



## PAPER

## Tracing trajectories of audio-visual learning in the infant brain

Alyssa J. Kersey<sup>1</sup> and Lauren L. Emberson<sup>2</sup>

1. Department of Brain and Cognitive Sciences, University of Rochester, USA

2. Department of Psychology, Princeton University, USA

**Abstract**

*Although infants begin learning about their environment before they are born, little is known about how the infant brain changes during learning. Here, we take the initial steps in documenting how the neural responses in the brain change as infants learn to associate audio and visual stimuli. Using functional near-infrared spectroscopy (fNIRS) to record hemodynamic responses in the infant cortex (temporal, occipital, and frontal cortex), we find that across the infant brain, learning is characterized by an increase in activation followed by a decrease. We take this U-shaped response as evidence of repetition enhancement during early stages of learning and repetition suppression during later stages, a result that mirrors the Hunter and Ames model of infant visual preference. Furthermore, we find that the neural response to violations of the learned associations can be predicted by the shape of the learning curve in temporal and occipital cortex. These data provide the first look at the shape of the neural response during audio-visual associative learning in infancy establishing that diverse regions of the infant brain exhibit systematic changes across the time-course of learning.*

**Research highlights**

- This study uses fNIRS to trace the patterns of activation during associative learning in infants.
- We find that, across multiple regions of the brain, infants show evidence of neural enhancement in the early stages of learning, followed by neural suppression in later stages of learning.
- These results mirror the canonical Hunter and Ames model of infant visual preference and may represent a neural signature that relates to infant looking preferences.
- Exploratory analyses find that the degree to which the infant brain exhibits these patterns can be used to predict the brain's response to a violation of the expectations built during learning.

**Introduction**

Decades of research have used behavioral methods to study infant learning and have established that human

infants are excellent learners in many circumstances (recent prominent examples, Gomez & Gerken, 1999; Marcus, Vijayan, Rao & Vishton, 1999; Saffran, Aslin & Newport, 1996; Smith & Yu, 2008; Stager & Werker, 1997). While this work firmly establishes that early learning has the power to shape neural and cognitive development, virtually nothing is known about how the infant brain changes during learning. The current paper examines how engaging in a complex learning task (audio-visual associative learning) shapes neural activity in the infant brain.

A common method for studying learning is to expose participants to some pattern of stimuli and then compare responses to stimuli that are consistent vs. those that differ from the previously experienced patterns. These studies provide evidence about what infants have learned, how they recognized previously learned stimuli, and how infants track novel information in the environment (in addition to the behavioral studies referenced above, see Nakano, Watanabe, Homae & Taga, 2009, for an excellent study of infant auditory novelty responses using fNIRS). However, measuring responses to novel stimuli is likely not targeting the same cognitive and

Address for correspondence: Lauren L. Emberson, 227 Peretsman Scully Hall, Department of Psychology, Princeton University, Princeton, NJ 08540, USA; e-mail: lauren.emberson@princeton.edu

neural systems that supported learning initially (see Karuza, Emberson & Aslin, 2014, for a relevant review of this distinction).

Importantly, the current paper diverges methodologically from previous work and does *not* examine neural activity during violations to a recently learned structure (e.g. Emberson, Richards & Aslin, 2015; Gervain, Macagno, Cogoi, Peña & Mehler, 2008). In contrast, the current paper examines neural activity starting from the first exposure to a novel audio-visual pairing and traces neural activity from this first exposure until the infant has become bored with or habituated to the stimuli. We then employ regression-based statistical methods to uncover the shape of neural responses over successive experiences, or repeated exposures, to the audio-visual stimuli (i.e. over the learning task or across the learning trajectory).

While there is little direct evidence for what neural changes correlate with learning trajectories in infancy, there is support in the literature for three possible patterns that might be observed. One possibility is that repeated exposures during the learning task will result in reductions of neural activity. Indeed, the phenomenon of *repetition suppression* has been well documented in the adult brain (sometimes called fMRI adaptation; see review in Grill-Spector, Henson & Martin, 2006). Moreover, repetition suppression appears to mirror the well-known behavioral phenomenon in the developmental literature of *habituation* where infants exhibit decreased looking time to familiar stimuli over successive exposures (Mather, 2013; Turk-Browne, Scholl & Chun, 2008). Thus, we might expect that infants who are repeatedly exposed to the same stimulus during a learning task will show decreases in neural activity in regions of the brain involved in learning. In fact, this is supported by Nakano and colleagues who found that the frontal cortex in 3-month-old infants exhibits a decrease in neural activity during habituation to a one syllable speech sound (Nakano *et al.*, 2009).

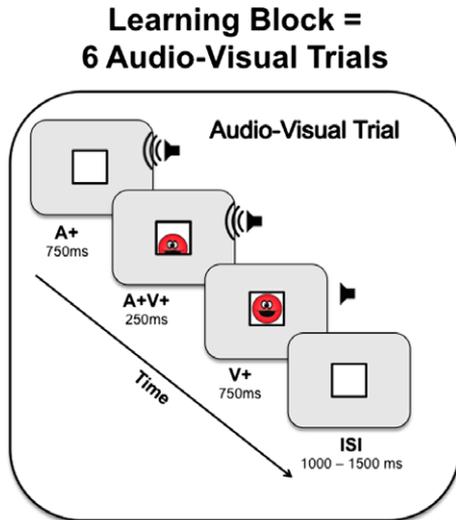
However, studies reporting repetition suppression effects tend to employ very simple stimuli (e.g. unrelated visual images, Grill-Spector, Kushnir, Edelman, Avidan, Itzhak *et al.*, 1999; one syllable speech sounds, Nakano *et al.*, 2009) and contexts where there is not much to be learned as a result of repetition. Thus, it is possible that repetition suppression arises from a neural disengagement as a result of very little information to learn from the environment. Relatedly, recent work has suggested that infants selectively attend or prolong their looking to situations where stimuli are learnable but, crucially, disengage their attention to rote repetition of a stimulus (Gerken, Balcomb & Minton, 2011; Kidd, Piantadosi & Aslin, 2012), suggesting that habituation too might be

driven in part by the simplicity of the stimuli infants are exposed to. These behavioral studies suggest that viewing simple stimuli will result in a different pattern of neural activity than experience with more complex stimuli endowed with learnable patterns. In fact, Turk-Browne, Yi, Leber and Chun (2007) found that the repetition of more perceptually complex or difficult to discriminate stimuli results in *repetition enhancement* in the adult brain. Thus, a second possibility is that learning will result in increases in neural activity. Indeed, previous research investigating neonate neural responses to learnable vs. non-learnable sequences of syllables (ABB vs. ABC rules) reported enhancement in neural activity to the learnable sequence and no change in the neural response to the unlearnable sequence (Gervain *et al.*, 2008; however, also see Wagner, Fox, Tager-Flusberg & Nelson, 2011).

Yet a third possibility is that regions involved in learning will show evidence of both neural suppression and enhancement. If repetition suppression and enhancement are related to attention and infant looking preferences, then a familiarity preference would be mirrored by greater activation to the repeated stimulus (neural enhancement), while a novelty preference would suggest a decreased response to the learning stimulus (neural suppression). Indeed, the seminal model of infant habituation, Hunter and Ames (1988), proposes that early stages of learning or encoding result in familiarity preferences, while later stages of learning result in novelty preferences.<sup>1</sup> A recent study in adults found evidence that perceptually complex conditions (similar to the conditions that resulted in repetition enhancement in Turk-Browne *et al.*, 2007) lead to initial repetition enhancement followed by repetition suppression (Muller, Strumpf, Scholz, Baier & Melloni, 2013). Therefore, a distinct and third possibility is that the infant brain could exhibit neural enhancement during early stages of learning and neural suppression during later ones (i.e. a non-linear relation between neural activation and stages of learning).

In the present study, we examined neural activity in 6-month-old infants using fNIRS during an audio-visual learning task. During this learning task, infants viewed successive blocks of audio-visual (AV) events. Each AV event paired a complex sound (either a squeaky horn or a rattle) with the appearance of a red, cartoon smiley face (Figure 1). Crucially, it has been well established that infants at this age can learn AV pairings under

<sup>1</sup> The Hunter and Ames model includes a 'task difficulty' factor that could potentially accommodate all three possibilities. However, to see only repetition suppression or repetition enhancement, learning would have to be either extremely simple or extremely complex.

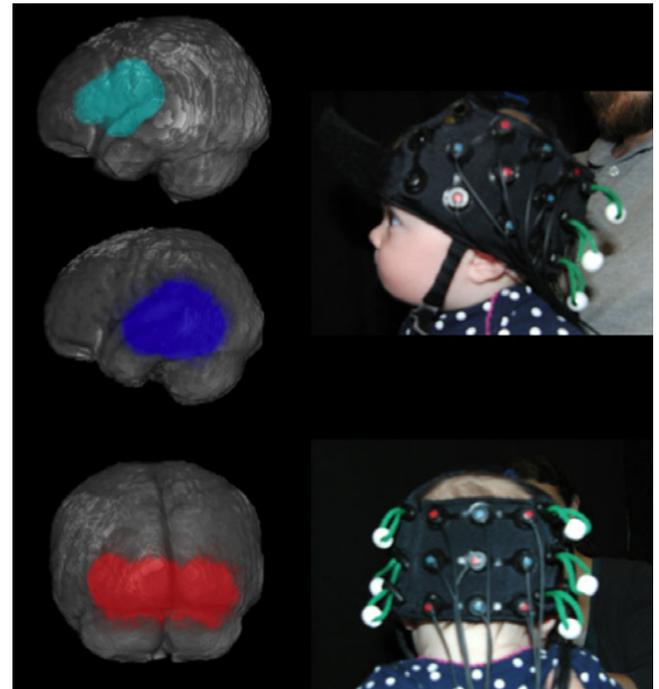


**Figure 1** Depiction of an audio-visual trial in a learning block. A learning block consisted of two separate audio-visual pairings: one sound was paired with the smiley face entering the box from the top, and a second sound was paired with the smiley face entering from the bottom of the box. Each trial was then repeated twice and present in a random order, resulting in six audio-visual trials per learning block.

similar experimental conditions (Gogate & Bahrick, 1998). Accordingly, the current study presumes that infants are engaging in AV learning during exposure and addresses the question of how neural activity changes over the time-course of learning.

Neural activity was recorded during learning using fNIRS, a neuroimaging modality that records the same physiological signal as fMRI (i.e. changes in blood oxygenation arising from neural activity) using near-infrared light. Unlike fMRI, this method permits relatively easy recording of these neural signals when infants are awake and learning (see Aslin, Shukla & Emberson, 2015). We recorded activity in three regions of interest (ROIs): the temporal lobe, occipital lobe and prefrontal cortex (Figure 2). ROIs were defined based on co-registration of the NIRS recordings with age-appropriate infant MR templates. Details of this method are reported in Emberson *et al.* (2015).

To trace the relation between learning and neural activity in three regions of the infant brain, neural responses to each successive block of AV experience during learning were estimated. Then, we employed regression-based statistical methods to determine the most robust relation between successive experiences and neural activity. Specifically, we investigated whether there is a significant change to the neural responses over learning and, if so, whether there is a linear decrease



**Figure 2** Left panel: Three regions of interest (ROI). The frontal ROI (teal) comprised two channels. The temporal ROI (blue) comprised five channels. The occipital ROI (red) comprised three channels. See Emberson, Richards and Aslin (2015) for details on the NIRS-MR co-registration supporting the creation of these ROIs. Right panel: Each infant who participated in the study was photographed to help determine the relation between the NIRS optodes and anatomical markers.

or increase (repetition suppression and repetition enhancement, respectively) or whether there is a significant non-linear relation between successive learning experiences suggesting a combination of repetition enhancement and suppression.

## Materials and methods

### Participants

Twenty-six infants, aged 5–7 months, were recruited from the database of interested participants for the Rochester Baby Lab. Infants were born no more than 3 weeks before their due date, had no major health problems, surgeries, history of ear infections, nor known hearing or vision difficulties. Out of 26 infants recruited, nine infants (34% of the total sample) were excluded for failing to watch the video to criterion ( $n = 5$ ), poor optical contact ( $n = 3$ , e.g. due to too much dark hair, see below for specific definitions for each of these exclusionary criteria) and

experimenter error ( $n = 1$ ). The final sample was 17 infants (mean age = 5.6 months,  $SD = 0.6$ , range = 5.0–6.9, 12 female, race: 14 white, three other or more than one, and no Hispanic infants). These are the same infants from the experimental group reported in Emberson *et al.* (2015).

#### *Stimulus presentation and experimental procedure*

The stimuli and stimulus presentation were the same as the experimental group reported in Emberson *et al.* (2015). Stimuli are included in the Supplementary Materials of this paper. All trials began with the presentation of a monochromatic grey screen with a white box (black bordered) presented in the middle. The box was 15.2 degrees of visual angle squared ( $11.5 \text{ cm}^2$  with the infant sitting approximately 43 cm from the screen). Immediately after, combinations of auditory and visual stimuli were presented. Auditory stimuli consisted of novel, non-speech auditory sounds that are similar to a squeaky honk from a clown horn and an unusual rattle sound. Visual stimuli consisted of a red cartoon smiley face that entered the screen from either the top or the bottom of a white box, moved into the box to touch the opposite side in 500 ms, and then exited the box in the same surface that it entered from in another 500 ms. Each sound was consistently and uniquely paired with one direction of movement for the visual stimulus, creating two pairs of audio-visual stimuli. Here infants could be learning either that a specific sound predicts a specific direction or location of the smiley face or infants could be learning that the presentation of a sound predicts the appearance of the smiley face. In either case, infants are associating previously unrelated auditory and visual information and engaging in cross-modal learning. The auditory stimulus was presented at the onset of each event for 1000 ms. The visual stimulus began 750 ms after the onset of the auditory stimulus and also lasted for 1000 ms. This resulted in overlap between these stimuli for 250 ms. Individual pairs of audio-visual stimuli were presented with equal frequency within each learning block and in randomized order. Stimuli were presented on a Tobii 1750 eye tracker screen measuring 33.7 by 27 cm and computer speakers placed directly below the screen but behind a black curtain. Sounds were presented between 64 and 67 dB using MATLAB for Mac (R2007b) and Psychtoolbox (3.0.8 Beta, SVN revision 1245).

The experiment included two types of stimulus presentation: learning blocks and single trials. Learning blocks consisted of six audio-visual events: three of each specific pairing, randomly ordered and each separated by a jittered interstimulus interval, ISI, (1–1.5 seconds, a jittered ISI was also included after the last stimulus of a

block; thus after every audio-visual stimulus there was an ISI of 1–1.5 seconds). Between blocks, baseline stimuli were presented (dimmed fireworks video, Watanabe, Homae, Nakano & Taga, 2008), and a calming instrumental version of ‘Camptown Races’ Baby Music (album released 2010) for a jittered inter-block interval, IBI, of 4–9 seconds (mean = 6.5 seconds). This IBI length has been previously shown to be sufficiently long to allow the neural response in infants to return to baseline (Plichta, Heinzel, Ehli, Pauli & Fallgatter, 2007).

The experiment started with the presentation of three learning blocks. Then, single event trials were presented intermittently between learning blocks. The same IBI separates single trials and learning blocks to separate neural signals to each. Single trials consisted of either an audio-visual event as seen six times in a learning block (one of the audio-visual pairings; AV) or the audio presented without the cartoon smiley face (visual omission trials; AV–). That is, the auditory stimulus began at the onset of the trial, but the cartoon smiley face did not move in and out of the white rectangle (see Figure S1). After the initial three learning blocks, four single trials were presented per subsequent learning block in randomized temporal order with two of each single trial type (equal frequency). The primary analyses here focus on the learning blocks; see Emberson *et al.* (2015) for details and analyses of the single trials. In an exploratory analysis presented after the primary analyses, data from visual omission trials served as the dependent measure in an exploratory analysis to index the strength of infants’ learning and provide a link between these two types of analysis.

The experiment was conducted in a darkened room with dark floor-to-ceiling curtains surrounding the infant and their caregiver with only the monitor visible. Infants sat on their caregivers’ laps. Caregivers were instructed not to interfere with the infant’s watching of the video but to make sure that they did not grab at the cap on their head (see next section) or rub up against them with the cap to move it. We also asked that they encourage the infant to be as still as possible but to allow the infant to move and stand up if it was necessary to keep the infant contentedly watching the video. The researchers watched the caregiver and infant from a video camera underneath the monitor.

#### *fNIRS recordings*

fNIRS recordings were conducted using a Hitachi ETG-4000 with 24 possible NIRS channels: 12 over the back of the head to record bilaterally from the occipital lobe, and 12 over the left side of the head to record from the left temporal lobe and prefrontal cortex. The channels were organized in two  $3 \times 3$  arrays, and the cap was

placed so that, for the lateral array, the central optode on the most ventral row was centered over the left ear and, for the rear array, the central optode on the most ventral row was centered between the ears and over the inion. This cap position was chosen based on which NIRS channels were most likely to record from temporal and occipital cortex in infants. Due to curvature of the infant head, a number of channels did not provide consistently good optical contact across infants (the most dorsal channels for each pad). We did not consider the recordings from these channels in subsequent analyses and only considered a subset of the channels (seven for the lateral pad over the ear and five for the pad at the rear array).

fNIRS recordings were collected at 10 Hz (every 100 ms). Using a serial port, marks were presented from MATLAB on the stimulus presentation computer to the Hitachi ETG-4000 using standard methods. Marks were sent for the start and end of each presentation type for the given experiment (e.g. blocks of AV trials). The raw data were exported from the Hitachi ETG-4000 to MATLAB (version 2006a for PC) for subsequent analyses with HomER 1 (Hemodynamic Evoked Response NIRS data analysis GUI, version 4.0.0) using the default preprocessing pipeline of the NIRS data. First, the raw intensity data are normalized to provide a relative (percent) change by dividing by the mean of the data (HomER 1.0 manual), thus any change from zero is meaningful and does not require an explicit baseline period. Then the data were low-pass and high-pass filtered (two separate steps) to remove high frequency noise such as Meyers waves and low frequency noise such as changes in blood pressure. Second, changes in optical density were calculated for each wavelength, and a PCA analysis was employed to remove motion artifacts. Finally, the modified Beer-Lambert law was used to determine the changes in ( $\Delta$ ) concentration of oxygenated and deoxygenated hemoglobin for each channel (the DOT.data.dConc output variable was used for subsequent analyses, see the HomER Users Guide for full details; Huppert, Diamond, Franceschini & Boas, 2009). Timing information (mark identity and time received by the ETG-4000 relative to the fNIRS recordings) was also extracted from the ETG-4000 data using custom scripts run in MATLAB R2007b.

### *fNIRS data analysis*

Extraction of the mean hemodynamic response for each learning block was conducted in MATLAB (version R2013a) with custom analysis scripts. First, we removed any additional motion artifacts. Following Lloyd-Fox, Blasi, Volein, Everdell, Elwell *et al.* (2009), an algorithm

was written such that for each trial when the concentration (either oxygenated or deoxygenated hemoglobin) was greater or less than  $\pm 5$  mM mm, the duration of the signal over  $\pm 5$  mM mm was determined. Then moving forward and backward, the algorithm searched for the place where the signal changes sign, if there had been 15 consecutive points where the slope between points has been less than 0.5 (indicating that the rapid change in signal has ceased), and if neither of these criteria are met, when there have been more than 200 points since the signal was greater or less than  $\pm 5$  mM mm. The benefits of this algorithm are that it does not rely on potentially biased and non-replicable 'hand-coding' of motion, and that it not only determines when motion was likely to have occurred (signal values beyond  $\pm 5$  mM mm), but also identifies the places where motion likely started (the steep rise or fall of the signal). An appropriate analogy is that you might use the peaks to identify mountains, but if you want to remove them, you need to also find out where their bases are. Once the segments of signal that were likely contaminated by motion were identified, they were removed by zeroing the signal. This is an appropriate method because the signal has been normalized in homER and has the benefit of maintaining timing and 'complete' data collection during the trial. This method was applied to all infants included in the study, resulting in an average of 0.58% of the data being excluded due to motion ( $SD = 0.72\%$ ; six infants had no data excluded, the maximum excluded for a single infant is 1.88%).

Next, the continuous data were segmented and sorted into individual trial types based on the timing of marks. Because the experiment ended when the infant became inattentive or fussy, we excluded any blocks at the end of the experiment that the infant did not fully attend to. Infants were required to watch a minimum of five AV blocks to be included in the experiment (see Participants for the number of infants excluded for not watching a sufficient amount of time). Five blocks was chosen as a cut-off to ensure that enough neural data were available for each infant. After exclusion, infants looked on average for 6.29 learning blocks (minimum = 5 blocks, maximum = 8 blocks,  $SD = 0.99$ , 5 blocks:  $n = 4$ , 6 blocks:  $n = 6$ , 7 blocks:  $n = 5$ , and 8 blocks:  $n = 2$ ).

Then, for each infant, the average concentration of oxygenated hemoglobin per channel was determined for each condition. Infants were excluded at this point if the data collected were still noisy. Noisiness of the data was based on a combination of visual inspection, notes on optical contact and the presence of hair, and output from the *otparex.m* script, which provided a measure of the number of bad channels. These infants were not excluded at the point where meaningful average information could

be seen in order to minimize experimenter bias to include or exclude participants who confirmed or denied experimental hypotheses. Moreover, the decision to include or exclude infants was made before group averages were determined and was not revisited.

Finally, the average and variance of responses for oxygenated hemoglobin were determined within each ROI for each infant for the 31 seconds of the learning block (defined a priori to start with initial stimulus presentation and continue into the jittered ISI interval to capture the entire hemodynamic response to the learning block). Analyses were conducted on the mean hemodynamic responses in RStudio (version 0.98) using the lmer (Bates, Maechler, Bolker & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff & Christensen, 2013) packages for R (version 3.1.1).

## Results

Our goal was to identify differences in neural activity (measured by functional changes in blood oxygenation) across learning blocks. The magnitude of the hemodynamic response for each learning block for each infant (average of the normalized changes in blood oxygenation from 0 to 31 seconds from the onset of the first stimulus) was calculated in each of our three regions of interest (ROIs): temporal cortex (five NIRS channels), occipital cortex (three NIRS channels), and frontal cortex (two NIRS channels). Our analyses unfolded in three stages. First, we tested for a linear relation between learning blocks and blood oxygenation. Next, we tested for a nonlinear relation between learning blocks and blood oxygenation. These analyses are primarily regression analyses, but we also include complementary post-hoc *t*-tests to clarify the main effects. Finally, we conducted exploratory analyses to test for a relation between individual differences in the shape of the learning trajectory and learning outcomes. Making connection to the broader literature, we used neural response to violations of the learned association as our measure of learning outcomes.<sup>2</sup>

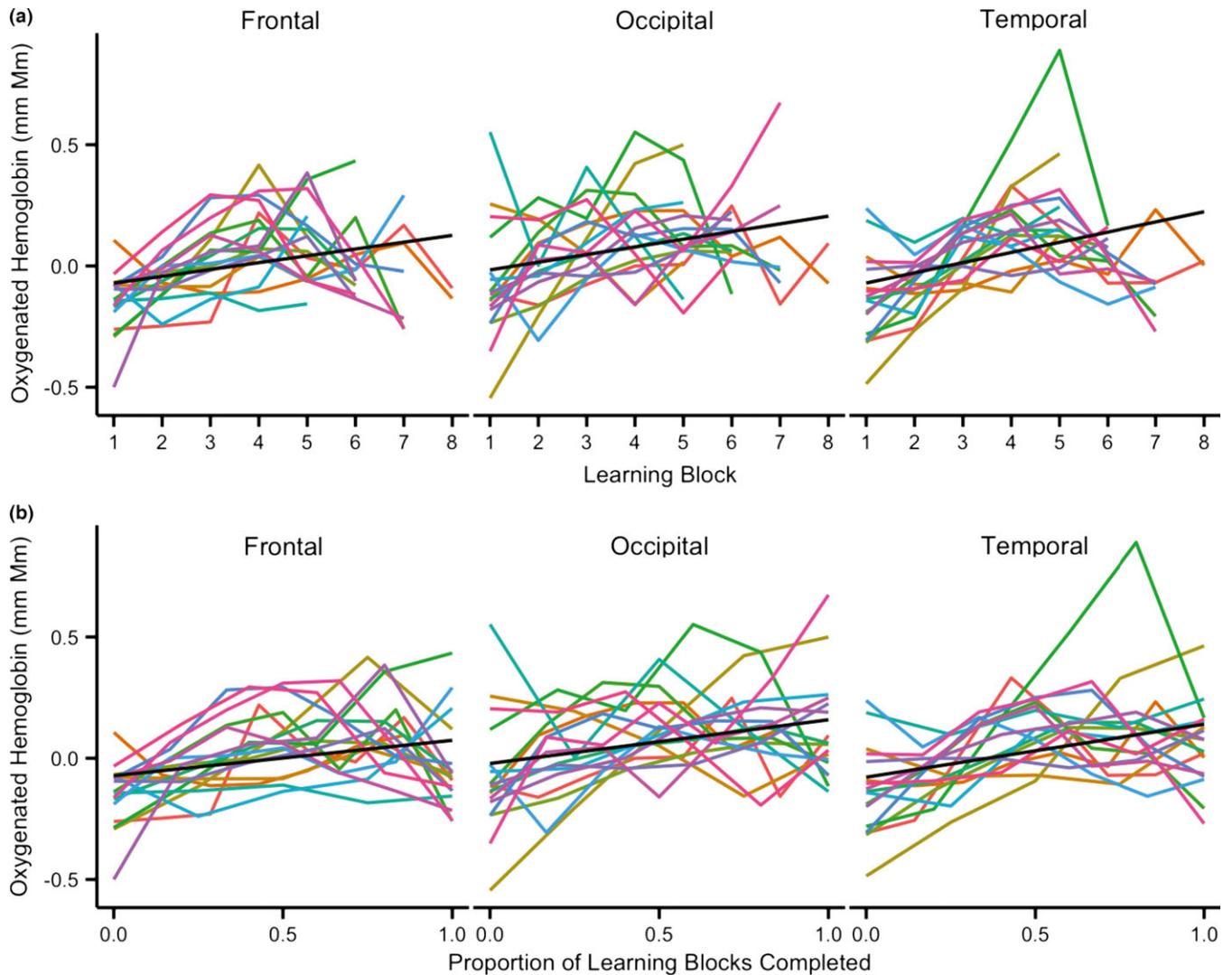
<sup>2</sup> Specifically, we measured neural activity during the presentation of an auditory cue without a visual cue (visual omission), which allowed us to measure how well infants were able to learn that auditory cues generally predict visual cues. Neural activity to a novel visual event could arise from a number of sources (e.g. a release from low-level visual adaptation vs. a violation of the audio-visual pair). However, the omission of visual information cannot lead to neural changes separate from the expectations produced by the audio cue and therefore this is a suitable measure of learning at the same level of specificity as the neural data during familiarization.

### Linear modeling of learning trajectory

To determine whether there was a linear relation between successive blocks and blood oxygenation, we fit linear mixed effects models to our data. We predicted mean oxygenation in each ROI using block number as the predictor. In each model, we also included random factors of infant and infant by block (e.g.  $1 + \text{Block} | \text{Infant}$ ) to control for any differences across participants and to consider changes by block as a within-Infant comparison (i.e. not treating each successive block as an independent observation). The linear mixed effect models revealed a main effect of block in both the temporal and frontal ROIs (*Temporal*:  $\beta = 0.04$ ,  $t = 2.61$ ,  $p^3 = .02$ ; *Frontal*:  $\beta = 0.03$ ,  $t = 2.49$ ,  $p = .03$ ; *Occipital*:  $\beta = 0.03$ ,  $t = 1.66$ ,  $p = .12$ ). The fixed effect of block in temporal and frontal ROIs indicates that the mean oxygenation increased across successive learning blocks of the experiment (Figure 3). The results from the occipital ROI are suggestive of a linear effect but do not reach significance in this model.

We further validated this increase in oxygenation across learning blocks by comparing the amount of oxygenation at the beginning and ending of learning. Because each infant watched between five and eight blocks, we performed two *t*-tests in each ROI (Figure 4): The first compared the oxygenation during the first block of learning to the oxygenation during the fifth block of learning, which was the minimum number of learning blocks completed by all infants (4/17 infants completed only five blocks); the second compared oxygenation during the first block of learning to oxygenation during each infant's last block of learning. This latter test allowed us to take into account that some infants may take longer than others to learn the audio-visual pairing and thus a longer learning time-course could result in a more lengthy change in the neural signal. In the temporal cortex, both the *t*-test comparing block 1 to block 5 and the *t*-test comparing block 1 to each infant's final block were significant (block 1 vs. block 5: block 1 mean =  $-0.13$ , block 5 mean =  $0.18$ ,  $t(16) = -3.65$ ,  $p < .01$ ; block 1 vs. end block: end block mean =  $0.05$ ,  $t(16) = -2.66$ ,  $p = .02$ ), indicating that oxygenation increased across the learning blocks. In the occipital blocks, the *t*-test comparing block 1 to block 5 was significant (block 1 mean =  $-0.08$ , block 5 mean =  $0.12$ ,  $t(16) = -2.13$ ,  $p < .05$ ) but, although the *t*-test comparing block 1 to each end infant's end block

<sup>3</sup> Because degrees of freedom are difficult to estimate for coefficients in mixed effects models, all *p*-values reported throughout the results section are estimated *p*-values. These values were estimated using the lmerTest package for R.



**Figure 3** Predicted data from linear mixed effect models for data plotted by (a) number of learning blocks and (b) proportion of learning blocks completed (i.e. a normalized learning time-course). Black line: Linear fit of the model; Colored lines: Individual data for each infant included in the analyses. Note: Excluding the infant in green with a sharp peak in the temporal channels did not qualitatively change these results.

revealed a similar pattern, the test was only marginally significant (block 1 mean =  $-0.08$ , end block =  $0.11$ ,  $t(16) = -1.84$ ,  $p = .08$ ). A similar result was found in the frontal region: the comparison of oxygenation in block 1 to block 5 revealed a significant increase over time (block 1 mean =  $-0.15$ , block 5 mean =  $0.08$ ,  $t(16) = -3.65$ ,  $p < .01$ ), and although the comparison of oxygenation in block 1 to each infant's final block revealed a similar pattern, again it was only marginally significant (block 1 mean =  $-0.15$ , end block mean =  $-0.04$ ,  $t(16) = -1.80$ ,  $p = .09$ ). Thus, we see evidence in all three ROIs of linear increases in activation over learning. This evidence is strongest in the temporal ROI where both the comparison between the first and fifth as well as the last blocks

is significant, but in all ROIs we see that activation in the fifth block is higher than the first block.

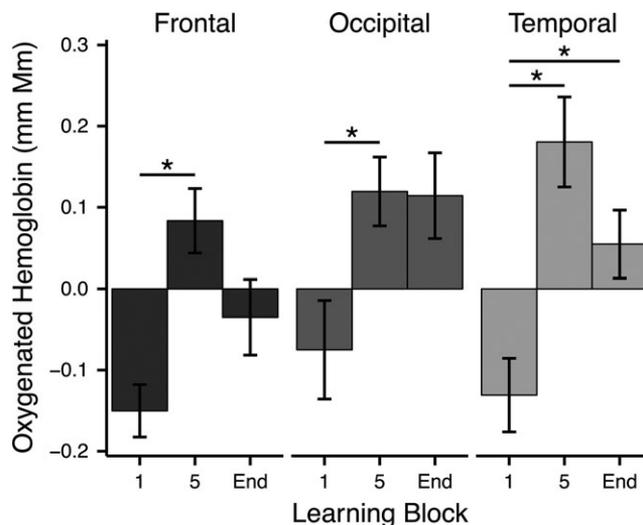
Thus far, we have used convergent statistical methods (mixed effect linear models and  $t$ -test for key points along the time-course) to show that there are systematic increases in activity for three ROIs throughout exposure in a learning task. This is suggestive of a progressive repetition enhancement during learning in an audio-visual associative task in infancy. However, the most robust results were found when comparing the first and the fifth blocks, with only marginal results for both the occipital and frontal ROIs when comparing the first and end blocks. We also find significant linear effects of block in only two of the three ROIs. These results combined

with a visual inspection of individuals' oxygenation curves across learning blocks suggest that perhaps instead of a linear trend, infants exhibited an inverted U-shaped response, characterized by a peak towards the middle (between blocks 4 and 5), followed by a decrease in the remaining blocks.

#### Non-linear modeling of learning trajectory

To examine the possibility of a non-linear relation between learning time-course and neural activity, we used a second mixed effects model to fit mean oxygenation per block with both a linear term of block (as in the previous models) as well as a non-linear term of block (2-degree polynomial). This model allows us to determine whether activation over learning blocks is non-linear and, if so, whether there is both a linear and a non-linear change over time. As with the previous models, we also include a random effect of individual infants and individual infants by block. The results of these models suggest both linear (block) and non-linear (block squared) changes in activation over block in all ROIs (*Temporal*: block:  $\beta = 0.21$ ,  $t = 6.60$ ,  $p < .01$ ; block squared:  $\beta = -0.02$ ,  $t = -5.92$ ,  $p < .01$ ; *Occipital*: block:  $\beta = 0.14$ ,  $t = 3.99$ ,<sup>4</sup>  $p < .01$ ; block squared:  $\beta = -0.01$ ,  $t = -3.72$ ,  $p < .01$ ; *Frontal*: block:  $\beta = 0.18$ ,  $t = 6.00$ ,  $p < .01$ ; block squared:  $\beta = -0.02$ ,  $t = -5.40$ ,  $p < .01$ ). A direct comparison of the simple models with only a linear effect of learning and the current models with both linear and non-linear effects of learning reveal that the models that included a U-shaped response over the course of learning fit significantly better than those models that only account for a linear change (*Temporal*:  $\chi^2 = 30.03$ ,  $p < .001$ ; *Occipital*:  $\chi^2 = 12.95$ ,  $p < .001$ ; *Frontal*:  $\chi^2 = 25.15$ ,  $p < .001$ ). The significantly better fit of the time-course by models with a non-linear term suggests that not only do the responses generally increase over the course of learning, but they also exhibit an inverted U-shaped pattern in which oxygenation peaks (between blocks 4 and 5) and then decreases throughout the remaining blocks (Figure 5). This pattern is consistent with the infant brain exhibiting repetition enhancement during early stages of learning, followed by repetition suppression once encoding or learning is complete. However, given the sharp decrease in the number of infants contributing data after five blocks (13/17 infants contributing data after five blocks), it is possible that the reduction in activity after the five blocks is because of the

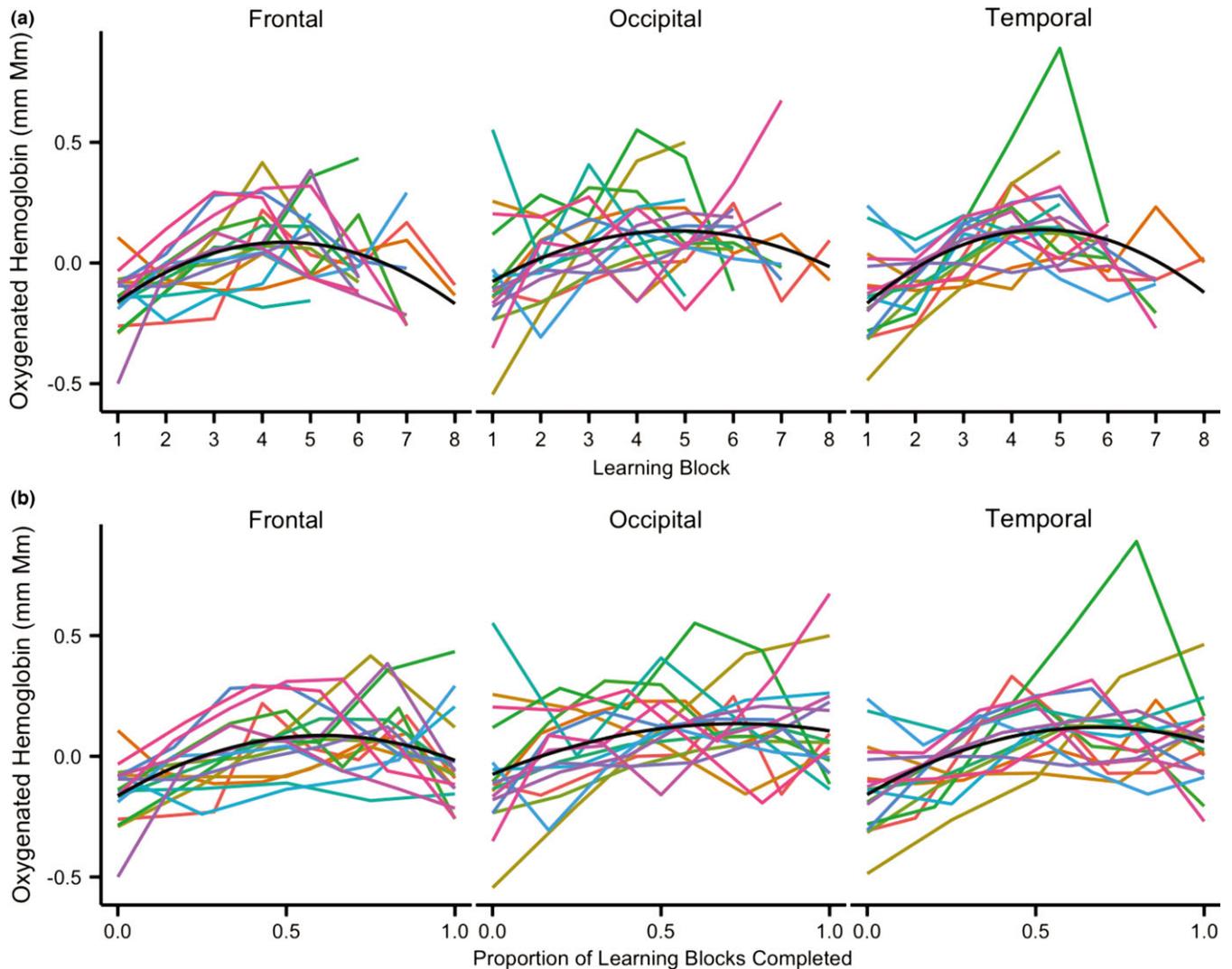
<sup>4</sup> Note that the linear effect of block is now significant for the occipital ROI whereas it was not significant in the model with only the linear term. Similarly, the linear effects for both the temporal and frontal ROIs have much higher  $t$ -values in this model, suggesting independent linear and non-linear effects of block on neural activation.



**Figure 4** Mean activation for key blocks throughout the experiment. Activation was compared for the first block and the fifth block (the last block completed by all infants) and the last block (End) for each individual infant. Asterisks (\*) denote significance at  $p < 0.05$ .

greater uncertainty in the sample. We return to this point in the Discussion.

To validate the inverted U-shaped pattern we conducted additional  $t$ -tests comparing the middle of learning to the beginning and ending of learning. Learning block 4 was selected as the middle of learning because the inverted U-shaped pattern peaked between blocks 4 and 5 and we wished to avoid any instances in which the middle block and end block would be the same for individual infants (i.e. for 4/17 infants block 5 was the final block so we decided *not* to also use it as an index of the middle of learning). These  $t$ -tests revealed a significant increase from the first learning block to the fourth learning block in all ROIs (*Temporal*: block 1 mean =  $-0.13$ , block 4 mean =  $0.16$ ,  $t(16) = -4.02$ ,  $p < .01$ ; *Occipital*:  $t(16) = -2.53$ ,  $p = .02$ , block 1 mean =  $-0.08$ , block 4 mean =  $0.13$ ; *Frontal*: block 1 mean =  $-0.15$ , block 4 mean =  $0.11$ ,  $t(16) = -4.56$ ,  $p < .01$ ). Comparisons of the fourth learning block to each infant's final block (ranging from block 5 to block 8) revealed a significant decrease from block 4 to the final block in the frontal channels (final block =  $-0.04$ ,  $t(16) = 2.28$ ,  $p = .04$ ) and a marginal decrease from block 4 to the final block in the temporal channels (final block mean =  $0.05$ ,  $t(16) = 1.95$ ,  $p = .07$ ). There were no differences between the middle of learning and the end of learning in the occipital channels (final block mean =  $0.11$ ,  $t(16) < 1.0$ ,  $p > .1$ ). These results are in line with the fits seen in Figure 5B, which depict the predicted data from a mixed effect model of mean oxygenation per proportion of learning blocks completed



**Figure 5** Predicted data from mixed effect models with a non-linear (2-degree polynomial) predictor for data plotted by (a) number of learning blocks and (b) proportion of learning blocks completed (i.e. a normalized learning time-course). Black line: Non-linear fit of the model; Colored lines: Individual data for each infant included in the analyses. Note: Excluding the infant in green with a sharp peak in the temporal channels did not qualitatively change these results.

with both a linear term of proportion of learning as well as a non-linear term of proportion of learning squared (2-degree polynomial). Together the results of these *t*-tests and the fits of the data using proportion of learning instead of number of blocks completed suggest that although the data do exhibit an inverted U-shaped pattern, the drop off at the end of learning may be subtle at best in occipital and temporal regions.

#### *Relation between learning trajectory and learning outcomes*

Finally, we conducted an exploratory analysis to determine whether individual differences in the changes of

neural activity across the learning blocks could relate to the changes in neural activity elicited by the novel visual omission trials. As highlighted in the Introduction, the majority of studies of learning focus on the relation between learned stimuli and responses to violations or novel stimuli. As reported in Emberson *et al.* (2015), infants who have learned the AV association exhibit robust occipital lobe responses to the omission of the visual event. We conducted an exploratory analysis to determine whether individual differences in AV learning trajectory (as modeled in this paper) would predict an individual's visual omission response within each neural region (i.e. whether trajectories in the occipital lobe predict the response of the occipital lobe to the visual

omission). Such a relation would lend further support that these neural changes are leading to the development of new representations that an infant is using to process future stimuli. The response to the visual omission trials was obtained by averaging a single analysis time window of 5–9 seconds after stimulus onset and then averaging across all visual omission trials in each ROI (see Emberson *et al.*, 2015, for more detail). We then fit each infant's data to a linear model with two terms (block and block-squared, see Figure S2 for a depiction of these fits) and extracted the coefficient for each term to index the shape of each infant's learning trajectory in each ROI. These terms were then entered into a multiple regression for each ROI to determine whether these two coefficients could predict the visual omission response. This model, in which the response to the visual omission trial was the dependent variable and the coefficients for block and block-squared were the two independent variables, was significant in the temporal and occipital ROIs (*Temporal*:  $R^2 = 0.58$ ,  $F(2, 14) = 9.64$ ,  $p < .01$ , block coefficient = 1.19,  $t = 4.21$ ,  $p < .01$ , block-squared coefficient = 8.06,  $t = 3.66$ ,  $p < .01$ ; *Occipital*:  $R^2 = 0.42$ ,  $F(2, 14) = 5.05$ ,  $p = .02$ , block coefficient = 0.94,  $t = 3.14$ ,  $p < .01$ , block-squared coefficient = 7.77,  $t = 3.07$ ,  $p < .01$ ; *Frontal*:  $R^2 = 0.19$ ,  $F(2, 14) = 1.60$ ,  $p > 0.1$ ).<sup>5</sup> This indicates that the shape of the learning trajectory predicts the visual omission responses in both the occipital and temporal ROIs. It is important to note that these models are all predicting responses within ROIs. Between ROI analyses where learning trajectories in the frontal or temporal ROIs were used to predict the occipital lobe response to visual omission trials were not significant.

## Discussion

The current study investigates changes in infant neural activity during audio-visual (AV) associative learning. To that end, we estimated the amount of neural activity during each learning block (exposure to six AV

trials, Figure 1) and employed regression-based statistical methods to determine the shape and direction of neural changes for each infant's learning time-course. We considered three possible relations between learning and neural changes. Across repeated learning experiences, we predicted that infants would exhibit (1) a decline of neural activity similar to repetition suppression; (2) an increase in neural activity similar to repetition enhancement; or (3) a non-linear change to neural activity where infants will first exhibit repetition enhancement followed by repetition suppression.

We do *not* find evidence that infants simply exhibit repetition suppression over the course of learning, and we find strong converging evidence that there are increases in neural activity across learning blocks. Specifically, regression-based statistics and targeted *t*-tests both reveal *positive* linear changes in neural activity over learning blocks regardless of whether we consider absolute block number or proportion of learning blocks for each infant (in order to accommodate the varying number of blocks viewed by the infants, 5–8 blocks). This extends the finding of increases in neural activity over learning from Gervain *et al.* (2008) to another learning task and numerous regions of the infant brain, suggesting that increases in neural activity with learning are part of a domain-general mechanism available early in life and not an example of domain-specific learning capacities.

While the evidence for increases in neural activity compared to block 1 are robust, there is also evidence that neural activity starts to decline at later stages of learning. In particular, we found that the fit of our statistical models significantly increased when we included a non-linear predictor indicating the presence of a non-linear relationship between learning and neural activity. In addition, targeted *t*-tests found that following a significant increase between blocks 1 and 4 in all ROIs, there is a significant decrease from block 4 to an infant's final block (which varied from 5 to 8) in the frontal ROI and a marginally significant decrease for the temporal ROI. Visual inspection of the data in these two ROIs also strongly suggests that neural activity starts to decline about two-thirds of the way through the learning task (regardless of whether we considered absolute block number or proportion of learning blocks to consider the learning trajectory for each infant). However, the evidence for a decline in occipital lobe activity is much weaker and both targeted *t*-tests and visual inspection suggest that the non-linear predictor is modeling a plateau or lack of increase in neural activity in later blocks and not repetition suppression.

<sup>5</sup> It is important to note that the coefficients for the block and block-squared predictors were highly correlated ( $R_s > 0.9$ ) which presents the problem of colinearity in the model. To confirm that the current results are robust and to control for colinearity, we ran a complementary set of models, which employed residualization to control for the shared variance in the block and block-squared term. First, we constructed a model where one of the predictors is used to predict the other predictor (e.g. Temporal Block Coefficients ~ Temporal Block Squared Coefficients). Then the residuals from this model are used to predict the Visual Omission response. These models, which help control for the colinearity of the two predictors, indicate that effects above continue to be robust for the Temporal and Occipital ROIs and continue to be absent for the Frontal ROI.

Lending additional support to a non-linear relationship between learning and neural activity, we conducted an exploratory analysis intended to determine whether a non-linear learning trajectory of individual infants would predict the robustness of that individual's novelty detection. The linking hypothesis here is that if a non-linear learning trajectory is crucial to the acquisition of audio-visual representations through learning, the degree of non-linearity of an individual infant's learning trajectory will predict how robustly an infant responds to a violation of the audio-visual pairing. The model of each infant's learning trajectory was used to predict their responses to a novel trial and find a significant relationship, but interestingly only in the temporal and occipital ROIs.

It is important to note that, even though we find robust evidence for non-linearity in the neural changes with learning in this task, we consistently find a significant linear predictor as well in these models. There are two major possibilities for the presence of both linear and a non-linear effects in these data. First, there could be a combination of processes underlying the relation between changes of neural activity and learning, with one being linear and positive and other being non-linear (starting with an increase and followed by a decrease or plateau in the case of the occipital lobe). A second and more parsimonious possibility is that, because the decline of neural activity later in learning is not as dramatic as the increase in neural activity early in learning, the linear term models this asymmetry. Moreover, it is possible that we simply do not have the methodological ability to capture the full decline of the neural activity, which depends on infant compliance after an infant has successfully learned the AV pairing and is no longer interested in the stimuli, since infants in this task effectively choose when they finish the task largely based on their interest and at the point of boredom. Indeed, it appears as if after a few blocks characterized by a decline or plateau in neural activity, they stop attending to the screen and the experiment is ended. Thus, if a decline or plateau in neural activity later in learning reflects the successful creation of a representation of the AV pairing and a decline in an infant's interest in the AV stimuli, then the nature of our experimental method would predict that we would not fully capture decreases in neural activity in later stages of learning.

Relatedly, while we do not find any evidence for a repetition suppression type effect underlying infant learning in the current task, a non-linear trajectory suggests that repetition suppression may be present during learning but may simply characterize one phase or stage of learning and specifically the later stages of

learning when representations are presumably largely formed. Thus, while on the surface our result runs counter to proposals by Mather (2013) and Turk-Browne *et al.* (2008), the current work simply places the link between repetition suppression and learning in infancy in a larger context by considering the entire learning time-course. It is important to note that we find no evidence of a significant decline in neural activity in the occipital ROI only in the frontal and temporal ROIs.

How does the current work relate to previous demonstrations of repetition suppression in infancy? Nakano *et al.* (2009) presented robust evidence for repetition suppression in the infant brain. In this study, infants heard repeated blocks of a single stimulus ('ba') and many regions including the temporal and frontal cortices exhibited decreasing responses over these repeated blocks. The repetition of a single stimulus is more akin to classic repetition suppression studies in adults where simple stimuli are repeated with no embedded patterns to learn (e.g. Grill-Spector *et al.*, 1999). As discussed in the Introduction, it has been demonstrated in adults that greater complexity can determine whether the experiment produces repetition enhancement vs. suppression (Muller *et al.*, 2013; Turk-Browne *et al.*, 2007). Given the differences in complexity between Nakano *et al.* (2009; repetition of a single syllable) and the current experiment (the association of two related and complex audio-visual stimuli), uncovering suppression vs. enhancement across the two studies, respectively, follows directly from the adult literature.

However, it is important to note that while Nakano *et al.* (2009) present a robust finding of repetition suppression, recent studies in developmental populations reveal mixed results. Although some studies have found evidence of repetition suppression to numerical stimuli (e.g. dot arrays or Arabic digits) in children (Cantlon, Brannon, Carter & Pelphrey, 2006; Vogel, Goffin & Ansari, 2015), other results have demonstrated the difficulty in exhibiting repetition suppression to more complex stimuli (particularly visual repetition suppression) in developmental populations. In a particularly compelling example, Scherf, Luna, Avidan and Behrmann (2011) examine repetition suppression of two types of visual stimuli in children, adolescents and adults. While adults exhibit robust repetition suppression for both stimuli, there is no evidence for repetition suppression to either stimulus type in children and an intermediate amount in adolescents. Given that research with children facilitates neural recordings even after children have reached later stages of learning (compared to infants), the lack of evidence for repetition suppression for complex stimuli in children suggests that there may be broader developmental changes in neural suppression

beyond differences in methodology. The developmental origins of such a difference in the neural response to repetition are currently unknown, but disentangling developmental and methodological differences across populations to address these questions directly will be an important direction for future research.

Concerning the relation between an infant's visual preference and increases in neural activity, Watanabe *et al.* (2008) found increased activity in lateral portions of occipital lobe and the prefrontal cortex when infants viewed a mobile versus a checkerboard (preferred and non-preferred visual stimuli, respectively). The authors propose that this increase may reflect an infant's visual preference or attention. Our current findings also suggest that visual preferences may correspond to increases in neural activity and, moreover, that neural changes during learning may mirror the canonical Hunter and Ames model of infant visual preference.

Finally, Watanabe, Homae, Nakano, Tsuzuki, Enkhtur *et al.* (2013) examined infant neural activity during either the visual only presentation of a novel mobile or the audio-visual presentation of this mobile. Broadly, the authors report greater responses to the audio-visual stimuli in many regions of the infant brain including some occipital regions and the frontal cortex.<sup>6</sup> While the authors attribute these increases to multisensory perception, it is also likely that these effects reflect audio-visual learning. This is particularly likely given that the infants were presented with a novel mobile that was quite complex. Certainly, current results suggest that these findings may reflect a learning process and not necessarily that the frontal cortex is supporting multisensory perception of known objects early in life.

The current results provide a foundation for building a more nuanced understanding of the neural processes that underlie learning in the infant brain. One potential direction of future work should evaluate the effects of task complexity on the shape of the learning trajectory. This work would further test the predictions of the Hunter and Ames (1988) model as, for example, extremely easy tasks should reduce the amount of time that infants exhibit repetition enhancement before exhibiting repetition suppression. Future work should also consider measuring the outcome of learning in other ways. For instance, here we used response to visual omission trials as an index of learning, but the neural response to other types of learning violations (e.g. mismatches of audio and visual pairings) may provide

<sup>6</sup> There were differences in activity in the temporal cortex as well, but this effect is hard to interpret because these two conditions differed on the basis of whether sound was presented.

additional insight. Finally, one important challenge is integrating more precise measures of infant attention or habituation and fNIRS recordings. Indeed, Aslin *et al.* (2015) highlight this as an important future advance for fNIRS. Integrating looking time measures of behavior and fNIRS is not trivial as the latter method considers neural activation for precisely timed and uniformly watched blocks of stimuli while the former relies on changes in looking time over the duration of the experiment. Thus, fNIRS studies end when looking time studies begin. Indeed, this difference in methodology could suggest that greater amounts of repetition suppression could have been seen if more controlled habituation techniques were employed in combination with fNIRS recordings. Indeed, the future integration of these methods will provide a powerful tool for considering the relationship between neural and behavioral changes involved in habituation and learning in infancy.

In sum, we find that during an audio-visual learning task, infants first show an increase in activation followed by a decrease in temporal, occipital, and frontal regions of the brain. This inverted U-shaped response during learning demonstrates that the infant brain shows evidence of both repetition enhancement and repetition suppression while building associations between audio and visual stimuli. Furthermore, we find that the shape of individual trajectories can be used to predict the response to novel stimuli in occipital and temporal ROIs, providing evidence that the shape of the learning trajectory is related to learning outcomes. These results are consistent with the canonical Hunter and Ames model of infant visual preference and are suggestive of a neural signature that may underlie looking preferences in infants.

## References

- Aslin, R.N., Shukla, M., & Emberson, L.L. (2015). Hemodynamic correlates of cognition in human infants. *Annual Review of Psychology*, **66**, 349–379.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67** (1), 1–48.
- Cantlon, J.F., Brannon, E.M., Carter, E.J., & Pelphrey, K.A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, **4**, 844–854.
- 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, USA*, **112**, 9585–9590.
- Gerken, L., Balcomb, F.K., & Minton, J.L. (2011). Infants avoid 'labouring in vain' by attending more to learnable than

- unlearnable linguistic patterns. *Developmental Science*, **14**, 972–979.
- Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences, USA*, **105**, 14222–14227.
- Gogate, L.J., & Bahrick, L.E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, **69**, 133–149.
- Gomez, R.L., & Gerken, L.A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, **70**, 109–135.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, **10**, 17–19.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., *et al.* (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, **24**, 187–203.
- 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.
- Huppert, T.J., Diamond, S.G., Franceschini, M.A., & Boas, D.A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, **48**, C280–C298.
- Karuz, E.A., Emberson, L.L., & Aslin, R.N. (2014). Combining fMRI and behavioral measures to examine the process of human learning. *Neurobiology of Learning and Memory*, **109**, 193–206.
- 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**, e36399.
- Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2013). lmerTest: tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version*, **2**.
- Lloyd-Fox, S., Blasi, A., Volein, A., Everdell, N., Elwell, C.E., *et al.* (2009). Social perception in infancy: a near infrared spectroscopy study. *Child Development*, **80**, 986–999.
- Marcus, G.F., Vijayan, S., Rao, S.B., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science*, **283**, 77–80.
- Mather, E. (2013). Novelty, attention, and challenges for developmental psychology. *Frontiers in Psychology*, **4**, 491.
- Muller, N.G., Strumpf, H., Scholz, M., Baier, B., & Melloni, L. (2013). Repetition suppression versus enhancement: it's quantity that matters. *Cerebral Cortex*, **23**, 315–322.
- Nakano, T., Watanabe, H., Homae, F., & Taga, G. (2009). Prefrontal cortical involvement in young infants' analysis of novelty. *Cerebral Cortex*, **19**, 455–63.
- Plichta, M.M., Heinzl, S., Ehli, A.-C., Pauli, P., & Fallgatter, A.J. (2007). Model-based analysis of rapid event-related functional near-infrared spectroscopy (NIRS) data: a parametric validation study. *NeuroImage*, **35**, 625–634.
- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926.
- Scherf, K.S., Luna, B., Avidan, G., & Behrmann, M. (2011). 'What' precedes 'which': developmental neural tuning in face- and place-related cortex. *Cerebral Cortex*, **21**, 1963–1980.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, **106**, 1558–1568.
- Stager, C.L., & Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, **388**, 381–382.
- Turk-Browne, N.B., Scholl, B.J., & Chun, M.M. (2008). Babies and brains: habituation in infant cognition and functional neuroimaging. *Frontiers in Human Neuroscience*, **2**, 16.
- Turk-Browne, N.B., Yi, D.-J., Leber, A.B., & Chun, M.M. (2007). Visual quality determines the direction of neural repetition effects. *Cerebral Cortex*, **17**, 425–433.
- Vogel, S.E., Goffin, C., & Ansari, D. (2015). Developmental specialization of the left parietal cortex for the semantic representation of Arabic numerals: an fMR-adaptation study. *Developmental Cognitive Neuroscience*, **12**, 61–73.
- Wagner, J.B., Fox, S.E., Tager-Flusberg, H., & Nelson, C.A. (2011). Neural processing of repetition and non-repetition grammars in 7- and 9-month-old infants. *Frontiers in Psychology*, **2**, 168.
- Watanabe, H., Homae, F., Nakano, T., & Taga, G. (2008). Functional activation in diverse regions of the developing brain of human infants. *NeuroImage*, **43**, 346–357.
- Watanabe, H., Homae, F., Nakano, T., Tsuzuki, D., Enkhtur, L., *et al.* (2013). Effect of auditory input on activations in infant diverse cortical regions during audiovisual processing. *Human Brain Mapping*, **34**, 543–565.

Received: 30 November 2015

Accepted: 11 July 2016

## Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

**Figure S1.** Depiction of part of the experimental timecourse and the Audio-Visual (AV) and Audio-No Visual (AV-) events that make up the learning blocks and single trials. Learning blocks and single trials were separated by inter-block intervals (IBI).

**Figure S2.** Fits of data from individual infants to a model with a linear term and a non-linear term for block (2-order polynomial).

**Data S1.** Relation between learning trajectory and learning outcomes.