M. J. G. Trafton (Ed.), *Proceedings of the 29th annual cognitive science society* (pp. 611–616). Austin, TX: Cognitive Science Society.
- Romberg, A., & Saffran, J. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906–914. <https://doi.org/10.1002/wcs.78>
- Saffran, J. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47(1), 172–196. <https://doi.org/10.1006/jmla.2001.2839>
- Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27–52. [https://doi.org/10.1016/S0010-0277\(98\)00075-4](https://doi.org/10.1016/S0010-0277(98)00075-4)
- Saffran, J., Newport, E., Aslin, R., Tunick, R., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8(2), 101–105. <https://doi.org/10.1111/j.1467-9280.1997.tb00690.x>
- Saffran, J., & Thiessen, E. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39(3), 484–494. <https://doi.org/10.1111/j.1467-9280.1997.tb00690.x>
- Sloan, L. K., Kim, H. I., & Johnson, S. P. (2015). Infants track transition probabilities of continuous sequences of faces but not shapes. Poster presented at the Biennial Conference of the Society for Research in Child Development, Philadelphia, PA.
- Sugden, N., & Moulson, M. (2018). These are the people in your neighborhood: Consistency and persistence in infants' exposure to caregivers', relatives', and strangers' faces across contexts. *Vision Research*. (in press). <https://doi.org/10.1016/j.visres.2018.09.005>
- Thiessen, E. D., & Erickson, L. C. (2013). Beyond word segmentation: A two process account of statistical learning. *Current Directions in Psychological Science*, 22, 239–243. <https://doi.org/10.1177/0963721413476035>
- Thiessen, E., Hill, E., & Saffran, J. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53–71. https://doi.org/10.1207/s15327078in0701_5
- Thiessen, E., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries in 7- to 9-month olds. *Developmental Psychology*, 39(4), 706–716. <https://doi.org/10.1037/0012-1649.39.4.706>

- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., ...Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168(3), 242–249. <https://doi.org/10.1016/j.psychres.2008.05.006>
- Turk-Browne, N., Jungé, J., & Scholl, B. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology-General*, 134(4), 552–563. <https://doi.org/10.1037/0096-3445.134.4.552>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Emberson LL, Misyak JB, Schwade JA, Christiansen MH, Goldstein MH. Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism(s). *Dev Sci*. 2019;22:e12847. <https://doi.org/10.1111/desc.12847>