

Overcoming the other-race effect in infancy with multisensory redundancy: 10–12-month-olds discriminate dynamic other-race faces producing speech

Nicholas J. Minar¹ | David J. Lewkowicz² 

¹Institute for the Study of Child Development, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA

²Department of Communication Sciences and Disorders, Northeastern University, Boston, MA, USA

Correspondence

David J. Lewkowicz, Department of Communication Sciences and Disorders, Northeastern University, 360 Huntington Ave., 503BK, Boston, MA 02115, USA.
Email: d.lewkowicz@northeastern.edu

Funding Information

This study was supported by Grant R01HD057116 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to DJL.

Abstract

We tested 4–6- and 10–12-month-old infants to investigate whether the often-reported decline in infant sensitivity to other-race faces may reflect responsiveness to static or dynamic/silent faces rather than a general process of perceptual narrowing. Across three experiments, we tested discrimination of either dynamic own-race or other-race faces which were either accompanied by a speech syllable, no sound, or a non-speech sound. Results indicated that 4–6- and 10–12-month-old infants discriminated own-race as well as other-race faces accompanied by a speech syllable, that only the 10–12-month-olds discriminated silent own-race faces, and that 4–6-month-old infants discriminated own-race and other-race faces accompanied by a non-speech sound but that 10–12-month-old infants only discriminated own-race faces accompanied by a non-speech sound. Overall, the results suggest that the ORE reported to date reflects infant responsiveness to static or dynamic/silent faces rather than a general process of perceptual narrowing.

RESEARCH HIGHLIGHTS

- We investigated the often-reported emergence of a perceptual insensitivity to faces from unfamiliar races – commonly known as the other-race effect (ORE)—by testing 4–6- and 10–12-month-old infants.
- We studied discrimination of either own- or other-race faces that could be either seen and heard uttering a speech sound, seen silently uttering a speech sound, or seen uttering a speech sound together with a non-speech sound.
- 4–6-month-old infants discriminated own- or other-race faces as long as the faces were accompanied by a speech or non-speech sound while 10–12-month-old infants discriminated own- or other-race faces but only when they were accompanied by a speech sound.
- Results show that older infants maintain their sensitivity to other-race faces as long as such faces are specified by the redundant and dynamic multisensory perceptual cues that are usually part of their everyday social experiences.

1 | INTRODUCTION

Typically, adults are more proficient at recognizing, discriminating, and remembering the faces of their own race than those of other races (Bothwell, Brigham, & Malpass, 1989; Lindsay, Jack, & Christian, 1991). This is known as the other-race effect (ORE) and studies have found that it emerges during the first year of life and that it reflects the tuning of the perceptual system by exposure to specific face categories (Kelly et al., 2009; Kelly, Quinn et al., 2007; Kelly et al., 2005; Liu et al., 2015; Sangrigoli & de Schonen, 2004b; Xiao et al., 2014; Xiao, Xiao, Quinn, Anzures, & Lee, 2013). In general, the ORE is characterized by the developmental narrowing of an initial ability to discriminate the faces of all races to a subsequently improved ability to discriminate own-race faces and a diminished ability to discriminate other-race faces.

To date, most of the studies that have investigated the emergence of the ORE have presented either static images (black and white photographs) of silent faces (Anzures et al., 2012; Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly, Liu et al., 2007; Kelly, Quinn et al., 2007; Kelly et al., 2005; Pascalis et al., 2005; Scott & Monesson, 2009) or dynamic but silent faces (Liu et al., 2011; Xiao et al., 2014). The findings from

The work reported here was performed when the authors were at Florida Atlantic University.

these studies have demonstrated that early experience plays a critical role in the tuning of the face processing system and that infants who are exposed mostly to a single-race face category cease discriminating other-race faces by the time they reach the end of the first year of life. These studies have contributed important information on the critical effects of early experience on face processing.

Unfortunately, the developmental picture that extant studies on the emergence of the ORE have painted may be incomplete because the test stimuli that have been used in these studies do not represent infants' daily experiences. Usually, infants are exposed to dynamic vocalizing faces during their daily social interactions (e.g., the peek-a-boo game). Studies have found that dynamic (silent) faces elicit more infant attention (Wilcox & Clayton, 1968) and better discrimination (Otsuka et al., 2009; Spencer, O'Brien, Johnston, & Hill, 2006) than static faces. In addition, studies have found that dynamic vocalizing faces are even more perceptually salient than static or dynamic silent faces. This is because such faces are usually specified by a variety of modality-specific vocal attributes such as pitch and timbre together with facial attributes such as color, shape, and texture as well as various amodal attributes such as intensity, duration, tempo, rhythm, gender, affect, and identity (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009; Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Munhall & Vatikiotis-Bateson, 1998). This multisensory, multi-attribute representation of dynamic vocalizing faces imbues them with multisensory redundancy which is known to increase perceptual salience which, in turn, has been found to enhance perception, learning, and memory (Bahrick & Lickliter, 2012; Grant & Seitz, 2000; Lewkowicz, 1988, 2004; Partan & Marler, 1999; Rosenblum, 2008; Rowe, 1999; Shams & Seitz, 2008; Stein & Stanford, 2008; Sumbly & Pollack, 1954; Summerfield, 1979; Thelen, Matusz, & Murray, 2014). Given this, it may be that older infants who have already undergone perceptual narrowing in the processing of static and/or dynamic/silent faces may still possess the ability to discriminate other-race faces if they are dynamic and vocalizing. Of course, this prediction implies, in part, that older infants can take advantage of the greater salience of redundantly specified faces.

Studies have shown that infants possess some rudimentary multisensory processing abilities at birth and that as they grow and acquire experience, they gradually acquire increasingly better multisensory processing abilities. Thus, for example, at birth infants can perceive intersensory relations based on low-level physical stimulus attributes such as intensity (Lewkowicz & Turkewitz, 1980) and temporal synchrony (Lewkowicz, Leo, & Simion, 2010). By 2 to 4 months of age, infants begin to exhibit the ability to match the auditory and visual attributes of isolated speech syllables (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003) and to detect the temporal alignment of audible and visible syllables (Lewkowicz, 2000b, 2010). By the second half of the first year of life, infants begin to perceive audiovisual spatiotemporal unity (Scheier, Lewkowicz, & Shimojo, 2003), distance (Walker-Andrews & Lennon, 1985), gender (Hillairet de Boisferon et al., 2015; Patterson & Werker, 2002; Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1994; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991), and affect (Walker-Andrews, 1986). Finally, by their first birthday,

infants exhibit the ability to match the auditory and visual attributes of fluent audiovisual speech on the basis of synchrony, prosody, and even the identity of their native language (Lewkowicz, Minar, Tift, & Brandon, 2015; Lewkowicz & Pons, 2013). Finally, as these various multisensory processing abilities are emerging, the ability to selectively and voluntarily attend to sources of multisensory redundancy also emerges. For example, by 8 months of age, infants begin to selectively attend to a talker's mouth (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017; Lewkowicz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowicz, 2015) which helps them discover various properties of complex audiovisual speech. In sum, the emergence of increasingly more sophisticated multisensory processing abilities permits infants to take advantage of increasingly more complex forms of multisensory redundancy (Bahrick & Lickliter, 2012; Lewkowicz, 2000a, 2014; Lewkowicz & Ghazanfar, 2009; Walker-Andrews, 1997).

If the ORE is considered in the context of the documented benefits of multisensory redundancy and of infants' improving ability to detect multisensory redundancy, the time-course of the developmental emergence of the ORE may be different for dynamic vocalizing faces than for static silent and/or dynamic silent faces. So far, it appears that the ORE begins emerging early in infancy and that it becomes a relatively stable feature of perceptual responsiveness by the end of the first year of life. Specifically, studies have found that the ORE begins to affect responsiveness as early as 3 months of age (Kelly et al., 2005; Sangrigoli & de Schonen, 2004b) and that by 9 months of age infants no longer spontaneously discriminate other-race faces (Kelly, Quinn et al., 2007). This decline in the ability to discriminate other-race faces appears to be the direct result of disproportionate early experience with own-race faces (Rennels & Davis, 2008; Sugden, Mohamed-Ali, & Moulson, 2014) which has been suggested to produce an attentional bias for the processing of own-race faces (Markant, Oakes, & Amso, 2016). Studies that have directly manipulated early experience have, in fact, shown that the effects of perceptual narrowing can be delayed or reversed by providing infants with extra experience with non-native faces (Anzures et al., 2012; Pascalis et al., 2005; Sangrigoli & de Schonen, 2004a; Scott & Monesson, 2009). Overall, these types of results suggest that the mechanisms underlying the ORE remain relatively plastic during its initial emergence in infancy. That is, it may be that the developmental trajectory of the ORE first documented by studies presenting static silent or dynamic silent faces mostly reflects responsiveness to these types of stimuli rather than a general diminution of perceptual plasticity. If so, then older infants who have been found to exhibit the ORE in response to static and dynamic silent other-race faces may be capable of discriminating dynamic vocalizing other-race faces.

The purpose of the current study was to investigate this possibility by testing 4-6- and 10-12-month-old Caucasian infants' discrimination of dynamic own- and other-race faces with a habituation/test procedure. Across the habituation and test phases of three experiments, we presented a visible and audible speech syllable being articulated repeatedly (Experiment 1), a visible-only version of the speech syllable (Experiment 2), or a visible version of the speech syllable together with a synchronous non-speech sound (Experiment 3). One group of infants in each experiment and at each age was tested for discrimination

of own-race faces while a second group was tested for discrimination of other-race faces.

The primary aim of this study was to determine whether infants can discriminate own- and other-race faces at each age and, if so, whether successful discrimination requires the presence of concurrent audible speech articulations or whether any concurrent sounds are sufficient for discrimination. Crucially, our experimental design ensured that the auditory context remained constant across the habituation and test trials in each experiment, meaning that either the identical sound was presented across the habituation and test phases (Experiments 1 and 3) or that no sound was presented across them (Experiment 2). This, in turn, ensured that successful discrimination could only be based either on the detection of facial features in the context of auditory cues or in the absence of auditory cues and that it could not be based on auditory cues alone. Finally, a secondary aim of this study was to obtain concurrent measures of selective attention to different parts of the face to determine where infants deployed their selective attention during the learning phase. For this, we employed an eye-tracking device to record point of gaze during the habituation phase.

2 | EXPERIMENT 1

2.1 | Method

2.1.1 | Participants

We tested two age groups in this experiment. One was a group of 4–6-month-old infants ($n = 28$; 15 boys; $M_{\text{age}} = 20.91$ weeks, range = 16.00–27.57 weeks) and the other was a group of 10–12-month-old infants ($n = 27$; 12 boys; $M_{\text{age}} = 47.49$ weeks, range = 42.14–53.43 weeks). All infants in this experiment, as well as those in the subsequent experiments, were full-term at birth, had a birth weight of 2500 grams or higher, and had a 5-minute APGAR score of 7 or higher. All infants were healthy at the time of testing, had no history of recent eye or ear infection, and came from Caucasian households. We tested an additional 13 infants but did not include their data because they were fussy (4), inattentive or the parent interfered with them (3), or they had a mixed racial background (6).

2.1.2 | Apparatus and stimuli

Three different types of videos comprised the stimuli for this experiment. One video showed a short segment of a Winnie-the-Pooh cartoon, which was presented at 70–74 dB (A-scale). The second video was an attention-getter and consisted of a silently and continuously expanding/contracting green disk. The final video was that of a face uttering the speech syllable /a/.

The face stimuli were created by making video recordings of several different female actors who considered themselves as either Caucasian or Asian (Chinese, Japanese, or Vietnamese). Each actor was filmed while she repeatedly uttered a speech syllable. We then used these video recordings to create four pairs of videos. Two of these video pairs consisted of two different Caucasian faces while the other two video

pairs consisted of two different Asian faces. All eight faces were similar in terms of their attractiveness/distinctiveness as determined by ratings obtained from 30 adult raters using a 7-point attractiveness rating scale and a rank-ordered scale of distinctiveness (see Kelly, Quinn et al., 2007, and Newell, Chiroro, & Valentine, 1999, for a similar approach). The stimulus pairs used in this experiment can be seen in Figure 1. As Figure 1 shows, each video depicted a woman's face (approximately from mid neck to the top of her head) looking directly at the camera without blinking. During the video, the woman uttered the syllable /a/ (presented at 63–67 dB, A-scale) every 3 s for a maximum of 60 s. Each actor's mouth articulation lasted between 0.94 and 1.28 s while she produced a vocalization lasting between 0.61 and 0.73 s. A separate female, whose face was not recorded, contributed the audio track that replaced the original vocalizations of each video. The duration of this woman's vocalization was 0.65 s. This procedure was adapted from previous studies (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999) and ensured that no idiosyncratic cues linked the voice to a particular face. The face display area measured 22.25×36.83 cm on the computer monitor and the faces themselves spanned roughly $\frac{1}{3}$ of the area (22.25×13.3 cm subtending approximately 22 degrees of visual angle). As can be seen in Figure 1, all actors had their hair pulled back in a ponytail and wore a headband at the time of recording so that their hair would not be visible.

We used two separate computer systems to run the experiment. The main one was a Window-based PC computer and the second one was a Tobii eye-tracking computer (Model T60 XL, 60 Hz sampling rate). We used the PC to run the experiment by using a custom-written program designed to implement the habituation/test procedure. This computer ran the experiment by controlling the presentation of the stimuli and recording infant looking via mouse presses performed by an experimenter. We used the Tobii eye-tracker to separately record

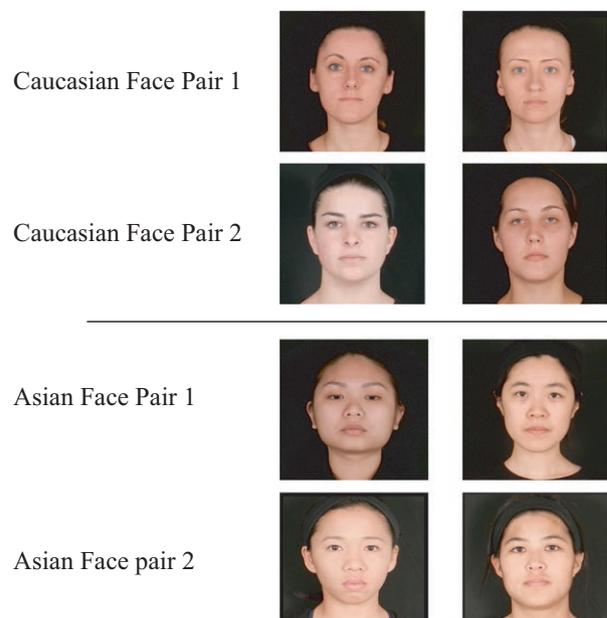


FIGURE 1 The face pairs presented in Experiments 1, 2 and 3 [The author(s) have obtained the individual's or parent's/guardian's free prior informed consent to publish this image.]



FIGURE 2 An example of a Caucasian and an Asian face with the eye, nose, and mouth AOIs overlaid [The author(s) have obtained the individual's or parent's/guardian's free prior informed consent to publish this image.]

infant point-of-gaze to measure selective attention. We also used the Tobii's 24-inch (60.9 cm) monitor and a speaker located along the bottom edge of the monitor to present the stimuli.

2.1.3 | Procedure

Testing took place in a dimly lit, sound-attenuated booth. Each infant sat in an infant seat or on his/her caregiver's lap approximately 50 cm from the stimulus-presentation monitor. If the infant was seated on the caregiver's lap, the caregiver wore headphones that played white noise and sunglasses that were occluded with tape to prevent the parent from hearing and seeing the videos and also to prevent the eye-tracker from accidentally recording the parent's eye gaze. Prior to the test session, infants were calibrated to a 5-point display presenting a bouncing object in each of the four corners and the center of the monitor. Successful calibration was achieved if infants calibrated to at least four of the five points (96% of all infants). The experimenter was located outside the booth and thus could not see nor hear the stimuli being presented. The experimenter observed the infants via a closed-circuit camera and monitor and, based on infant looks at the stimulus-presentation monitor, controlled the onset and offset of the stimuli via clicks of the mouse attached to the PC computer running the habituation/test program.

The habituation/test procedure used in this experiment was based on a 1 s look-away criterion. This meant that the stimulus was presented for as long as infants were looking at the stimulus-presentation monitor and until they looked away from it for more than 1 s or until 60 s elapsed. Once they met this criterion, the trial ended and the attention-getter reappeared to reorient the infants back to the stimulus-presentation monitor. The experiment began with a pre-test trial during which infants could see and hear a segment of a Winnie-the-Pooh video for a maximum of 60 s. The purpose of this trial was to assess the infant's initial level of attention. As soon as this trial ended, either because the infant reached the maximum trial length or met the look-away criterion, the habituation phase began. During this phase, infants could see and hear the actor repeatedly uttering the syllable /a/. Habituation trials continued until the total amount of looking during the three most recent trials decreased to less than 50% of the total amount of looking during the first three trials.

As soon as an infant reached the habituation criterion, the next trial constituted the start of the test phase. This phase consisted of four trials of alternating Familiar and Novel test trials, with the constraint that the Familiar test trial always be the first test trial. During the Familiar test trials we presented the same stimulus that was presented during the habituation phase whereas during the Novel test trials we presented a novel face that belonged to the same race category

as the face presented during the habituation phase. During the Novel test trials, the novel face could be seen and heard uttering the identical syllable that was presented during the habituation phase. The experiment ended with a Post-test trial during which we presented the Winnie-the-Pooh movie again. Responsiveness in this trial was used to measure the terminal level of attention. It was also used to help determine whether a failure to exhibit response recovery in the Novel test trials was due to a true failure to discriminate between the Familiar and Novel test trials. Given that the stimulus presented in the Post-test trial was so different from the other stimuli, it was expected that infants would exhibit response recovery in this test trial even if they failed to exhibit response recovery in the Novel test trials.

At each age, we formed two groups: an own-race group and an other-race group and counterbalanced face pair across the infants within each group. Also, we counterbalanced the faces used as the habituation and test stimuli for each face pair across the infants in each group.

To collect the point-of-gaze data, we defined four areas-of-interest (AOI): the whole face, eyes, nose, and mouth. The eye AOI was delineated by one horizontal line above the eyebrows and another horizontal line through the top of the nasal bridge along with two vertical lines at the edge of the actor's hairline on both sides of the face. The nose AOI was delineated by one horizontal line at the top of the actor's nasal bridge, just below the eye AOI, and another horizontal line just under the actor's nostrils along with two vertical lines just outside the actor's nostrils. Finally, the mouth AOI was delineated by one horizontal line just below the nose AOI, and another horizontal line running through the center of the chin along with two vertical lines located halfway between the right and left corners of the mouth. The eye and mouth AOIs corresponded to the AOIs defined by Lewkowicz and Hansen-Tift (2012) and the nose AOI corresponded with the AOI defined by Xiao et al. (2013). Figure 2 provides an example of one Caucasian and one Asian face overlaid with these AOIs.

2.2 | Results and discussion

First, we conducted a preliminary analysis to determine whether any infants exhibited spontaneous regression to the mean during the first test trial (i.e., once they reached the habituation criterion). Specifically, we examined the duration of looking during the first Familiar test trial separately for each age group. Any infant whose duration of looking in this trial exceeded the mean duration of looking in this trial for the whole group by two standard deviations was excluded from further analyses. Based on this preliminary analysis, we excluded the data of four 4–6-month-olds and two 10–12-month-olds from any further analyses.

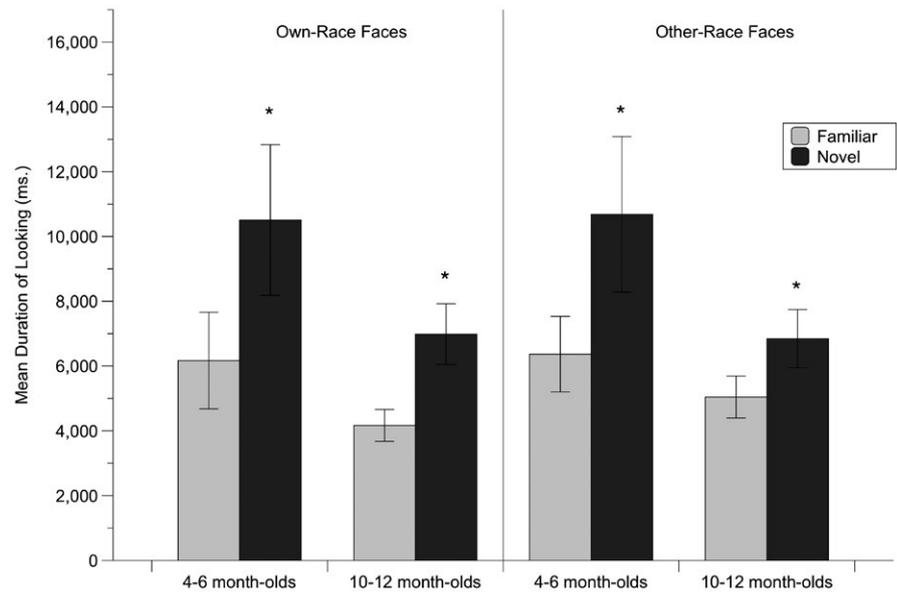


FIGURE 3 Mean duration of looking at Familiar and Novel faces as a function of age and race of face in Experiment 1. Error bars represent SEMs and asterisks represent statistically significant results

To test our principal *a priori* hypothesis that both age groups would exhibit response recovery when presented with novel faces, we first collapsed the duration of looking scores for the two Familiar and the two Novel test trials to yield a single mean duration of looking score for each respective type of test trial. Figure 3 shows the mean duration of looking at the Familiar and Novel faces as a function of age, separately for own-race and other-race faces. To test our hypothesis that infants would exhibit response recovery when presented with novel faces of either race, we used planned, paired-samples, *t* tests (one-tailed) to compare responsiveness across the Familiar and Novel test trials at each age, respectively. These tests indicated that the 4–6-month-old infants exhibited significant response recovery when presented with novel Caucasian faces, $t(12) = 3.18$, $p = .004$, Cohen's $d = 0.51$, as well as novel Asian faces, $t(12) = 2.01$, $p = .034$, Cohen's $d = 0.59$. Similarly, the *t* tests indicated that the 10–12-month-old infants exhibited response recovery when presented with novel Caucasian faces, $t(13) = 3.81$, $p = .001$, Cohen's $d = 1.04$, as well as novel Asian faces, $t(10) = 2.57$, $p = .014$, Cohen's $d = 0.90$.

In sum, the results from this experiment indicate that infants of both ages successfully discriminated own- and other-race dynamic faces which were accompanied by a redundant audible speech utterance. Critically, the findings demonstrate that 10–12-month-old infants can discriminate other-race faces when they can be seen and heard uttering a speech sound. This finding contrasts with previously reported findings that infants no longer discriminate other-race faces by this age when they are tested with static or dynamic and silent faces (Bar-Haim et al., 2006; Hayden, Bhatt, Reed, Corbly, & Joseph, 2007; Kelly, Quinn et al., 2007; Kelly et al., 2005; Sangrigoli & de Schonen, 2004b). Our findings suggest that the ORE obtained in previous studies probably reflects infants' response to static and/or silent dynamic faces.

It is important to note that the audible speech utterances that accompanied the visible speech articulations could not have served as a discriminative cue in this experiment because the utterances were identical across the Familiar and Novel test trials. Therefore, the most

reasonable interpretation of the successful discrimination obtained here is that it was based on the dynamic, multisensory character of the faces. Specifically, the concurrent and redundant speech utterances probably increased the perceptual salience of the face and this probably increased the infants' attentional focus to the point that it made it possible for them to engage in deeper processing of facial feature information. The next experiment tested this possibility.

3 | EXPERIMENT 2

If the successful face discrimination obtained in Experiment 1 depends on the accompanying and redundant speech utterances, then infants should fail to discriminate silent faces even if they can be seen articulating speech. If, however, dynamic facial cues are sufficient for discrimination, then the infants should discriminate the faces given that face-motion cues facilitate infants' discrimination of faces (Otsuka et al., 2009; Spencer et al., 2006; Wilcox & Clayton, 1968). Thus, we repeated Experiment 1 except that this time we presented the faces in silence. We hypothesized that the younger infants may discriminate both types of faces because of their broad perceptual tuning for all face categories. In contrast, we hypothesized that the older infants may not discriminate other-race faces because they may require that other-race faces be redundantly specified by concurrent visual and auditory attributes to overcome the effects of perceptual narrowing.

3.1 | Method

3.1.1 | Participants

We tested two groups of infants in this experiment. One was a group of 4–6-month-old infants ($n = 24$, 15 boys; $M_{\text{age}} = 20.20$ weeks, range = 16.14–27.57 weeks) and the other was a group of 10–12-month-old

infants ($n = 24$, 12 boys; $M_{\text{age}} = 48.04$ weeks, range = 42.00–53.42 weeks). In addition, we tested four other infants but excluded them from data analysis due to fussiness (2), mixed racial background (1), or a recent eye infection (1).

3.1.2 | Apparatus and stimuli

Here, the apparatus and stimuli were identical to those used in Experiment 1 except that we presented silent versions of the habituation and test videos.

3.1.3 | Procedure

The procedure was identical to that used in Experiment 1.

3.2 | Results and discussion

The preliminary analysis indicated that one 4–6-month-old infant and one 10–12-month-old infant exhibited spontaneous regression to the mean. The data from these two infants were removed from all subsequent analyses. Figure 4 shows the data from this experiment. As before, we collapsed the duration of looking scores for the two Familiar and the two Novel test trials to derive a single mean duration of looking score for each respective type of test trial. We then compared the data from the Familiar and Novel test trials, separately at each age, with planned, paired-samples (one-tailed), t tests. Results indicated that the 4–6-month-old infants did not exhibit response recovery to either the novel Caucasian face, $t(10) = 0.22$, $p = .414$, Cohen's $d = 0.06$, or to the novel Asian face, $t(11) = 0.08$, $p = .471$, Cohen's $d = 0.03$. In contrast, results indicated that the 10–12-month-old infants exhibited response recovery to the novel Caucasian face, $t(11) = 2.09$, $p = .030$, Cohen's $d = 0.53$, but not to the novel Asian face, $t(10) = 0.57$, $p = .290$, Cohen's $d = 0.17$.

We conducted two additional secondary analyses to gain insights into the failures to discriminate. First, we wanted to determine

whether the 4–6-month-old infants' failure to discriminate both types of faces reflected a true failure to discriminate the stimuli presented in the Familiar and Novel test trials or whether it reflected fatigue. Second, we wanted to determine whether the stimuli in this experiment elicited less attention than those in Experiment 1. To determine if the infants were fatigued, we compared responsiveness in the Familiar test trials versus responsiveness in the Post-test trial by way of a paired (one-tailed) t test. Results indicated that the 4–6-month-old infants exhibited significant response recovery in the Post-test trial, $t(22) = 21.86$, $p < .001$, Cohen's $d = 5.77$, indicating that their failure to discriminate the faces was a true failure to discriminate rather than fatigue. To determine whether the infants in this experiment paid less attention, we compared attention across the two experiments at both ages, by separately examining three measures: the total looking time during the habituation phase, the number of habituation trials, and the average looking time per habituation trial (calculated by dividing the total looking time during habituation by the number of habituation trials for each infant). We analyzed each of these measures with separate ANOVAs, with Age and Experiment as between-subjects factors. The analyses indicated no significant main effects for the three measures: total looking during habituation, $F(1, 91) = 1.05$, $p = .308$, $\eta_p^2 = .02$, number of habituation trials, $F(1, 91) < .01$, $p = .961$, $\eta_p^2 < .02$, and average looking time per habituation trial, $F(1, 91) = .51$, $p = .475$, $\eta_p^2 < .01$. Furthermore, there were no significant Age \times Experiment interactions for the three measures: total looking during habituation, $F(1, 91) = 0.34$, $p = .56$, $\eta_p^2 < .01$, number of habituation trials, $F(1, 91) = 1.86$, $p = .176$, $\eta_p^2 = .02$, and average looking time per habituation trial, $F(1, 91) = 0.34$, $p = .56$, $\eta_p^2 < .01$. These results indicate that infants devoted similar amounts of attention, both, across the two experiments and across the two ages.

The findings from this experiment indicate that dynamic visual cues alone are not sufficient to elicit discrimination in the younger infants nor are they sufficient to elicit discrimination of other-race

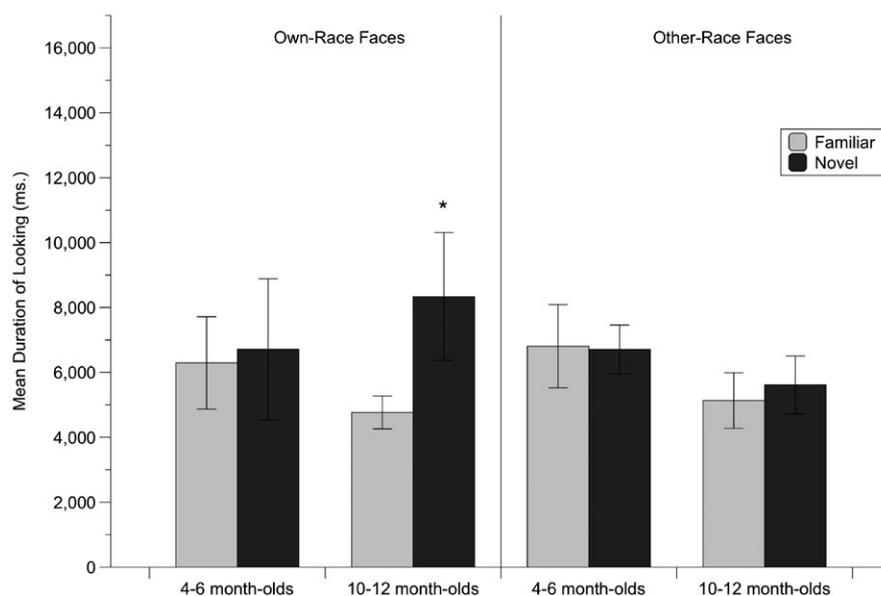


FIGURE 4 Mean duration of looking at Familiar and Novel faces as a function of age and race of face in Experiment 2. Error bars represent SEMs and the asterisks represent statistically significant results

faces in the older infants. Furthermore, when the results from the principal analysis are combined with the results from the secondary analysis it is clear that the difference in outcome across Experiments 1 and 2 reflects the specific nature of the faces. That is, when the faces were dynamic and accompanied by a redundant audible speech utterance (Experiment 1), they were sufficiently distinct to be discriminable for both age groups. When, however, the faces were dynamic but silent (Experiment 2), neither the own-race nor the other-race faces were sufficiently distinct to be discriminable for the 4–6-month-old infants and only the own-race faces were sufficiently distinct to be discriminable for the 10–12-month-old infants.

4 | EXPERIMENT 3

The question of primary interest in the current study was whether the ORE, as currently understood in the extant literature, reflects infant responsiveness to static and/or silent dynamic faces. The data from Experiments 1 and 2 showed that infants who typically exhibit the ORE, namely 10–12-month-old infants, can discriminate dynamic other-race faces as long as they are accompanied by a redundant speech utterance. If so, what specific form must the redundancy take for the older infants to overcome the ORE? Must the redundancy be represented by equivalent speech information across the modalities or can it be a more general form of redundancy that is specified by facial speech cues together with temporally synchronized sounds? To answer this question, we repeated Experiment 1 except that this time we presented the faces together with a temporally synchronized non-speech sound.

4.1 | Method

4.1.1 | Participants

We tested two groups of infants in this experiment. One was a group of 4–6-month-old infants ($n = 25$, 15 boys; $M_{\text{age}} = 20.20$ weeks, range = 16.14–27.57 weeks) and the other was a group of 10–12-month-old infants ($n = 24$, 15 boys; $M_{\text{age}} = 49.28$ weeks, range = 42.42–53.28 weeks). We tested 10 additional infants but excluded them from data analysis due to fussiness ($n = 3$), mixed racial background ($n = 6$), or a recent health concern ($n = 1$).

4.1.2 | Apparatus and stimuli

We used the same apparatus that we used in Experiment 1 and presented the same habituation and test videos as those in Experiment 1. Here, however, the faces were accompanied by a computer-generated “boing” sound (65 dB, A scale; created with Adobe Audition CS6 software). The duration of this sound was the same as the duration of the vocal utterance used in Experiment 1 (0.65 s) and its onset and offset were synchronized with the onset and offset of the visible speech articulation

4.1.3 | Procedure

The procedure was identical to that used in Experiment 1.

4.2 | Results and discussion

The preliminary analysis indicated that three 4–6-month-old infants and one 10–12-month-old infant exhibited regression to the mean. The data from these four infants were removed from all subsequent analyses.

The data from this experiment can be seen in Figure 5. As in the prior experiments, we compared the combined scores from the Familiar and Novel test trials with planned, paired-samples t tests. These tests indicated that the 4–6-month-old infants discriminated the Caucasian faces, $t(10) = 2.29$, $p = .02$, Cohen's $d = 0.75$, and the Asian faces, $t(10) = 1.85$, $p = .047$, Cohen's $d = 0.75$. In contrast, the paired-samples t tests indicated that the 10–12-month-old infants discriminated the Caucasian faces, $t(11) = 3.00$, $p = .006$, Cohen's $d = 0.32$, but that they did not discriminate the Asian faces, $t(10) = 1.13$, $p = .144$, Cohen's $d = 0.42$.

The data from this experiment indicated that younger infants discriminated both own-race and other-race dynamic faces even though they were accompanied by synchronous non-speech sounds but that the older infants only discriminated own-race dynamic faces when they were accompanied by non-speech sounds. In other words, the data from the older infants suggest that they require that other-race faces be accompanied by redundant speech attributes, rather than just sounds, to be able to successfully discriminate other-race faces.

4.2.1 | Selective attention to different areas of the face in Experiments 1, 2, and 3

To determine where infants deployed their attention during learning, we examined point-of-gaze during the first three habituation trials. Not all infants yielded usable point-of-gaze data even though they may have yielded usable data from the habituation/test procedure (see previous habituation analysis). Here, infants were excluded if they met one of three criteria: (1) they exhibited atypical levels of fixation compared to other infants of the same age group and experimental condition, (2) the eye-tracker failed to collect a minimum of 2 seconds of looking over the whole experiment, or (3) they failed to calibrate to at least four of the five calibration points prior to starting the experiment. To implement the first exclusion criterion, we examined infants' raw captured looking times using the boxplot method of outlier detection as described by Cohen, Cohen, West, and Aiken (2002). Overall, we eliminated the data from three 4–6-month-old infants and one 10–12-month-old infant based on the first criterion, four 4–6-month-olds and one 10–12-month-old infant based on the second criterion, and four 4–6-month-olds and two 10–12-month-olds based on the third criterion. One infant's data were lost due to equipment error. Thus, the three experiments yielded usable point-of-gaze data from 57 infants in the 4–6-month-old group ($n = 24$ in Experiment 1, $n = 19$ in Experiment 2, and $n = 14$ in Experiment 3) and

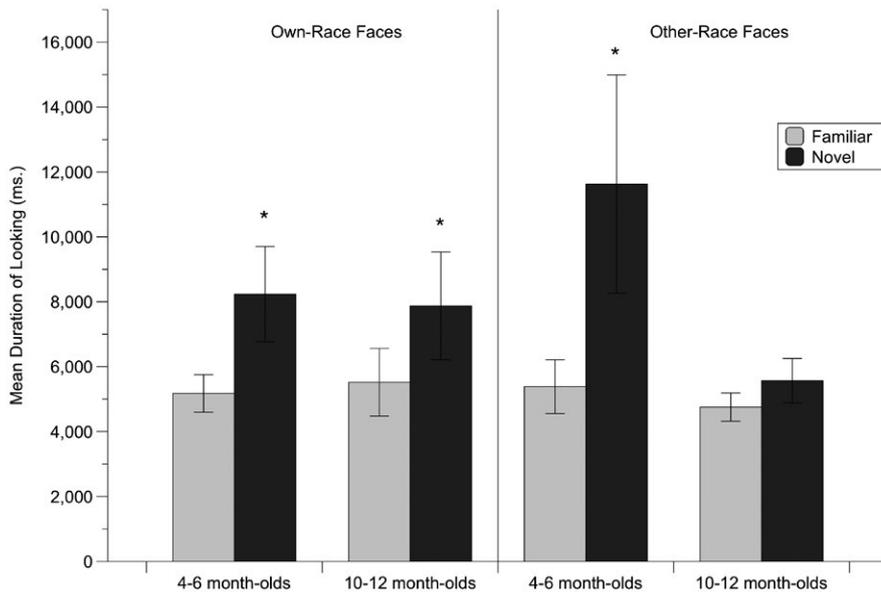


FIGURE 5 Mean duration of looking at Familiar and Novel faces as a function of age and race of face in Experiment 3. Error bars represent SEMs and the asterisks represent statistically significant results

67 infants in the 10–12-month-old group ($n = 23$ in Experiment 1, $n = 22$ in Experiment 2, and $n = 22$ in Experiment 3).

We calculated each infant's proportion-of-total-looking time (PTLT) scores for the eyes, nose, and mouth AOI by dividing the amount of total looking to each of these AOIs by the total amount of looking to the face AOI. We then entered the PTLT scores into a mixed, repeated-measures ANOVA, with AOI (eyes, nose, and mouth) as the within-subjects factor and Age (4–6 and 10–12 months of age), Face-Race (Caucasian or Asian), and Experiment (1, 2, or 3) as between-subjects factors. The ANOVA yielded a significant main effect of AOI, $F(2, 111) = 19.33, p < .001, \eta_p^2 = .26$, a significant main effect of Age, $F(1, 112) = 4.48, p = .036, \eta_p^2 = .04$, a significant AOI \times Age interaction, $F(2, 111) = 14.70, p < .001, \eta_p^2 = .21$, and a significant Age \times Face-Race interaction, $F(1, 112) = 5.65, p = .019, \eta_p^2 = .05$. There were no other significant effects. Figure 6 depicts the mean PTLT scores as a function of age and AOI. As can be seen, the younger infants looked equally at the three AOIs whereas the older infants looked

longer at the mouth AOI as opposed to the other two. Paired-samples t tests confirmed this by showing that looking to the eyes and mouth did not differ in the 4–6-month-olds, $t(56) = 0.37, p = .714$, Cohen's $d = 0.08$, but that looking to the mouth was greater than to the eyes in the 10–12-month-olds, $t(66) = 9.55, p < .001$, Cohen's $d = 2.02$. The pattern was the same when the AOI \times Race interaction was examined. These results indicate that, regardless of the actor's race and whether it was accompanied by a speech syllable or a non-speech sound, the 4–6-month-old infants distributed their attention equally to the three regions of the face during the initial learning phase but that the 10–12-month-old infants allocated most of their attention to the mouth.

5 | GENERAL DISCUSSION

Previous studies of infant response to other-race faces have found that the ORE is fully established by the end of the first year of life. Here,

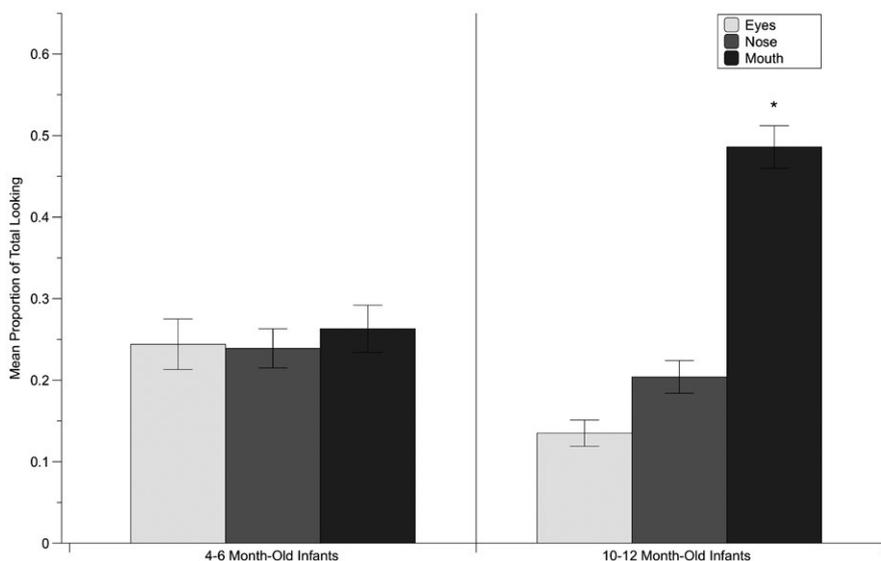


FIGURE 6 The proportion of total looking time (PTLT) that infants spent looking at the eyes, nose, and mouth in the two age groups, collapsed across the three experiments. Error bars represent SEMs and the asterisks represent statistically significant results

we investigated whether the developmental timing of the emergence of the ORE reported to date may reflect a decline in responsiveness to static and/or dynamic silent faces rather than a general, experience-based, decline in perceptual sensitivity to an infrequently experienced face category. Therefore, we hypothesized that infants may not exhibit a response decline to other-race faces when tested with dynamic vocalizing faces because these are the sorts of faces that infants normally experience in their everyday social environment and because such faces are perceptually more salient. Thus, we habituated 4-6- and 10-12-month-old Caucasian infants either to an own-race or an other-race face that they either saw and heard uttering an /a/ speech syllable (Experiment 1), only saw it uttering the /a/ syllable (Experiment 2), or saw it uttering the /a/ syllable and heard a synchronous non-speech sound (Experiment 3). During the test trials, infants saw a novel person's face from the same race category and in Experiments 1 and 3 they also heard an audible stimulus that was the identical audible stimulus that they heard during habituation. This ensured that the only discriminative cue in Experiments 1 and 3 was facial identity.

Findings indicated that when infants saw and heard a face uttering a speech syllable, both the 4-6-month-olds and 10-12-month-olds discriminated own-race and other-race faces. When the infants saw a face silently uttering a speech syllable, the 4-6-month-olds did not discriminate either type of face while the 10-12-month-olds only discriminated own-race faces. Finally, when the infants saw a face uttering a speech syllable while they heard a synchronous non-speech sound, the 4-6-month-olds discriminated both types of faces whereas the 10-12-month-olds once again only discriminated own-race faces. Overall, the fact that the older infants discriminated other-race faces when they were accompanied by a speech sound confirmed our hypothesis that the ORE so far reported in extant studies reflects responsiveness to static/silent and dynamic/silent stimuli and not the closing of a sensitive period for face processing.

Overall, our findings also provide interesting insights into the development of the processing of multisensory dynamic faces. We found that 4-6-month-old infants discriminated dynamic own-race and other-race faces when they were accompanied by a sound and other studies have found that infants also can discriminate static/silent other-race faces (Anzures et al., 2012; Bar-Haim et al., 2006; Kelly, Liu et al., 2007; Kelly, Quinn et al., 2007; Kelly et al., 2005; Pascalis et al., 2005; Scott & Monesson, 2009) as well as dynamic/silent other-race faces (Liu et al., 2011; Xiao et al., 2014). Considered together, these data suggest that dynamic facial cues *per se* may not be essential for face discrimination. We also found, however, that the 4-6-month-olds in our study did not discriminate silent but dynamic own- and other-race faces, indicating that when these infants see a face articulating a speech syllable, they can only discriminate it when the face is accompanied by an auditory stimulus. Thus, these findings suggest that a concurrent auditory stimulus enhances the processing of facial feature information in the younger infants and that, at this age, infants expect to hear a sound when they see a face with a moving mouth. This non-specific expectation is consistent with findings that infants of this age have not yet developed sufficient native-language expertise (Lewkowicz, 2014; Lewkowicz & Ghazanfar, 2009; Maurer & Werker, 2014) and that they have not yet

learned an association between particular types of faces and the particular types of sounds that they make (Lewkowicz & Ghazanfar, 2006; Lewkowicz et al., 2010; Lewkowicz, Sowinski, & Place, 2008).

The non-specific nature of the 4-6-month-old infants' expectations about dynamic faces articulating a speech sound appears at first glance to be inconsistent with the findings from prior studies reporting that infants of this age have specific expectations about the source of sounds. For example, one study found that 5-month-old infants associate human faces with human vocalizations and monkey faces with monkey vocalizations (Vouloumanos, Druhen, Hauser, & Huizink, 2009). It should be noted, however, that the faces presented in this particular study were static and, thus, that infants did not have to map dynamic visible and audible attributes. Other studies, which did not test explicitly for expectations but rather for infants' ability to perceive intersensory equivalence have found that young infants can perceive the equivalence of dynamic visible and audible speech syllables (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003) but not the equivalence of dynamic visible speech syllables with audible, non-speech sounds (Kuhl, Williams, & Meltzoff, 1991). Together, these findings suggest that young infants possess experience-dependent multisensory unity expectations for static faces and vocalizations as well as for dynamic human faces and speech utterances. Of course, this interpretation is not consistent with our finding that the 4-6-month-old infants did not exhibit evidence of speech-specific multisensory unity expectations. Therefore, the most reasonable conclusion is that task demands determine whether infants exhibit a multisensory unity expectation or not. In our study, the infants' task was to detect the facial features that differentiated dynamic faces and to discriminate them in the context of accompanying but unchanging auditory information. Thus, the task was to focus on facial features and not on the association between the facial features and accompanying information. In contrast, in the studies described above, the task required infants to remember previously acquired intersensory associations or to extract equivalent dynamic auditory and visual stimulus features and then map them onto one another. Such a task-based interpretation of these various findings is consistent with the results of many studies of multisensory processing indicating that responsiveness to multisensory inputs depends on task requirements (Murray, Lewkowicz, Amedi, & Wallace, 2016).

The multisensory unity expectation implications of the current results are especially interesting in light of evidence from adults that such expectations play an important role in multisensory responsiveness (Barenholtz, Lewkowicz, Davidson, & Mavica, 2014; Welch, 1999; Welch & Warren, 1980). Some of the best-known examples of such expectations are the McGurk illusion (McGurk & MacDonald, 1976) and the ventriloquism illusion (Bertelson & Radeau, 1981). Both illusions illustrate the strong tendency of adults to unify conflicting auditory and visual inputs into unitary percepts because of a lifetime of experience with redundant/congruent multisensory inputs which normally do not induce illusory percepts. When might multisensory unity expectations begin emerging in development? Our findings suggest that a multisensory unity expectation for faces articulating speech sounds emerges by 12-14 months of age. This is when infants only discriminate other-race

faces that are associated with a speech syllable. This is interesting because the developmental emergence of this specific expectation by this age is consistent with the fact that this is when infants first begin exhibiting audiovisual speech- and language-processing expertise (Lewkowicz & Ghazanfar, 2009; Maurer & Werker, 2014). Crucially, however, it appears that the timing of the emergence of specific multisensory unity expectations depends on the domain of processing. For example, when infants have to process the spatiotemporal relations inherent in an object-based ambiguous audiovisual event, they exhibit evidence of a multisensory unity expectation as early as 6 months of age. That is, when infants see two objects passing through each other while they hear a sound at the point of their spatiotemporal coincidence, they respond as if the objects bounce against each other at 6 and 8 months of age but not at 4 months of age (Scheier et al., 2003). This “bounce” illusion demonstrates that by 6 months of age infants resolve a spatiotemporally conflicting event by unifying its auditory and visual attributes.

The 10–12-month-old infants’ failure to discriminate other-race faces associated with a non-speech sound suggests that a multisensory unity expectation, together with the effects of perceptual narrowing, renders the discrimination of other-race faces more challenging. That is, an accompanying and arbitrary non-speech sound may force 12–14-month-old infants to devote greater attentional resources to resolving the categorical incongruence introduced by the violation of the multisensory unity expectation and may prevent them from focusing on facial feature differences. This conclusion is supported by the results from Experiment 3 showing that when the faces were own-race and, thus, presumably easier to process because of their familiarity, the older infants discriminated them even though their visible articulations did not match the auditory stimulus. It seems that older infants’ ability to discriminate dynamic faces uttering a speech syllable is modulated by their early experience and that when the faces are relatively unfamiliar, they rely on categorically redundant (i.e., matching) speech information for discriminating. They do so even when the speech information cannot be utilized as a discriminative cue. Presumably, the categorically congruent multisensory information increases the overall perceptual salience of the facial feature information and enables the infants to focus on the facial features sufficiently to enable them to detect differences based on those features.

The eye-gaze data provide interesting insights into the mechanisms underlying the emergence of specific multisensory unity expectations. We found that whereas the younger infants deployed equal amounts of their selective attention to the eyes, nose, and mouth of a face visibly uttering a speech syllable, the older infants deployed most of their selective attention to the mouth. Obviously, distributing one’s attention to all regions of a face uttering speech makes it difficult to detect, extract, and learn the relations between specific aspects of visible and audible attributes of the face, much less of the audiovisual speech that it produces. On this account, selectively deploying one’s attention to the mouth—the source of audiovisual speech—promotes the discovery of the links between the visible and audible attributes of the face, including the visible and audible attributes of speech. Interestingly, the shifting attentional strategy that we found across the two age groups is consistent with other findings showing that infants shift their selective attention from the eyes to the mouth during the first year of life

(Hillairet de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). It should be noted, however, that this pattern of developmental shifting of attention reflects infants’ response to fluent audiovisual speech rather isolated speech syllables. Thus, even though this pattern of shifting attention across early development is sure to facilitate the acquisition of increasingly more specific multisensory unity expectations, there are likely to be differences in responsiveness to isolated syllables as opposed to fluent speech. Nonetheless, there is evidence that motion of any sort in the mouth region elicits different patterns of attention from the absence of motion. This is evident in studies examining the emergence of the other-race effect in Chinese infants which found that infants attend increasingly more to the nose of both own- and other-race static faces as development progressed (Liu et al., 2015). The fact our older infants exhibited greater looking to the mouth indicates that motion plays an important role in infant deployment of selective attention to faces.

Of course, the key finding in the current study is that the 10–12-month-old infants discriminated other-race faces as long as they were accompanied by a redundant speech cue. This supports our hypothesis that the specific nature of the stimuli used to test for the emergence of the ORE determines how older infants respond to different face categories. Our results help adjudicate the question of whether the previously reported findings of the ORE emerging by the end of the first year of life reflect the general effects of restricted early experience with own-race as opposed to other-race faces or whether they reflect the specific types of stimuli used in prior studies. Our finding that 10–12-month-old infants discriminated other-race faces as long as those faces were specified by redundant identity cues related to the speech syllable *per se* rather than merely temporal synchrony cues is also important because the distinction between redundant identity versus redundant synchrony cues is crucial. The importance of this distinction is illustrated by findings that newborns can perceive the multisensory redundancy of faces and voices when redundancy is defined by temporal synchrony cues and that they perceive it without processing redundantly specified identity cues (Lewkowicz et al., 2010). It is further illustrated by findings that 12–14-month-old infants can perceive the much more complex multisensory redundancy defined by fluent speech prosody cues and, most importantly, that they can do this even in the absence of synchrony cues (Lewkowicz et al., 2015).

In conclusion, the fact that 10–12-month-old infants can discriminate dynamic other-race faces producing a speech syllable indicates that the ORE is not fully established by the end of the first year of life. Although this is contrary to the results from prior studies reporting that the ORE emerges by this time (Kelly et al., 2009; Kelly, Quinn et al., 2007; Kelly et al., 2005; Liu et al., 2015; Sangrigoli & de Schonen, 2004b; Xiao et al., 2014; Xiao et al., 2013), it is consistent with the results from “training”, adoption, and selective attention studies. The training studies have found that infants who are given extra experience with less frequently experienced categories of information maintain their perceptual sensitivity to those categories relative to infants who do not receive such training (Pascalis et al., 2005; Scott & Monesson, 2009). Studies of Korean adults who were adopted as children by Caucasian

families in France also have shown that they, like their Caucasian counterparts, have difficulty discriminating Korean faces (Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005). Finally, infants whose attentional focus to own-race and other-race faces is experimentally manipulated exhibit recognition of other-race faces (Markant et al., 2016). Together, these findings show that the perceptual system is plastic and that it remains open to the effects of experience well into childhood. Our findings provide additional evidence to support this conclusion. It is too early to tell whether the process of perceptual narrowing reflects a single sensitive period or multiple ones that depend on domain and/or sensory modality (Maurer & Werker, 2014). Nonetheless, it is becoming clear that the sensitive period for a particular category of information depends on specific early experience. Given that infants usually experience talking (i.e., dynamic, audiovisual) faces in their daily social interactions (Fausey, Jayaraman, & Smith, 2016) and that they become better at detecting multisensory redundancy and at profiting from the greater perceptual salience created by such redundancy as they get older (Lewkowicz, 2014), it is not surprising that even older infants can discriminate other-race faces when tested with dynamic other-race faces articulating a speech syllable. This, in turn, means that we must take infants' typical experiences into account if we want to achieve a clear understanding of the interaction between early experience, perceptual narrowing, and the developmental emergence of perceptual expertise.

ORCID

David J. Lewkowicz  <http://orcid.org/0000-0001-6685-3178>

REFERENCES

- Anzures, G., Wheeler, A., Quinn, P.C., Pascalis, O., Slater, A.M., Heron-Delaney, M., & Lee, K. (2012). Brief daily exposures to Asian females reverses perceptual narrowing for Asian faces in Caucasian infants. *Journal of Experimental Child Psychology*, *112*, 484–495.
- Bahrick, L.E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A.J. Bremner, D.J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–206). Oxford: Oxford University Press.
- Barenholtz, E., Lewkowicz, D.J., Davidson, M., & Mavica, L. (2014). Categorical congruence facilitates multisensory associative learning. *Psychonomic Bulletin & Review*, *21*, 1346–1352.
- Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R.M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, *17*, 159–163.
- Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception and Psychophysics*, *29*, 578–584.
- Bothwell, R.K., Brigham, J.C., & Malpass, R.S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin*, *15*, 19–25.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*, e1000436.
- Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Taylor & Francis.
- Fausey, C.M., Jayaraman, S., & Smith, L.B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107.
- Grant, K.W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America*, *108*, 1197–1208.
- Hayden, A., Bhatt, R.S., Reed, A., Corbly, C.R., & Joseph, J.E. (2007). The development of expert face processing: Are infants sensitive to normal differences in second-order relational information? *Journal of Experimental Child Psychology*, *97*, 85–98.
- Hillairet de Boisferon, A., Dupierrix, E., Quinn, P.C., Lævenbruck, H., Lewkowicz, D.J., Lee, K., & Pascalis, O. (2015). Perception of multisensory gender coherence in 6- and 9-month-old infants. *Infancy*, *20*, 661–674.
- Hillairet de Boisferon, A., Tift, A.H., Minar, N.J., & Lewkowicz, D.J. (2017). Selective attention to the source of audiovisual speech in infancy: The role of temporal synchrony and linguistic experience. *Developmental Science*, *23*, <https://doi.org/10.1111/desc.12381>
- Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). Putting the face to the voice: Matching identity across modality. *Current Biology*, *13*, 1709–1714.
- Kelly, D.J., Liu, S., Ge, L., Quinn, P.C., Slater, A.M., Lee, K., & Pascalis, O. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants. *Infancy*, *11*, 87–95.
- Kelly, D.J., Liu, S., Lee, K., Quinn, P.C., Pascalis, O., Slater, A.M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology*, *104*, 105–114.
- Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*, 1084–1089.
- Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Gibson, A., Smith, M., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*, F31–F36.
- Kuhl, P.K., & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.
- Kuhl, P.K., Williams, K.A., & Meltzoff, A.N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 829–840.
- Lewkowicz, D.J. (1988). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*, 155–171.
- Lewkowicz, D.J. (2000a). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281–308.
- Lewkowicz, D.J. (2000b). Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. *Child Development*, *71*, 1241–1257.
- Lewkowicz, D.J. (2004). Perception of serial order in infants. *Developmental Science*, *7*, 175–184.
- Lewkowicz, D.J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, *46*, 66–77.
- Lewkowicz, D.J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology*, *56*, 292–315.
- Lewkowicz, D.J., & Ghazanfar, A.A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy Sciences, USA*, *103*, 6771–6774.
- Lewkowicz, D.J., & Ghazanfar, A.A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, *13*, 470–478.
- Lewkowicz, D.J., & Hansen-Tift, A.M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences, USA*, *109*, 1431–1436.
- Lewkowicz, D.J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match non-human primate faces and voices. *Infancy*, *15*, 46–60.
- Lewkowicz, D.J., Minar, N.J., Tift, A.H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, *130*, 147–162.
- Lewkowicz, D.J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *International Journal of Behavioral Development*, *37*, 90–94.

- Lewkowicz, D.J., Sowers, R., & Place, S. (2008). The decline of cross-species intersensory perception in human infants: Underlying mechanisms and its developmental persistence. *Brain Research*, 1242, 291–302.
- Lewkowicz, D.J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental Psychology*, 16, 597–607.
- Lindsay, D.S., Jack, P.C., & Christian, M.A. (1991). Other-race face perception. *Journal of Applied Psychology*, 76, 587–589.
- Liu, S., Quinn, P.C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same- and other-race faces as revealed by eye tracking in 4- to 9-month-olds. *Journal of Experimental Child Psychology*, 108, 180–189.
- Liu, S., Xiao, W.S., Xiao, N.G., Quinn, P.C., Zhang, Y., Chen, H., & Lee, K. (2015). Development of visual preference for own- versus other-race faces in infancy. *Developmental Psychology*, 51, 500–511.
- Markant, J., Oakes, L.M., & Amso, D. (2016). Visual selective attention biases contribute to the other-race effect among 9-month-old infants. *Developmental Psychobiology*, 58, 355–365.
- Maurer, D., & Werker, J.F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, 56, 154–178.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 229–239.
- Munhall, K.G., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 123–139). Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis.
- Murray, M.M., Lewkowicz, D.J., Amedi, A., & Wallace, M.T. (2016). Multisensory processes: A balancing act across the lifespan. *Trends in Neurosciences*, 39, 567–579.
- Newell, F.N., Chiroro, P., & Valentine, T. (1999). Recognizing unfamiliar faces: The effects of distinctiveness and view. *Quarterly Journal of Experimental Psychology: Section A*, 52, 509–534.
- Otsuka, Y., Konishi, Y., Kanazawa, S., Yamaguchi, M.K., Abdi, H., & O'Toole, A.J. (2009). Recognition of moving and static faces by young infants. *Child Development*, 80, 1259–1271.
- Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, 283, 1272–1273.
- Pascalis, O., Scott, L.S., Kelly, D.J., Shannon, R.W., Nicholson, E., Coleman, M., & Nelson, C.A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences, USA*, 102, 5297–5300.
- Patterson, M.L., & Werker, J.F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22, 237–247.
- Patterson, M.L., & Werker, J.F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, 81, 93–115.
- Patterson, M.L., & Werker, J.F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6, 191–196.
- Pons, F., Bosch, L., & Lewkowicz, D.J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, 26, 490–498.
- Poulin-Dubois, D., Serbin, L.A., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology*, 30, 436–442.
- Rennels, J.L., & Davis, R.E. (2008). Facial experience during the first year. *Infant Behavior and Development*, 31, 665–678.
- Rosenblum, L.D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, 17, 405–409.
- Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, 58, 921–931.
- Sangrigoli, S., & de Schonen, S. (2004a). Effect of visual experience on face processing: A developmental study of inversion and non-native effects. *Developmental Science*, 7, 74–87.
- Sangrigoli, S., & de Schonen, S. (2004b). Recognition of own-race and other-race faces by three-month-old infants. *Journal of Child Psychology and Psychiatry*, 45, 1219–1227.
- Sangrigoli, S., Pallier, C., Argenti, A.M., Ventureyra, V.A.G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, 16, 440–444.
- Scheier, C., Lewkowicz, D.J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, 6, 233–244.
- Scott, L.S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science*, 20, 676–680.
- Shams, L., & Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12, 411–417.
- Spencer, J., O'Brien, J., Johnston, A., & Hill, H. (2006). Infants' discrimination of faces by using biological motion cues. *Perception*, 35, 79–89.
- Stein, B.E., & Stanford, T.R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9, 255–266.
- Sugden, N.A., Mohamed-Ali, M.I., & Moulson, M.C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology*, 56, 249–261.
- Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212–215.
- Summerfield, A.Q. (1979). Use of visual information in phonetic perception. *Phonetica*, 36, 314–331.
- Thelen, A., Matusz, P.J., & Murray, M.M. (2014). Multisensory context portends object memory. *Current Biology*, 24, R734–R735.
- Vouloumanos, A., Druhen, M., Hauser, M., & Huizink, A. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences, USA*, 106, 18867–18872.
- Walker-Andrews, A.S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology*, 22, 373–377.
- Walker-Andrews, A.S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, 121, 437–456.
- Walker-Andrews, A.S., Bahrack, L.E., Raglioni, S.S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, 3, 55–75.
- Walker-Andrews, A.S., & Lennon, E.M. (1985). Auditory-visual perception of changing distance by human infants. *Child Development*, 56, 544–548.
- Welch, R.B. (1999). Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perception. In G. Aschersleben, T. Bachman, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events*, Vol. 129 (pp. 371–387). Amsterdam: Elsevier.
- Welch, R.B., & Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88, 638–667.
- Wilcox, B.M., & Clayton, F.L. (1968). Infant visual fixation on motion pictures of the human face. *Journal of Experimental Child Psychology*, 6, 22–32.
- Xiao, N.G., Perrotta, S., Quinn, P.C., Wang, Z., Sun, Y.-H.P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, 5, 633.
- Xiao, W.S., Xiao, N.G., Quinn, P.C., Anzures, G., & Lee, K. (2013). Development of face scanning for own- and other-race faces in infancy. *International Journal of Behavioral Development*, 37, 100–105.

How to cite this article: Minar NJ, Lewkowicz DJ. Overcoming the other-race effect in infancy with multisensory redundancy: 10–12-month-olds discriminate dynamic other-race faces producing speech. *Dev Sci*. 2017;e12604. <https://doi.org/10.1111/desc.12604>