Cognition 193 (2019) 104019

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Infants use knowledge of emotions to augment face perception: Evidence of top-down modulation of perception early in life

Naiqi G. Xiao*, Lauren L. Emberson

Department of Psychology, Princeton University, Princeton, NJ, USA

ARTICLE INFO

Keywords:
Top-down
Perception
Infants
Emotion
Face perception
Cross-modal perception
Perceptual development

ABSTRACT

While top-down modulation is believed to be central to adult perception, the developmental origins of this ability are unclear. Here, we present a direct, behavioral investigation of top-down modulation of perception in infancy using emotional face perception as a test case. We investigated whether 9-month-olds can modulate their face perception based on predictive, auditory emotional cues without any training or familiarization procedure. Infants first heard a 3-second emotional vocal sound (happy/angry) while their gaze was held in the center of the screen. Then, they were presented with a pair of emotional and neutral faces images without any audio sound. The faces were small (4.70° × 5.80°) and presented in randomized locations outside their focus of attention. We measured the initial latency to shift gaze to look at a congruent emotional face as an index of infants' pre-attentive perception of these faces. We found that infants' face perception was augmented by preceding emotional cues: They were faster to look at the emotional face after hearing an emotionally congruent sound than an incongruent one. Moreover, the emotional sounds boosted perception of congruent faces 200 ms after the onset of the faces. These top-down effects were robust for both happy and angry emotions, indicating a flexible and active control of perception based on different top-down cues. A control study further supported the view that the Congruency effect is due to a top-down influence on face perception rather than a rapid matching of cross-modal emotional signals. Together, these findings demonstrate that top-down modulation of perception is already quite sophisticated early in development. Raw data is available on Github (https://github.com/naiqixiao/CuedEmotion.git).

1. Introduction

Given the central role of top-down modulation in adult perception (Bar, 2004; Gilbert & Li, 2013), the developmental emergence of this phenomenon is essential to understand, yet top-down influences on perception in infancy has received very little empirical focus to date (Emberson, 2017; Hadley, Rost, Fava, & Scott, 2014; Markant & Scott, 2018). Here, we present a direct behavioral test of top-down modulation of perception in infancy. We determined that infants' prior knowledge of emotions gained outside the laboratory could be used in a top-down fashion to quickly and flexibly modulate face perception at 9 months of age.

In the 20th century, the adult brain was largely considered a feed-forward system in which information is passed successively from one specialized neural region to another (e.g., Norris, McQueen, & Cutler, 2000; Tanaka, 1993, 1996). Recent decades of work have uncovered the importance of dynamic neural networks in human cognition. These networks include feedback and long-range neural connections that are able to engender top-down modulation of perceptual systems (Gilbert & Li, 2013), among other functions. Top-down neural modulation increases neuron response efficiency (Kok, Jehee, & de Lange, 2012; Manita et al., 2015) and plays a crucial role in supporting perception especially when sensory input is weak or ambiguous (e.g., Bullier, 2001; Chen et al., 2014; Grassi, Zaretskaya, & Bartels, 2017; Hupé et al., 1998; Schwiedrzik & Freiwald, 2017), with even very long-range connections can modulate visual perception (e.g., from the frontal lobe, Summerfield et al., 2006).

We define **top-down** as neural signals that feedback along established cortical hierarchies (cf. Emberson, 2017; Gilbert & Li, 2013). Consider the domain under investigation: emotion and face perception. A likely neural pathway by which emotional information can influence face perception is via the amygdala. The human amygdala encodes social and emotional information across sensory modalities (Schirmer & Adolphs, 2017), which receives feedforward inputs from high-level visual cortices (e.g., Pessoa & Adolphs, 2010). Thus, we consider 1) the amygdala to be higher in the visual hierarchy than high-level visual

* Corresponding author.
E-mail addresses: nx@princeton.edu, nx@kangleelab.com (N.G. Xiao).

https://doi.org/10.1016/j.cognition.2019.104019
Received 20 September 2018; Received in revised form 21 June 2019; Accepted 25 June 2019
Available online 08 July 2019

cortices (e.g., inferior temporal cortex, IT), and 2) neural signals transferred from amygdala to visual cortices to be feedback or top-down (e.g., Anderson & Phelps, 2001). While we consider the amygdala as the source of top-down information to the visual system in the context of emotional information, the originating neural region (i.e., the contributor of top-down information) will vary relative to task (e.g., for non-emotional stimuli), developmental availability, and other factors.

Top-down modulation could be crucial to the development of perceptual systems. Perceptual systems undergo protracted development, which greatly limits the fidelity of perceptual signals early in life. Given that adults rely on top-down signals mostly when sensory input is weak or ambiguous (Bar, 2004), it is possible that infant perception could greatly benefit from top-down signals if they were available early in life. Moreover, top-down signals would provide an effective way for infants to leverage their expanding knowledge about the world into adaptive changes in their perceptual abilities. However, to date, very little is known about top-down influences on developing perceptual systems.

Despite some neural connectivity studies suggesting that top-down modulation might not be available, or might be extremely limited, early in life (Batardière et al., 2002; Gao et al., 2015; Gilmore, Knickmeyer, & Gao, 2018; Menon, 2013; Price et al., 2006), recent functional neuroimaging studies suggest that these weak and developing neural networks might not limit an infant's capability for engaging in top-down of visual perception. Converging findings have established that infants can use their recent experience to generate top-down signals that modulate neural activity in their visual systems (fNIRS: Emberson, Richards, & Aslin, 2015; EEG/ERP: Gliga, Volein, & Csibra, 2010; Kouider et al., 2015; Vogel, Monesson, & Scott, 2012). Particularly relevant for the current study, Vogel et al. (2012) found that face perception related ERP components (N290 & P400) were modulated by preceding emotional vocal sounds in 9-month-olds. Thus, despite the weak long-range neural connections available in infancy, neuroimaging studies suggest there may already be top-down modulation of perceptual systems early in life.

Building on this neuroimaging evidence, it is essential to determine whether infant's perception, as measured behaviorally, is affected by top-down signals. It is possible that the top-down neural modulation found in developmental neuroimaging studies (e.g., Emberson et al., 2015, Vogel et al., 2012) is too weak or delayed to modify perception. Moreover, in the case of work by Emberson and that by Kouider and colleagues, these neural changes could represent prediction errors rather than alterations in perception based on top-down signals, given that these signals were generated when novel combinations of stimuli were presented. Ultimately, to determine whether these neural signals shape the development of perception, a behavioral demonstration of top-down modulation of perception is necessary in young infants.

By using behavioral measures closely indexing perception, previous studies have demonstrated that experience can alter infants' perception but cannot determine whether these perceptual changes arise from top-down influences. Not all changes in perception based on experience or familiarity are top-down, as strengthening of representations can occur through purely feedforward mechanisms and do not rely on feedback connections (e.g., Scott, Pascalis, & Nelson, 2007). A series of studies have found that 3-month-olds' scene segmentation is shaped by prior visual experience (e.g., Needham, Cantlon, & Ormsbee Holley, 2006; Quinn, Schyns, & Goldstone, 2006). Learning an audio-visual association enabled 3-month-olds to imagine a visual event based on the associated auditory cue (Spelke, 1981). However, in these cases, it is not clear whether these experience-based changes in perception involve top-down modulation (i.e., require feedback neural signals) or whether they occur through bottom-up mechanisms (e.g., passive strengthening representations based on sensory input).

The current study presents the most direct behavioral investigation of top-down modulation of perception in infancy to date. We capitalized on infants' prior knowledge of the relation between emotions in facial expression and vocalizations. This knowledge is available to infants starting in the first few months of life, as revealed in the cross-modal matching of emotional signals (e.g., Kahana-Kalman & Walker-Andrews, 2001). Exploiting this early available knowledge, we examined whether infants can use emotional vocal sounds to generate top-down signals that would augment infants' face perception.

We developed a novel paradigm that allows for a more direct behavioral measurement of perception relying on attentional orientation. The measurement of perception on each trial also allows for within-subjects comparisons of how different emotional cues affect infant face perception. Infants first hear an emotional vocal sound while their gaze is drawn to the center of the screen with a neutral attention-getter. Then, two face images are presented outside their foveal visual field: One is an emotional face; the other is a neutral face. The peripherally-presented faces would provide relatively weak bottom-up sensory input (Fig. 1). We hypothesized that top-down signals provided by the emotional cues would augment the perception of the congruent emotional face despite weak sensory input, causing the infant to look more quickly

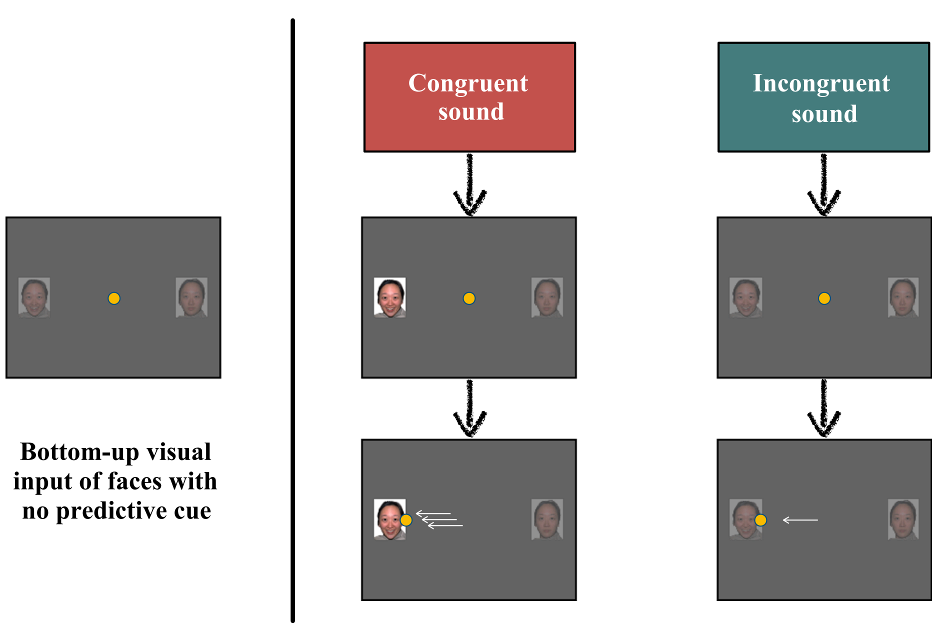

Fig. 1. Schematic of infant perception and eye gaze shifting in this task. Perception is presented as the relative salience of the two face images (i.e., image transparency); eye gaze is the yellow dot. Left: Bottom-up input is weak when infants' attention is held at screen center with no differences in perception between the two faces. Center: If infants are capable of top-down modulation of face perception, preceding congruent emotional vocalizations will boost perception of a congruent emotional face. This better perception of the congruent faces results in attentional orienting (initial looking) to that face. Right: Incongruent emotional sounds do not augment perception of the emotional face and will not guide infants' initial looking. Top-down modulation of perception will be manifested by faster initial looking latency in the congruent sound condition (center panel) than that in the incongruent condition (right panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

at the congruent emotional face (as it is perceived more clearly than the other face). Thus, we used the time that infants took to shift attention to the emotional face (initial looking latency) from screen central to index relative differences in face perception between the two faces (Fig. 1). A faster shift of attention to the emotional face in the congruent condition compared to the incongruent condition would be evidence of top-down modulation on face perception.

2. Experiment 1

2.1. Method

2.1.1. Participants

Eighteen infants (8 females, 10 males) participated in the current experiment. These infants were 9 months of age (M = 272.89 days, SD = 18.44 days). Sample size was determined by referring to recent infant studies on top-down effect on attention (e.g., Richardson & Kirkham, 2004; Tummeltshammer & Amso, 2018). All of the participants were Asians who had not seen the face images used in the current experiment before their participation. Three additional infants participated in the experiment but were excluded from the final sample because insufficient data was collected (i.e., less than 1 trial per condition) due to crying (n = 1) or not looking at the screen (n = 2). This study was conducted in Asia and was approved by IRB of Princeton University and that of Zhejiang Sci-Tech University.

2.1.2. Materials and procedure

At the beginning of each trial, participants heard a 1-second vocal sound (e.g., laughing and grumbling, Sauter, Eisner, Ekman, & Scott, 2010) repeated 3 times for a total of 3 s. During this time, participants would see a light grey ball at the center of a blank screen. The ball changed its size slowly so as to attract infants' attention. At 500 ms before the vocal sounds ended, this ball was replaced by colorful rings twice the size of the ball. This change in visual display attracted infants' attention to screen center at the point when they were presented with the face images. Eye tracking data was analyzed starting at the onset of the face images.

As shown in Fig. 2, two face images were presented on each side of the screen as the vocal sound terminated. These face images were 150 pixels wide and 180 pixels high, occupying 3.43% of the screen area. They were presented at 358 pixels left or right to the center of a 17-inch screen (resolution: 1024 × 768 pixels). Based on the distance

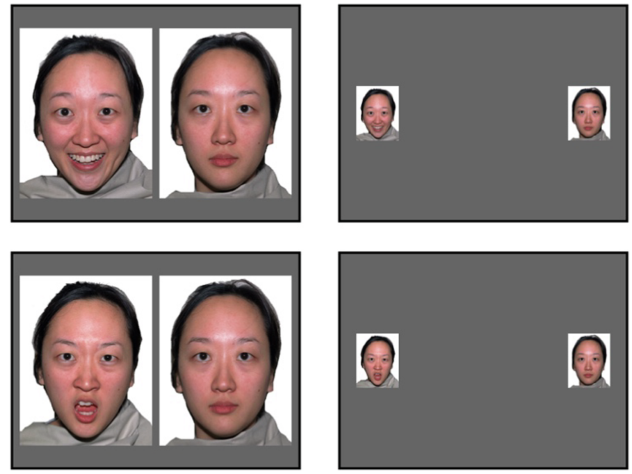

Fig. 3. Demonstration of the size and layout of face images (Upper: happy face, Lower: angry face) in previous cross-modal matching studies as summarized in Table 1 (Left, image width: 15.02° visual angle) and those in the current study (Right, image width: 4.70° visual angle).

estimation from the eye tracker (approximately 60 cm from the screen for all infants), each image subtended 4.70° (width) × 5.80° (height) of visual angle and was located towards the periphery at 11.50° of visual angle away from the center of the screen. All face images were presented in color against a grey background. The face images showed up for 3 s without any auditory stimulation accompanied. Face stimuli were 4 Asian female actresses from the NimStim Face Set (Face# 15, 16, 17, & 19, Tottenham et al., 2009).

Trials followed a 2 × 2 design (vocal sound × emotional facial expression), but we used a variety of faces across trial types. Four different Asian actresses (NimStim Face Set #15, 16, 17, & 19, Tottenham et al., 2009) were used, each with their own neutral, angry, and happy expression. Infants were given no training on these stimuli prior to the experiment. On a given trial, the faces depicted the same person but with two facial emotions: One with an emotional expression (happy or angry) and one with a neutral expression. Given the two possible emotional sounds and two possible emotional faces, there were four conditions: Congruent happy face (happy vocal sound-happy face), Congruent angry face (angry vocal sound-angry face), Incongruent happy face (angry vocal sound-happy face), and Incongruent angry face

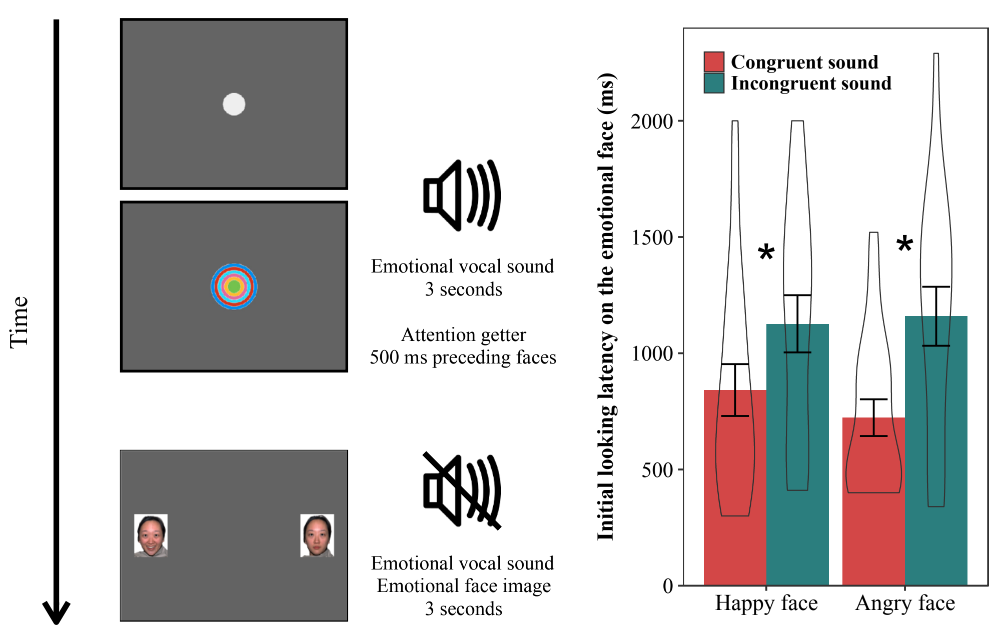

Fig. 2. A representative trial for Expt. 1 (Left) and infants' initial looking latency to the emotional face (Right). Error bars represent unit standard errors. The overlaid shapes represent the distribution of individual response for each condition. The asterisks indicate a significant difference between the Congruent and Incongruent conditions (Tukey's HSD, FDR corrected p < .050).

Table 1
Experimental parameters used by previous studies on infants' cross-modal matching emotions vs. the current study (the bottom row). Four studies with identical experimental settings to the listed studies were not included due to space constraints.

Study	Presentation order	Orientating attention to the center	Presentation duration	Stimuli width	Stimuli location	Measurement
Walker and Walker (1982)	Simultaneous	No	120 s	~30.00°	Foveal (3.81°)	Proportional looking
Soken and Pick (1999)	Simultaneous	Yes	60 s	32.61°	Near peripheral (11.78°)	Proportional looking
Kahana-Kalman and Walker-Andrews (2001)	Simultaneous	Yes	25 s	12.18°	Foveal (0°)	Proportional looking
Montague and Walker-Andrews (2002)	Simultaneous	No	25 s	25.75°	Foveal (0°)	Proportional looking
Zieber et al. (2013)	Simultaneous	No	15 s	7.63°	Near peripheral (12.90°)	Proportional looking
The current study	**Sequential: Auditory preceding**	**Yes**	**3 s**	**4.70°**	**Near Peripheral (11.5°)**	**Initial looking latency**

(happy vocal sound-angry face) conditions. Each of these four conditions included four trials. The order in which the images were presented was randomized. Thus, there were $4 \times 4 = 16$ trials in total, with a fully random presentation order.

The use of audio and visual emotional stimuli in the current experiment is methodologically distinct from previous studies on infants' cross-modal matching of emotional information (Kahana-Kalman & Walker-Andrews, 2001; Montague & Walker-Andrews, 2002; Soken & Pick, 1999; Walker, 1982; Zieber, Kangas, Hock, & Bhatt, 2013). **These crucial differences allow us to quantify infant face perception** as opposed to infant attention to faces (Table 1). First, the emotional vocal sounds were presented before face images in the current study with no presentation of vocal sounds during face presentations (sequential presentation); studies of cross-modal matching typically present face and vocal sounds simultaneously. Second, we used an attention getter to ensure that infant attention was oriented to the center of the screen during the onset of the face stimuli. As a result, participants had equal bottom-up input for both of these faces to ensure there was no systematic bias in infant attentional orienting to one face over another in the absence of top-down modulation based on emotional cues. The emotional face location was randomized across trials so infants could not predict where an emotional face will be presented. Moreover, to ensure that each face was outside the locus of infant attention, the face stimuli were substantially smaller (4.70° visual angle) than those in the previous ones (7.63–32.61° visual angle). The face images were presented at near-peripheral visual field (11.50° from the attended location). These two designs were crucial in order to attribute any changes in initial looking behavior to biases in infant perception of these faces based on the vocal sounds. Furthermore, the small size of the faces reduced the bottom-up perceptual signal from each face, which encouraged infants to rely on any top-down signal they were able to generate to modulate their perception (Hupé et al., 1998). Finally, our trials were very short (3 s face presentation compared to 15–120 s in previous studies). This rapid presentation could increase infants' reliance on top-down signals because it reduced the availability of bottom-up perceptual information and included time pressure. These details in the experimental design enabled the experiment to attribute differences in shifting eye gaze to differences in face *perception* as opposed to other processes (e.g., cross-modal matching, memory recall, and visual attention).

2.2. Results and discussion

2.2.1. Initial looking latency is biased by preceding vocal cues

Initial looking latency was defined as the time-point of the first gaze sample that landed on the emotional face AOI. Since we are measuring the initial looking behavior to small stimuli presented peripherally, this measures an infant's ability to perceive the two faces in their near-peripheral vision rather than other abilities, such as spatial attention or a detection of the correspondence between the audio and visual stimuli, which requires that the two be presented simultaneously (i.e., cross-modal matching, see Expt. 2 for a control addressing this point). By comparing the latency between the Congruent and Incongruent trials, we determined whether preceding emotional vocal sounds (a potential top-down cue) influence face perception. We also analyzed the initial fixation location before the face images appeared, which confirmed that infants started the trial with a central fixation (for details see Supplemental Materials).

A repeated-measures ANOVA was used to examine the influence of Congruency between the vocal sounds (Congruent vs. Incongruent) and face expressions (Happy vs. Angry) on initial looking latency. Both independent variables were within-subject. As shown in Fig. 2, the initial looking latency in the Congruent trials ($M = 782$ ms, $SD = 409$ ms) was faster than that in the Incongruent trials ($M = 1143$ ms, $SD = 523$ ms), which was supported by a significant main effect of Congruency ($F(1, 17) = 13.18$, $p = .002$, $\eta_p^2 = .44$). Moreover, we did not find that this auditory influence on visual perception differed by the two emotions ($F(1, 17) = 0.11$, $p = .747$, $\eta^2_P = .01$) or interacted ($F(1, 17) = 0.44$, $p = .514$, $\eta^2_P = .03$).

We further examined the Congruency effect within happy and angry face trials with Tukey's HSD post-hoc test. The results showed that the Congruency effect existed in both the happy ($z = 2.01$, FDR corrected $p = .045$) and angry face trials ($z = 2.56$, FDR corrected $p = .021$). This confirms the generalization of this effect across two emotion contexts and that the effect is not driven by an infant preference for a given emotional face type.

Infants' initial looking latency indicated that they looked towards emotional faces faster when they were preceded by Congruent emotional vocal sounds, as compared to when they were cued by Incongruent emotional vocal sounds. These findings demonstrate that hearing an emotional vocal sound boosts infants' ability to perceive the corresponding emotional face within the visual display. It is possible this top-down effect is derived from an increase in perceptual sensitivity of facial expression by boosting the neural response representativeness and accuracy of corresponding facial expressions (e.g., Kok et al., 2012; Manita et al., 2015). It is also possible that the top-down influence is underlain by a modification on the perceptual decision process (Summerfield & de Lange, 2014) in which infants are more likely to respond to the corresponding facial expression. Although it is difficult to identify the specific perceptual components that are augmented by top-down signals, the current finding suggests a top-down influence on perceptual systems in infants.

2.2.2. Looking preference is biased by preceding vocal cues

Looking preference (as measured by proportion of looking time) is widely used in prior cross-modal emotion perception studies (e.g., Montague & Walker-Andrews, 2002; Zieber et al., 2013) and has been established previously for the current stimuli (Zieber et al., 2013). This measure allows us to use infants' visual attention to infer whether infants can perceptually differentiate emotional contents in visual stimuli (e.g., face images, facial motion, and body gestures). It is typically considered a measure of outcomes of perceptual processes rather than a direct measure of face perception *per se*. As summarized in Table 1, the current experimental design differs from cross-modal matching in several ways, including the key experimental parameters of stimuli presentation duration, stimuli size, and stimuli presentation order. Despite these differences, we determined whether infants still exhibit a looking preference for the emotional face indicated by the emotional sound beyond their initial looking latency.

A visual preference score for the emotional face was calculated on each trial based on the number of eye-tracking samples (N) that fell within the AOIs: $N_{\text{emotional face}}/(N_{\text{emotional face}} + N_{\text{neutral face}})$. A score of 50% means infants looked at the emotional and neutral faces for equal amounts of time; a score above 50% means infants looked mostly at the emotional face; vice versa for a score below 50%.

We first examined whether the emotional vocal sounds influenced the visual preference with a 2 (Congruency of the emotional vocal sounds: Congruent vs. Incongruent) × 2 (Face expression: Happy vs. Angry) repeated-measures ANOVA on the visual preference for emotional faces. The preference score was significantly higher in the Congruent condition ($M = 64.96\%$, $SD = 15.32\%$) as compared to that in the Incongruent condition ($M = 49.97\%$, $SD = 15.00\%$, $F(1, 17) = 18.97$, $p < .001$, $\eta^2_P = .53$, Fig. 4). Infants tended to look proportionally longer at the emotional face when preceded by Congruent vocal sounds versus Incongruent vocal sounds. We further examined the Congruency effect within happy and angry face trials with Tukey's HSD post-hoc test. The results showed that this Congruency effect (i.e., difference in looking to the emotional face in the Congruent vs. Incongruent conditions) was significant in both the happy face trials ($z = 2.93$, FDR corrected $p = .003$) and angry face trials ($z = 2.94$, FDR corrected $p = .003$). We found no significant effect of Face expression ($M_{\text{Happy face}} = 57.45\%$, $SD_{\text{Happy face}} = 17.46\%$, $M_{\text{Angry face}} = 57.49\%$, $SD_{\text{Angry face}} = 16.45\%$, $F(1, 17) < 0.01$, $p = .992$, $\eta^2_P < .01$) and no interaction of Face expression and Congruency ($F(1, 17) = 0.45$, $p = .511$, $\eta^2_P = .03$).

One-sample *t*-tests indicated that visual preference for the happy face was significantly above chance (50%) after they heard a happy vocal sound ($t(17) = 3.92$, $p = .001$). By contrast, no preference was found after they heard an Incongruent angry vocal sound ($t(17) = -0.46$, $p = .653$). Similarly, a significant preference for the angry face was found after infants heard an angry vocal sound ($t(17) = 4.27$, $p < .001$). No such preference for angry faces was found in the Incongruent condition ($t(17) = 0.28$, $p = .782$). These results indicated that a preceding emotional vocal sound could guide infants to primarily look at face images with congruent facial expression. This was not driven by a preference to look at emotional faces in general because there was no looking preference in the Incongruent conditions in which no face matched the vocal sound.

2.2.3. Looking preference is driven by initial looking behavior

Here, we integrate these two measures and examine the proportion of looking time over the time-course of the trial. Given the initial boost of face perception based on the vocal cue (revealed through initial looking latency), we hypothesize that the difference in visual preference will emerge early in the trial. The presence of proportion looking preferences early in the face presentation suggest that the overall looking preference in this task arises from these initial perceptual changes rather than from a comparison process between the two signals requiring multiple saccades.

To this end, we performed a time-course analysis of infants' visual preference by segmenting the 3-second face presentation phase into fifteen 200 ms time windows. The 200 ms duration for each time window was chosen because infants were capable of initiating a saccade around 200 ms (Hood & Atkinson, 1993). Given that previous analyses consistently reveal no differences between face expressions, we collapsed the preference scores of the happy and angry expression conditions by the Congruency of the emotional vocal sounds. This was particularly important as this fine-grained temporal analysis is more sensitive to the inherent noisiness of infant data. We then performed a series of paired-sample *t*-tests between the Congruent and Incongruent conditions within each time window to determine if infants were exhibiting a visual preference (see Supplemental Materials for details). As shown in Fig. 4 (right panel), the Congruency effect emerges in the 200–400 ms time-window after face onset, which is the first time-window in which infants performed saccades to face images. This Congruency effect persisted throughout the first 1400 ms ($ts \geq 2.61$, FDR corrected $ps \leq .036$, note: no baby looked at either face from 0 to 200 ms, as the trials started with central fixation and this is faster than a reactive saccade can be programmed at this age). Starting from 1400 ms after the onset, no significant Congruency effect was observed ($ts \leq 2.21$, FDR corrected $ps \geq .065$) until the 2800 ms ($t = 2.80$, FDR corrected $p = .025$). Given that the location of the emotional face cannot be predicted in each given trial, this early Congruency effect provides further evidence that the preceding vocal cue biases infants face perception rather than affecting some other system or a downstream ability such as visual attention or perceptual comparison.

Considering this effect in more detail, we examined visual preference in the Congruent and Incongruent conditions (i.e., whether infants were looking at the emotional face greater than 50%). In the Congruent conditions, infants showed a preference for the emotional faces over the neutral face from 400 ms after onset (one-sample *t*-test, $ps \leq .046$, Fig. 5). However, an opposite looking preference was found in the Incongruent condition within the initial face presentation period (from 200 ms to 1000 ms after onset, one-sample *t*-test, $ps \leq .001$). Infants preferred looking at the neutral face over the emotional face that was Incongruent with the emotional sounds. The visual preference for the neutral face was replaced by a preference for the emotional face from 1800 to 2600 ms (one-sample *t*-test, $ps \leq .043$). These findings

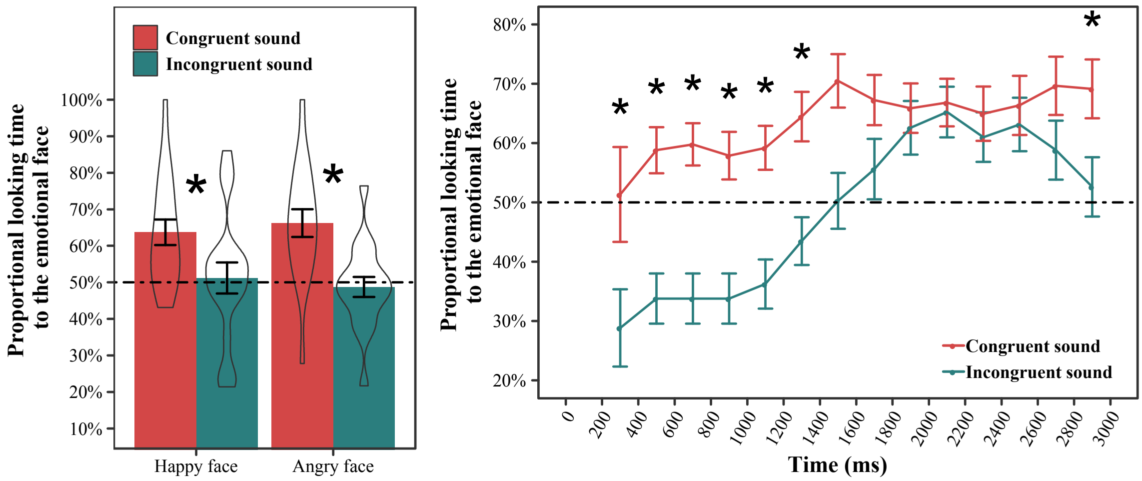

Fig. 4. Mean proportional looking to the emotional faces in the Congruent and Incongruent conditions (Left) and mean proportional looking to the emotional for each time window (Right). Error bars represent unit standard errors. The asterisks indicate a significant difference between the Congruent and Incongruent conditions (FDR corrected $p < .050$). The overlaid shapes represent the distribution of individual response. The dash line represents the chance level (50%).

suggest that, in the early stages of face presentation phase, infants' perception of the two faces was biased by the emotional vocal sounds. The face whose facial expression was closest to that of the emotional sound was boosted. In the Congruent conditions, the facial expression in the emotional face was more consistent with the vocal sound as compared to that in the neutral face. By contrast, in the Incongruent conditions, although none of the faces showed the facial expression matching that of the preceding vocal sound, the neutral face expression was closer to the emotion presented in the vocal sound as compared to that in the Incongruent emotional face. Therefore, infants exhibited visual preference for the emotional face in the Congruent conditions and for the neutral face in the Incongruent conditions.

3. Experiment 2

Expt. 1 provided evidence for the top-down modulation of infants' face perception based on emotional cues. This effect was revealed through biases in infant's initial looking latency based on preceding emotional sounds. However, previous research has shown that infants are able to engage in cross-modal matching where, during simultaneous presentation of vocalizations and emotional sounds, their attention is guided to the corresponding emotional face (e.g., Montague & Walker-Andrews, 2002). Cross-modal matching is not thought to be a specific top-down process and may be mediated by bottom-up processes in which higher-level associative regions (like the amygdala) compare the two inputs and then guide attention/eye movements (as opposed to using feedback connections to modulate face perception regions).

While there are numerous differences between the current paradigm and cross-modal matching paradigms (see Fig. 3, Table 1), Expt. 2 clarifies the difference between top-down changes in perception found in Expt. 1 and cross-modal matching. Here, we kept the perceptual signals identical but changed their presentation from being **sequential** (audio preceding visual stimulation) to **simultaneous**. With simultaneous presentation, infants were able to engage in cross-modal matching, but it was more difficult for top-down modulation of face perception as the signals are being presented simultaneously. If the

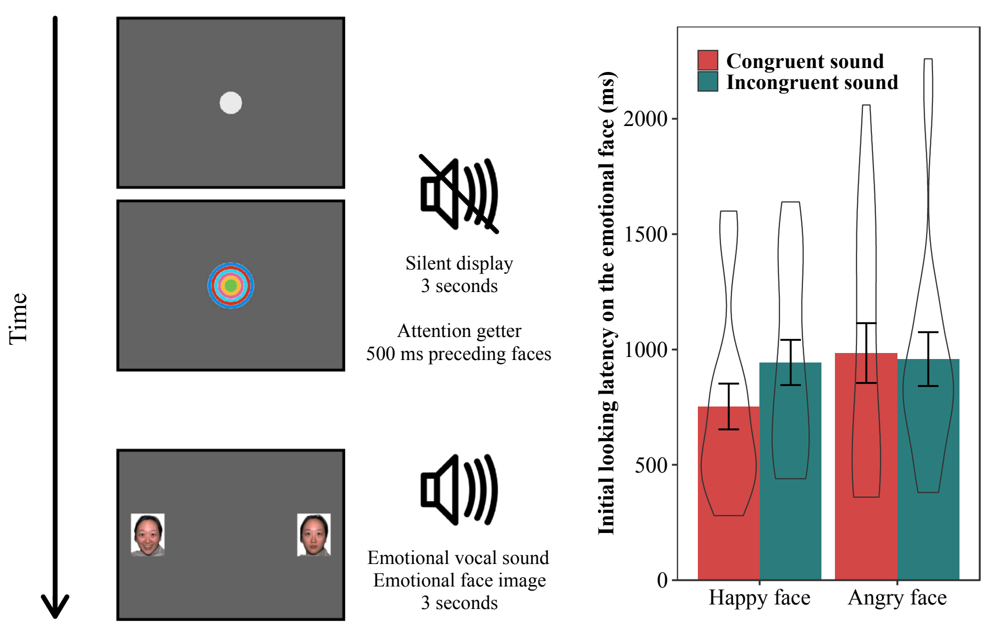

Fig. 5. A representative trial for Expt. 2 (Left) and infants' initial looking latency to the emotional face (Right). The overlaid shapes represent the distribution of individual response for each condition. Error bars represent unit standard errors.

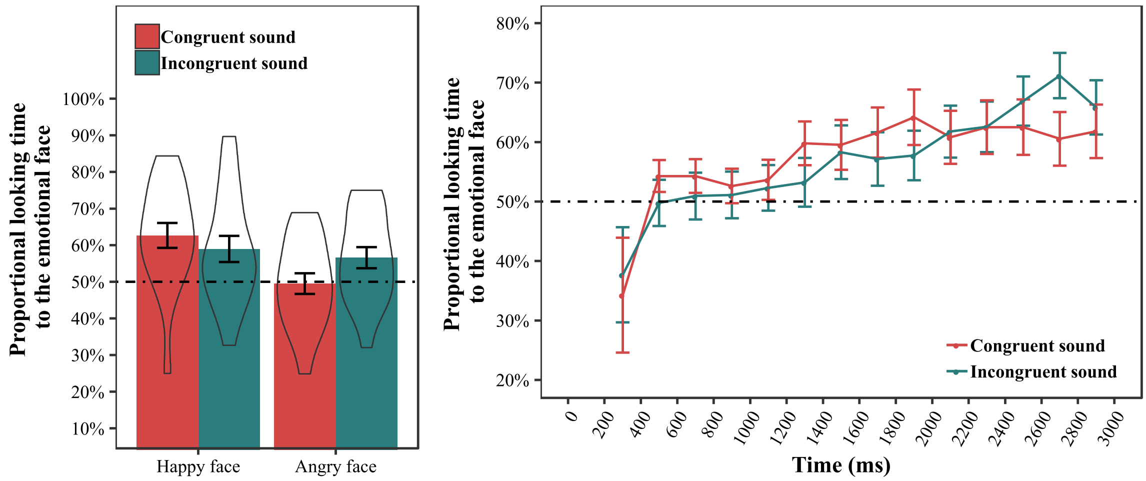

Fig. 6. Mean proportional looking to the emotional faces in the Congruent and Incongruent conditions (Left) and mean proportional looking to the emotional for each time window (Right). Error bars represent unit standard errors. The overlaid shapes represent the distribution of individual response. The dash line represents the chance level (50%).

findings of Expt. 1 were caused by the cross-modal matching, we would find highly similar results as Expt. 1 (albeit a bit delayed). However, if cross-modal matching is a distinct process from top-down effects on perception (as we propose), we should find different effects from Expt. 1.

3.1. Method

3.1.1. Participants

Eighteen infants (11 females, 7 males) participated in the current experiment. These infants were 9 months of age (M = 263.80 days, SD = 21.95 days). All of the infants were Asian, who had not seen the face images used in the current experiment before their participation. Four additional infants participated in the experiment but were excluded from the final sample due to insufficient data collection (i.e., less than 1 trial per condition) resulting from crying (n = 1) or not looking at the screen (n = 3). None of these infants participated in Expt. 1.

3.1.2. Procedure and Materials

The procedure and materials in Expt. 2 were identical to those in Expt. 1 except the onset time of the vocal sounds. Specifically, the vocal sounds and face images were presented concurrently for 3 s (Fig. 5, left panel).

3.2. Results and discussion

3.2.1. Simultaneous presentation of vocal sounds doesn't bias initial looking behavior

A similar 2 (Congruent vs. Incongruent) × 2 (Happy vs. Angry) repeated-measures ANOVA was performed on the initial looking latency (Fig. 5, right panel). The initial looking latency in the Congruent trials (M = 806 ms, SD = 440 ms) was not different from that of the Incongruent trials (M = 880 ms, SD = 429 ms, $F(1, 17) = 0.54$, p = .474, η_p^2 = .03, Fig. 5). The main effect of Face expression was significant, $F(1, 17) = 4.78$, p = .043, η_p^2 = .22, in which the looking latency in the happy face trials (M = 758 ms, SD = 349 ms) was faster than that in the angry face trials (M = 927, SD = 494 ms). The interaction was not significant, $F(1, 17) = 0.22$, p = .646, η_p^2 = .01. When the vocal sound and the emotional faces were represented simultaneously, infants' initial look latency at the emotional face was not affected by the vocal sounds they heard.

To further strengthen our evidence of top-down modulation on face perception in Expt. 1, we compared the Congruency effect between Expts. 1 and 2. A mixed ANOVA on the initial looking latency was conducted with Congruency and Facial expression as within-subject independent variables and Experiment as a between-subject independent variable. Supporting a difference between experiments (i.e., sequential vs. simultaneous presentation of the emotional sounds), we found a marginally significant interaction between Experiment and Congruency, $F(1, 34) = 4.08$, p = .051, η_p^2 = .03. This interaction, though marginally significant, indicated that the Congruency effect was different between the two experiments. This result indicates that the mechanisms driving infants' initial looking differs between the two experimental settings. Thus, the Congruency effect on infants' initial looking latency found in Expt. 1 reflects a top-down modulation of perception rather than a rapid detection of cross-modal correspondence. The results also revealed a main effect of Congruency, $F(1, 34) = 9.39$, p = .004, η_p^2 = .06, but we did not find any other effect reached significance ($ps \geq 0.174$).

3.2.2. Proportional looking time for simultaneous presentation

A similar 2 (Congruent vs. Incongruent) × 2 (Happy vs. Angry) repeated-measures ANOVA was performed on looking preference across the 3-seconds stimulation presentation phase. The main effect of Congruency was not significant ($M_{\text{Congruent}}$ = 56.08%, $SD_{\text{Congruent}}$ = 14.68%, $M_{\text{Incongruent}}$ = 57.76%, $SD_{\text{Incongruent}}$ = 13.61%, $F(1, 17) = 0.23$, p = .635, η_p^2 = .01). The main effect of Face expression was significant, $F(1, 17) = 7.08$, p = .02, η_p^2 = .29. The looking preference for emotional faces over neutral ones was larger in the happy face conditions (M = 60.81%, SD = 14.65%) than that in the angry face conditions (M = 53.03%, SD = 12.51%). The interaction was not significant ($F(1, 17) = 2.67$, p = .121, η_p^2 = .14). Thus, the simultaneously played emotional vocal sounds did not affect infants' visual preference between an emotional face and a neutral face across trial types (Fig. 6, left panel).

3.2.3. Time-course of infants' preference is unaffected by simultaneous vocal cues

Similar to Expt. 1, we collapsed the proportional looking in the happy and angry face trials to focus on the Congruency effect. As shown in the right panel of Fig. 6, we did not find significant Congruency effect in any time window ($ts \leq 2.45$, FDR corrected $ps \geq .355$). Infants did not exhibit the Congruency effect in their face looking preference throughout the 3-second face presentation period.

Together, the results of Expt. 2 did not show any Congruency effect on face perception like we found in Expt. 1. Specifically, infants did not exhibit classical cross-modal matching (longer looking at the congruent face), suggesting infants were unable to perform cross-modal matching with such small face images of short durations on a near-peripheral layout. The failure to find the Congruency effect in Expt. 2 further confirms that infants are using preceding emotional vocal sounds to augment visual perception of faces when vocal sounds precede the visual display, as in Expt. 1. It should also be noted that the current study is not designed to find cross-modal matching effects between vocal cues and facial expressions. As summarized in Table 1, there are many methodological differences between the current experiment and previous cross-modal matching experiments, even beyond the simultaneity of the vocal and visual presentation, including the length of the trial and the size of the facial stimuli.

4. Experiment 3

Expt. 1 established that 9-month-olds' face perception can be flexibly modulated by top-down signals engendered by emotional vocal sounds. A prior neuroimaging study reported an age-related change in neural markers of top-down modulation of face perception from 5 to 9 months of age (e.g., Vogel et al., 2012). To further strengthen the evidence of top-down modulation of face perception and to provide convergent evidence of developmental change in top-down modulation, we employed the same paradigm as Expt. 1 with 6.5-month-old infants. We hypothesized that infants at this age would not show evidence of top-down modulation of face perception (i.e., no differences in initial looking latency based on preceding emotional cues).

4.1. Method

4.1.1. Participants

Eighteen infants (7 females, 11 males) participated in the current experiment. These infants were 6.5 months of age ($M = 199.78$ days, $SD = 22.63$ days). All of the infants were Asians who had not seen the face images used in the experiment before their participation. Four additional infants participated in the experiment but were excluded from the final sample due to insufficient data collection (i.e., less than 1 trial per condition) resulting from crying ($n = 1$) or not looking at the screen ($n = 3$). Expt. 3 was performed at the same time as Expt. 1. None of the participants participated in Expt. 1 or 2.

4.1.2. Procedure and materials

The procedure and materials in Expt. 3 were identical to those in Expt. 1.

4.2. Results and discussion

4.2.1. Initial looking latency is not affected by corresponding vocal cues in six-month-olds

The same repeated-measures Congruency × Face expression ANOVA was performed. We found a significant interaction ($F(1, 17) = 4.66$, $p = .046$, $\eta_p^2 = .22$), indicating that any effect of congruency differed between the happy face and angry face trials (Fig. 7, upper left panel). Specifically, there was no significant difference for the happy faces ($t(17) = 0.63$, $p = .536$). The looking latency in the Incongruent condition was faster than that of the Congruent condition ($t(17) = 2.54$, $p = .021$), which was opposite to the hypothesized direction. Moreover, no main effect reached significance: Congruency, $F(1, 17) = 1.73$, $p = .206$, $\eta_p^2 = .09$; Face expression, $F(1, 17) = 0.21$, $p = .656$, $\eta_p^2 = .01$. These results suggest that infants at 6.5 months of age could not use the emotional vocal sounds to bias their perception of emotional faces, unlike infants at 9 months of age.

We further examined age difference in the Congruency effect by conducting a mixed ANOVA on the initial looking latency with Congruency and Facial expression as within-subject independent variables and Age (9 vs. 6.5 months) as a between-subject independent variable. We found a significant interaction between Age and Congruency, $F(1, 34) = 10.20$, $p = .003$, $\eta_p^2 = .23$, suggesting an age-related development in top-down modulation of perception. We did not find any other effect reached significance ($ps \geq 0.211$).

4.2.2. Looking preference is not affected by preceding vocal cues in six-month-olds

The same repeated-measures ANOVA on the preference for emotional faces showed that neither the main effect of Congruency ($M_{\text{Congruent}} = 57.17\%$, $SD_{\text{Congruent}} = 18.96\%$, $M_{\text{Incongruent}} = 56.87\%$, $SD_{\text{Incongruent}} = 18.47\%$, $F(1, 17) = 0.004$, $p = .951$, $\eta_p^2 < .001$) nor the main effect of Face expression reached significance ($M_{\text{Happy face}} = 58.01\%$, $SD_{\text{Happy face}} = 17.58\%$, $M_{\text{Angry face}} = 56.03\%$, $SD_{\text{Angry face}} = 19.73\%$, $F(1, 17) = 0.16$, $p = .697$, $\eta_p^2 = .01$). The interaction was also not significant ($F(1, 17) = 0.03$, $p = .856$, $\eta_p^2 = .002$). These results indicated that the preceding emotional vocal sounds did not affect 6.5-month-olds visual preference between an emotional face and a neutral face (Fig. 7, upper right panel). Moreover, one-sample t-tests indicated that infants did not show visual preference for emotional faces in any condition ($ps \geq .059$).

4.2.3. Looking preference is not affected by initial looking behavior in six-month-olds

We examined the time-course of proportional looking in the 6.5-month-olds with the same analysis used in Expt. 1. We did not find significant Congruency effect in any time window ($ts \leq 2.38$, FDR corrected $ps \geq 0.406$), suggesting 6.5-month-olds were not able to use the emotional vocal sounds to guide their attention to the corresponding emotional faces throughout the entire face presentation period (Fig. 7, lower panel). Moreover, we found that infants looked longer at the emotional face in the 200–400 ms time window with a one-sample t-test ($Mean = 69.25\%$, $SD = 38.92\%$, $t(15) = 2.16$, $p = .047$). This initial looking preference indicated that infants could tell the difference between an emotional face and a neutral one.

In sum, the absence of the Congruency effect in the current experiment indicated that infants were not able to use the emotional vocal sounds to modulate their perception of facial expression at 6.5 months. This failure to use vocalization as a top-down signal did not arise from an insufficient knowledge of the emotional vocal sounds and facial expressions; infants as young as 3–6.5 months of age were able to associate emotional vocal sounds with facial expression or even with more abstract body gestures using these same stimuli (Kahana-Kalman & Walker-Andrews, 2001; Zieber et al., 2013). Thus, the age-difference in the Congruency effect suggests that the top-down influence on face perception may develop between 6.5 and 9 months of age, which further strengthens our hypothesis that the top-down processing found in the infant brain supports adaptive developmental change in perception.

5. General discussion

The current study presents the most direct behavioral investigation of top-down modulation of perception in young infants to date. We specifically developed a method to index infant face perception on each trial. With this novel paradigm, we found that infants employed their knowledge of emotion, cued auditorily, to bias their face perception via top-down modulation. We ruled out the possibility that the observed effects were due to rapid cross-modal matching of emotional signals (Expt. 2).

This behavioral finding of a top-down influence on perception in infants corroborates our latest understanding of early availability of feedback neural connections (Emberson, 2017; Emberson et al., 2015; Kouider et al., 2015). Current findings indicate that these connections are available and influence behavior in the first year of postnatal life. The top-down signals engendered by preceding emotional sounds effectively modulated infants' perception of faces. This top-down modulation may modify the responsiveness of the neurons representing the two faces (i.e., a neural and an emotional face) presented in a visual field. The activity of neurons representing the corresponding face

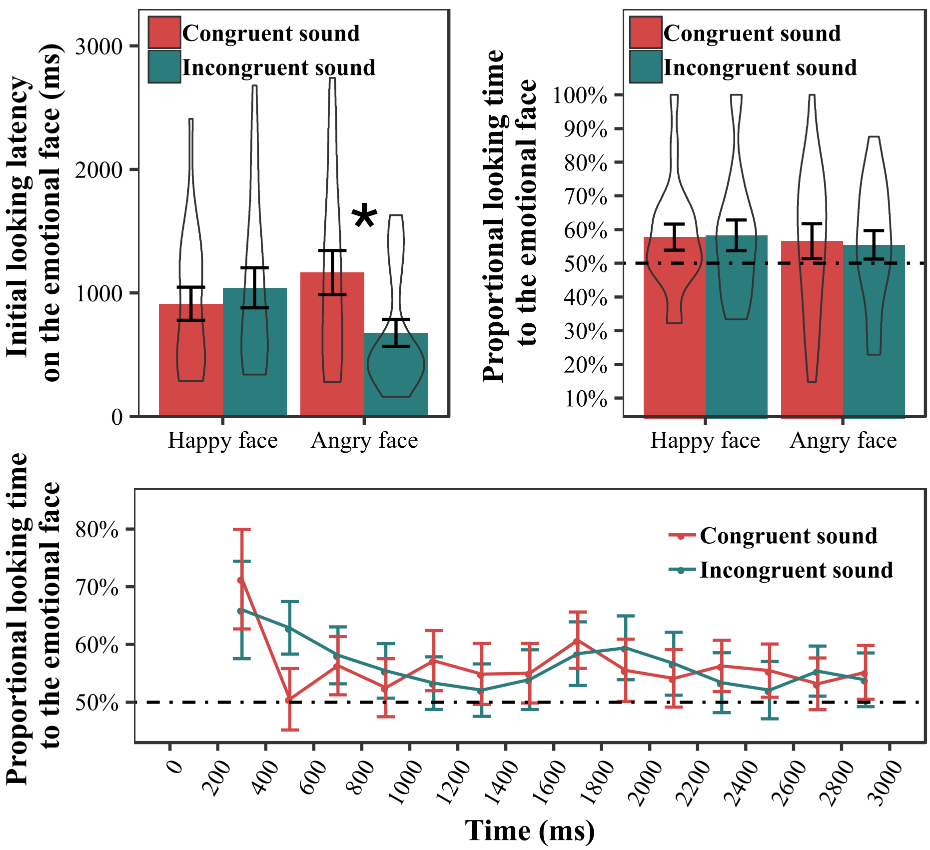

Fig. 7. Mean latency to look at the happy and angry faces (Upper Left), mean proportional looking to the emotional faces (Upper Right), and mean proportional looking to the emotional face for each time window (Lower). Error bars represent unit standard errors. The asterisks indicate a significant difference between the Congruent and Incongruent conditions (FDR corrected $p < .050$). The overlaid shapes represent the distribution of individual response. The dash line represents the chance level (50%).

would be enhanced by the top-down signals. By contrast, the activity of neurons representing the irrelevant face would be inhibited by the top-down signals. As a result, the relative differences in perception of the two faces leads to differences in attentional orienting as infants perceive the congruent face first or more strongly.

The current behavioral findings closely follow previous neuroimaging findings. While not originally considered in the framework of top-down modulation, Vogel et al. (2012) used a similar sequential paradigm to show that 9-month-olds' face perception-related ERP components (N290 and P400) were modulated by the congruency of preceding emotional vocal sounds. Our behavioral findings closely follow this important result, and, with other neuroimaging studies with infants (e.g., Emberson et al., 2015; Kouider et al., 2015), suggest that top-down modulation is available early in life and is able to modulate perception, as evidenced behaviorally.

Though both imaging and behavioral evidence convergently indicate the availability of top-down modulation of perception early in life, the specific source of top-down signals is still unclear. As with all top-down effects (defined here as the modulation of perceptual cortices through feedback connections), there could be numerous sources. In the adult brain, the visual system is subject to top-down modulation from systems such as the frontal cortex, reward circuitry, and influence from memory systems such as the hippocampus, in addition to attentional systems (e.g., Emberson, 2017, 2019).

The amygdala is a highly likely source of top-down modulation in the current task. In this case, the emotional signals in vocal sounds activate the amygdala, which in turn generates top-down signals to modulate face perception in the Fusiform Face Area (FFA). It is established in adults and non-human primates that the amygdala has direct feedback connections to the FFA and other visual cortices (e.g., Amaral, Behniea, & Kelly, 2003; Anderson & Phelps, 2001; Herrington, Taylor, Grupe, Curby, & Schultz, 2011). Through feedback connections, the amygdala has the ability to modulate visual perception based on emotional signals (Furl, Henson, Friston, & Calder, 2013; Phelps, Ling, & Carrasco, 2006). A broad range of feedforward and feedback neural connections between the amygdala and cortical regions develop early in life (Johnson, 2005). These early-developed reciprocal connections not only allow infants to perceive emotional signals in various perceptual domains (Walker, 1982), but, we propose, also result in effective top-down modulation in perceptual systems.

The most commonly considered source of top-down modulation is the attentional system. While we do not believe that attention is the source of top-down modulation in the current study, we cannot rule it out in all cases. Previous research has found that top-down attention can guide infants' eye movements to predicted spatial locations (e.g., Richardson & Kirkham, 2004; Tummeltshammer & Amso, 2018). This mechanism is not a possible explanation for the current findings because the spatial information of the faces was independent of the audio cues. Similarly, paradigms that have elicited covert shifts of attention in young infants (e.g., 4-month-olds, Johnson, Posner, & Rothbart, 1994) also employed targets that were spatially predictable and only involved the presentation of a single target at a time. For covert spatial attention to be involved in the current paradigm, infants would have to attend to two targets (either in parallel or in series) without moving their eyes and with no valid spatial cue to help them focus on a single one. Then, after perceptual analysis, they could move their eyes in a biased fashion towards the congruent face. To our knowledge, there is no evidence that such a sophisticated ability is available to infants, nor would they employ this without a strong set of instructions (as in an adult study). Thus, infants' use of covert attention to perform this task is highly unlikely.

Another possible role of attention is through the activation of a perceptual template. In this case, the emotional sound would activate a working memory representation of the corresponding emotional face as the to-be searched target, which could then augment perception of the corresponding face. While this type of attention has been demonstrated in the mature brain (Chelazzi, Duncan, Miller, & Desimone, 1998; Roelfsema, Lamme, & Spekreijse, 1998), it has not been established in infants (e.g., Wu & Zhao, 2017) making it a less plausible explanation than the direct relationship between the amygdala and the FFA, proposed above.

The current findings highlight the surprising flexibility in top-down modulation of perception in the first year of postnatal life. Infants had to rely on emotional sounds (presented for only 3 s) to adjust their perception of facial expression of multiple actresses without training or spatial contingency on a trial-by-trial basis. Their ability to do this indicates a highly flexible and powerful top-down mechanism is available to infants starting early in life. Moreover, the time-course analysis revealed a looking preference for the neutral face when the audio sound was incongruent with either of the faces. This looking preference suggests that top-down modulation could guide infants' looking at the most probable face rather than a specific matching or template effect (i.e., when one hears a happy vocalization, the more probable source is a neutral face relative to an angry face). While previous work has demonstrated changes in perception based on experience, this type of flexibility in perception has never been demonstrated before (e.g., Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Lew-Williams & Saffran, 2012; Needham et al., 2006; Quinn et al., 2006). This kind of flexibility in perception is a major benefit to the use of top-down signals.

Moreover, current findings suggest that top-down modulation of perception could be broadly used by infants in their daily lives. Prior studies have used a familiarization or training session to provide infants with new knowledge and examined how this knowledge affected infants' perceptual performance or neural responses in either identical or highly similar settings (Emberson et al., 2015; Kouider et al., 2015). By contrast, the current study did not include any training or familiarization procedure, and, instead, examined whether infants can use knowledge acquired outside the lab to modify perception in novel settings with novel stimuli. The ability to achieve robust top-down modulation of perception in this scenario suggests that infants are readily capable of using their knowledge to modulate their perception in a top-down fashion in their daily lives.

Unlike 9-month-olds, 6.5-month-olds did not show evidence of top-down modulation of face perception. While the inability of 6.5-month-olds to benefit from this top-down augmentation may seem incompatible with functional imaging results that demonstrate top-down neural modulation in 6-month-olds (Emberson et al., 2015), this, in fact, suggests that neither all types of top-down modulation, nor the conditions in which they are being tested, are equal. The current task is unique for its lack of a familiarization procedure, which requires the use of top-down modulation in novel settings with novel stimuli. This generalization of top-down modulation might not be developed in young infants. In line with this hypothesis, neuroimaging studies that have demonstrated top-down neural modulations in perceptual systems in 6-month-olds employ training procedures and do not require any generalization (Emberson et al., 2015). In contrast, studies examining top-down modulation of perception in novel settings reported it only in 9-month-olds, not in younger infants (Barry-Anwar, Hadley, & Scott, 2018; Grossmann, Striano, & Friederici, 2006; Vogel et al., 2012).

Another (not mutually exclusive) possibility is that top-down modulation of perception becomes increasingly specialized with age. For young infants, top-down modulation of perception may be limited to general perceptual processing, such as detection of an upcoming visual events. In contrast, the ability to modulate more complex and specialized visual processing (e.g., face and object correspondences) is not developed at this stage of life. With increased experience, older infants may become increasingly capable of top-down modulation of higher-level visual processing, such as face perception. This specialization in development account is supported by current neural and behavioral evidence. Apart from the current study, top-down modulation for specific faces and objects has been found in neuroimaging studies with infants older than 9 months of age (Gliga et al., 2010; Kouider et al., 2015; Vogel et al., 2012) but not in younger infants (Vogel et al., 2012). Studies with younger infants do not test for modulation of specific objects nor specialized objects, like different faces (e.g., Emberson et al., 2015). More studies are needed to examine this specialization developmental account to disentangle any potential specialization of top-down modulation of perception and the specialization of perceptual representations themselves via fine-grained perceptual manipulations (e.g., face inversion effect).

This work broadly stems from the theoretical accounts of Emberson (2017), Hadley et al. (2014), and Markant and Scott (2018), which propose that top-down processes may be available starting early in life and that top-down mechanisms may guide perceptual development. The current findings support all of these accounts. We also provide some initial evidence that these effects emerge from 6.5 to 9 months of age, which suggests that the emergence of top-down processes may be related to perceptual narrowing. The potential source of this top-down modulation of perception (the amygdala directly to the FFA or via attentional template matching) could differentiate the proposals. Scott, Markant, and colleagues propose the reciprocal interaction of attention and perceptual learning in the developmental of face perception. They propose that several different developmental tasks (e.g., language development & motor development) may influence the interaction of attention and perceptual learning. Emberson (2017) proposes that there are many sources of top-down modulation in perceptual development and could accommodate either possible origin of top-down modulation of perception.

In sum, we find that infants exhibit an active, flexible, and adaptive ability to use knowledge of emotions to modulate face perception. This finding suggests that top-down modulation of perception does not emerge later in life (e.g., with the emergence of robust long-range or feedback neural connections) but is available and quite sophisticated early in development. This encourages the consideration of how top-down mechanisms can support the development of face perception and perceptual development more broadly.

Acknowledgements

This research was supported by grants from the National Institutes of Health (R00 4R00HD076166-02), McDonnell Foundation (220020505) and the National Science Foundation of China (31771233). We thank Ramesh S. Bhatt for sharing audio stimuli, Paul C. Quinn, Richard N. Aslin, and Sabine Kastner for their comments on the early versions of the manuscript. We also thank Meiyu Li, Zhijun Wu, Chun Su, Yuwen Fang, and Shaoying Liu for their assistance in data collection.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.104019.

References

Amaral, D. G., Behniea, H., & Kelly, J. L. (2003). Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. *Neuroscience, 118*, 1099–1120. https://doi.org/10.1016/S0306-4522(02)01001-1.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411*, 305–309. https://doi.org/10.1038/35077083.

Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience, 5*, 617–629. https://doi.org/10.1038/nrn1476.

Barry-Anwar, R., Hadley, H., & Scott, L. S. (2018). Differential neural responses to faces paired with labels versus faces paired with noise at 6- and at 9-months. *Vision Research*. https://doi.org/10.1016/j.visres.2018.03.002.

Batardière, A., Barone, P., Knoblauch, K., Giroud, P., Berland, M., Dumas, A.-M., & Kennedy, H. (2002). Early specification of the hierarchical organization of visual cortical areas in the macaque monkey. *Cerebral Cortex, 12*, 453–465. https://doi.org/10.1093/cercor/12.5.453.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*, 298–304. https://doi.org/10.1111/j.0956-7976.2005.01531.x.

Bullier, J. (2001). Feedback connections and conscious vision. *Trends in Cognitive Sciences,* *5*, 369–370. https://doi.org/10.1016/S1364-6613(00)01730-7.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology, 80*, 2918–2940. https://doi.org/10.1152/jn.1998.80.6.2918.

Chen, M., Yan, Y., Gong, X., Gilbert, Charles D., Liang, H., & Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. *Neuron, 82*, 682–694. https://doi.org/10.1016/j.neuron.2014.03.023.

Emberson, L. L. (2017). How does experience shape early development? Considering the role of top-down mechanisms. In J. B. Benson (Ed.), Advances in Child Development and Behavior (Vol. 52, pp. 1–41). JAI.

Emberson, L. L. (2019). How does learning and memory shape perceptual development in infancy? *Journal of Learning and Motivation, 70*, 129–160. https://doi.org/10.1016/bs.plm.2019.03.003.

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 of the United States of America, 112*, 201510343–201519590. https://doi.org/10.1073/pnas.1510343112.

Furl, N., Henson, R. N., Friston, K. J., & Calder, A. J. (2013). Top-down control of visual responses to fear by the amygdala. *Journal of Neuroscience, 33*, 17435–17443. https://doi.org/10.1523/Jneurosci.2992-13.2013.

Gao, W., Alcauter, S., Elton, A., Hernandez-Castillo, C. R., Smith, J. K., Ramirez, J., & Lin, W. (2015). Functional network development during the first year: Relative sequence and socioeconomic correlations. *Cerebral Cortex, 25*, 2919–2928. https://doi.org/10.1093/cercor/bhu088.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews: Neuroscience, 14*, 350–363. https://doi.org/10.1038/nrn3476.

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Reviews: Neuroscience, 19*, 123–137. https://doi.org/10.1038/nrn.2018.1.

Gliga, T., Volein, A., & Csibra, G. (2010). Verbal labels modulate perceptual object processing in 1-year-old children. *Journal of Cognitive Neuroscience, 22*, 2781–2789. https://doi.org/10.1162/jocn.2010.21427.

Grassi, P. R., Zaretskaya, N., & Bartels, A. (2017). Scene segmentation in early visual cortex during suppression of ventral stream regions. *Neuroimage, 146*, 71–80. https://doi.org/10.1016/j.neuroimage.2016.11.024.

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science, 9*, 309–315. https://doi.org/10.1111/j.1467-7687.2006.00494.x.

Hadley, H., Rost, G. C., Fava, E., & Scott, L. S. (2014). A mechanistic approach to cross-domain perceptual narrowing in the first year of life. *Brain Sciences, 4*, 613–634. https://doi.org/10.3390/brainsci4040613.

Herrington, J. D., Taylor, J. M., Grupe, D. W., Curby, K. M., & Schultz, R. T. (2011). Bidirectional communication between amygdala and fusiform gyrus during facial recognition. *Neuroimage, 56*, 2348–2355. https://doi.org/10.1016/j.neuroimage.2011.03.072.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development, 16*, 405–422. https://doi.org/10.1016/0163-6383(93)80001-O.

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature, 394*, 784. https://doi.org/10.1038/29537.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience, 6*, 766–774. https://doi.org/10.1038/nrn1766.

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1994). Facilitation of saccades toward a covertly attended location in early infancy. *Psychological Science, 5*(2), 90–93.

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development, 72*, 352–369. https://doi.org/10.1111/1467-8624.00283.

Kok, P., Jehee, J. F., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron, 75*, 265–270. https://doi.org/10.1016/j.neuron.2012.04.034.

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nat Commun, 6*, 8537. https://doi.org/10.1038/ncomms9537.

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition, 122*, 241–246. https://doi.org/10.1016/j.cognition.2011.10.007.

Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., ... Murayama, M. (2015). A top-down cortical circuit for accurate sensory perception. *Neuron, 86*, 1304–1316. https://doi.org/10.1016/j.neuron.2015.05.006.

Markant, J., & Scott, L. S. (2018). Attention and perceptual learning interact in the development of the other-race effect. *Current Directions in Psychological Science, 27*, 163–169. https://doi.org/10.1177/0963721418769884.

Menon, V. (2013). Developmental pathways to functional brain networks: Emerging principles. *Trends in Cognitive Sciences, 17*, 627–640. https://doi.org/10.1016/j.tics.2013.09.015.

Montague, D. P. F., & Walker-Andrews, A. S. (2002). Mothers, fathers, and infants: The role of person familiarity and parental involvement in infants' perception of emotion expressions. *Child Development, 73*, 1339–1352. https://doi.org/10.1111/1467-8624.00475.

Needham, A., Cantlon, J. F., & Ormsbee Holley, S. M. (2006). Infants' use of category knowledge and object attributes when segregating objects at 8.5 months of age. *Cognitive Psychology, 53*, 345–360. https://doi.org/10.1016/j.cogpsych.2006.05.003.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–325. https://doi.org/10.1017/S0140525X00003241.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*, 773–783. https://doi.org/10.1038/nrn2920.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science, 17*, 292–299. https://doi.org/10.1111/j.1467-9280.2006.01701.x.

Price, D. J., Kennedy, H., Dehay, C., Zhou, L., Mercier, M., Jossin, Y., ... Molnár, Z. (2006). The development of cortical connections. *European Journal of Neuroscience, 23*, 910–920. https://doi.org/10.1111/j.1460-9568.2006.04620.x.

Quinn, P. C., Schyns, P. G., & Goldstone, R. L. (2006). The interplay between perceptual organization and categorization in the representation of complex visual patterns by young infants. *Journal of Experimental Child Psychology, 95*, 117–127. https://doi.org/10.1016/j.jecp.2006.04.001.

Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General, 133*, 46–62. https://doi.org/10.1037/0096-3445.133.1.46.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*, 376. https://doi.org/10.1038/26475.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 2408–2412. https://doi.org/10.1073/pnas.0908239106.

Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice, and touch: Comparisons and convergence. *Trends in Cognitive Sciences, 21*, 216–228. https://doi.org/10.1016/j.tics.2017.01.001.

Schwiedrzik, C. M., & Freiwald, W. A. (2017). High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron, 96*, 89–97.e84. https://doi.org/10.1016/j.neuron.2017.09.007.

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science, 16*, 197–201. https://doi.org/10.1111/j.1467-8721.2007.00503.x.

Soken, N. H., & Pick, A. D. (1999). Infants' perception of dynamic affective expressions: Do infants distinguish specific expressions? *Child Development, 70*, 1275–1282. https://doi.org/10.1111/1467-8624.00093.

Spelke, E. S. (1981). The infant's acquisition of knowledge of bimodally specified events. *Journal of Experimental Child Psychology, 31*, 279–299. https://doi.org/10.1016/0022-0965(81)90018-7.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews: Neuroscience, 15*, 745–756. https://doi.org/10.1038/nrn3838.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science, 314*, 1311–1314.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262*, 685. https://doi.org/10.1126/science.8235589.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109–139. https://doi.org/10.1146/annurev.ne.19.030196.000545.

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*, 242–249. https://doi.org/10.1016/j.psychres.2008.05.006.

Tummeltshammer, K., & Amso, D. (2018). Top-down contextual knowledge guides visual attention in infancy. *Developmental Science, 21*, e12599. https://doi.org/10.1111/desc.12599.

Vogel, M., Monesson, A., & Scott, L. S. (2012). Building biases in infancy: The influence of race on face and voice emotion matching. *Developmental Science, 15*, 359–372. https://doi.org/10.1111/j.1467-7687.2012.01138.x.

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology, 33*, 514–535. https://doi.org/10.1016/0022-0965(82)90063-7.

Wu, R., & Zhao, J. (2017). Prior knowledge of object associations shapes attentional templates and information acquisition. *Frontiers in Psychology, 8*, 1–6. https://doi.org/10.3389/fpsyg.2017.00843.

Zieber, N., Kangas, A., Hock, A., & Bhatt, R. S. (2013). Infants' perception of emotion from body movements. *Child Development, 85*, 675–684. https://doi.org/10.1111/cdev.12134.