Review Article

Tamara L. Watson[1,2]
Rachel A. Robbins[1]
Catherine T. Best[2,3]

[1] *School of Social Science and Psychology*
University of Western Sydney
New South Wales, Australia
E-mail: t.watson@uws.edu.au

[2] *MARCS Institute*
University of Western Sydney
New South Wales, Australia

[3] *School of Humanities and*
Communication Arts
University of Western Sydney
New South Wales, Australia

Infant Perceptual Development for Faces and Spoken Words: An Integrated Approach

ABSTRACT: *There are obvious differences between recognizing faces and recognizing spoken words or phonemes that might suggest development of each capability requires different skills. Recognizing faces and perceiving spoken language, however, are in key senses extremely similar endeavors. Both perceptual processes are based on richly variable, yet highly structured input from which the perceiver needs to extract categorically meaningful information. This similarity could be reflected in the perceptual narrowing that occurs within the first year of life in both domains. We take the position that the perceptual and neurocognitive processes by which face and speech recognition develop are based on a set of common principles. One common principle is the importance of systematic variability in the input as a source of information rather than noise. Experience of this variability leads to perceptual tuning to the critical properties that define individual faces or spoken words versus their membership in larger groupings of people and their language communities. We argue that parallels can be drawn directly between the principles responsible for the development of face and spoken language perception.* *© 2014 The Authors. Dev Psychobiol Published by Wiley Periodicals, Inc.* Dev Psychobiol 56: 1454–1481, 2014.

Keywords: *infant perceptual development; speech perception; spoken word recognition; face recognition; perceptual narrowing; face space; perceptual assimilation*

Language has long been held to be a defining ability that distinguishes humans from other animals. That is, it has been considered to be *species-specific* (e.g., Chomsky, 2006; Deacon, 1997). Relatedly, language acquisition has often been assumed to require an elaborated, specialized neural module that is uniquely devoted to language and thus divorced from more general cognitive skills (Coltheart, 1999; Fodor, 1983). The acquisition mechanisms have been thought to be *domain-specific*. Yet language is not the sole focus of such claims for biological specialization of our perceptual and cognitive skills. Human face recognition is another capability that has also been posited to be specialized (domain-specific) and species-specific (e.g., de Schonen & Mathivet, 1989; Morton & Johnson, 1991). More recent research, however, has shown that both abilities undergo substantial "tuning" by environment-specific experience during the first year of life. Specifically, as infants develop they tend to show: both a "narrowing" of perceptual ability away from discrimination between less experienced stimuli, and an "elaboration" or increase in discrimination and categorization ability for often experienced stimuli. This experience-based perceptual tuning poses some challenge to claims that language and face recognition are biological specializations. Moreover, certain language-like abili-

Manuscript Received: 20 February 2013
Manuscript Accepted: 24 June 2014

Correspondence to: Tamara L. Watson
Article first published online in Wiley Online Library (wileyonlinelibrary.com): 4 August 2014
DOI 10.1002/dev.21243 • © 2014 The Authors. *Dev Psychobiol* published by Wiley Periodicals, Inc.

ties (Gervain & Mehler, 2010), as well as the ability to recognize individual human faces (Peirce, Leigh, daCosta, & Kendrick, 2001), have been demonstrated in other animals, leading some theorists to question species-specificity for both abilities. There has also been increasing experimental evidence on the development of spoken language perception and face recognition in infancy that indicate the processes involved may not be entirely domain-specific (e.g., see Bahrick & Lickliter, 2012; Bulf, Johnson, & Valenza, 2011; Pascalis, de Haan, & Nelson, 2002; Scott & Monesson, 2009; Scott, Pascalis, & Nelson, 2007; Walker-Andrews, 1997). Importantly to the present article, this evidence also suggests fundamental parallels in the ways the two skills emerge in infants. The purpose of this article, therefore, is to examine the parallels in the development of spoken word and face perception in infancy and outline a proposal of their theoretical implications.

Many things that humans can do outwardly appear to involve quite separate and qualitatively different skills, such as language, face recognition, music-making, dancing, mathematical calculation, etc. Yet similar developmental trajectories for any seemingly distinct pair of abilities could offer a clue that both skills may be underwritten by a common fundamental set of mechanisms deployed to subserve disparate functions. As the link between perceptual behavior and the sensorimotor and cognitive functions of the brain is increasingly revealed by research on infant development, we envisage that these "fundamental mechanisms" could be functioning at many different levels of resolution spanning neuroscience and psychology. For example, the means by which neural connectivity in sensory cortex is shaped based on experienced stimulation during infancy could be common across sensory domains. Despite the experienced input differing across the sensory modalities, any common constraint on the development of neural connectivity patterns would result in the development of these sensory skills sharing important characteristics that would be apparent in the developmental trajectory of both skills.

Based on the integrative review of the development of face and spoken word perception skills that we present here, we propose that such a fundamental set of mechanisms underpins both of these abilities and is evident in the perceptual behavior of developing infants. We posit that these mechanisms/principles organize incoming sensory information into meaningful domain-relevant categories, such as the phonemes (consonants and vowels) of words in native-language speech, or the faces of individuals and subgroups within our social circle. Additionally, we will outline the importance of systematically structured variability in the natural environmental input for the infant's acquisition of these perceptual abilities. We propose that utilizing such natural variability might lead to similar outcomes in terms of perceptual spaces, or internal representations, for speech and faces. Importantly, we posit that these variability-based perceptual spaces are organized around the critical dimensions of variation that perceivers discover across the variable surface details of the speech and faces they experience in their environment (see, e.g., Best, in press; see also, E. J. Gibson, 1969; J. J. Gibson & E. J. Gibson, 1955). Not only should these perceptual spaces be based around the dimensions of variation, we propose that the perceptual space should be considered to be composed primarily of these dimensions, rather than being composed of a suite of exemplars or even of norms that specify the central tendency of each dimension.

WHAT KIND OF LEARNING IS INVOLVED?

The type of developmental learning we are talking about is fine grained, the foundation for skilled discrimination between and categorization of individual tokens (e.g., spoken words or individual faces) that exist within a crowded environmental space. The environmental space can be considered crowded if the inputs or signals needing individuation share a high level of similarity across multiple dimensions of structured variability. For example, faces vary within certain constraints along several visible dimensions, yet all share a common configuration with two eyes arranged above a nose that is above a mouth, in which small differences in, for example, spacing between the eyes, can change their appearance dramatically. In spoken language, multi-dimensional variability is also ubiquitous. Very small physical differences in production of the consonants or vowels of a word (i.e., *phonetic* differences) can signal large differences in meaning. For example, the single phonetic difference between the words PARK and BARK is that the vocal cords start vibrating to produce voicing a few tens of milliseconds later in P than in B. From within this type of crowded environmental space one person or word often also needs to be discriminated/recognized across a wide range of transformations or systematic variations. For example, the identity of a person needs to be recognized despite a change in facial expression, lighting and pose relative to the viewer. The recognition of individual words needs to be maintained across differences in the speech styles and emotional expression of the voice of an individual speaker, and across speakers including those who speak with different accents. Moreover, the same visual or auditory input may need to be used in a variety of ways. For example, the sex of the face or of the speaker may also need to be established across all the variations previously outlined, as may race, age etc. In short, a range of categorizations can and must be made from the same object, involving many variations in appearance and/or sound. The distinctions that need to be made about the environmental input can be pictured as being supported by a "perceptual space" that describes this information within the perceptual/cognitive system of the perceiver. In the case of spoken word recognition and face recognition we can call these internal perceptual spaces the perceiver's *word space* and *face space* (Valentine, 1991), respectively. To achieve an extremely flexible perceptual proficiency, many different aspects of the systematic variability between experienced instances in the environment will need to be characterized within the perceptual space. Very different aspects of the incoming information signal that the same word has been spoken by a male compared to a female, relative to those signaling that the same word has been spoken by two females with different accents. Likewise, the cues signaling that two people share the same facial expression will be different from those that signal that a person is from a different race than the perceiver is. While other kinds of perceptual activities can also require skilled perception (e.g., musical abilities), we focus on the acquisition of face recognition and spoken word and phoneme perception skills. They are both obligatory to developing and maintaining social relationships, and thus are central to us as humans.

COMMON DEVELOPMENTAL PROGRESS ACROSS SENSORY DOMAINS

Researchers often acknowledge that spoken language and face recognition are comparable in that perceptual narrowing occurs in both areas (e.g., Lewkowicz & Ghazanfar, 2009; Pascalis et al., 2002; Scott et al., 2007). Perceptual narrowing is the observation that young infants have the ability to sense a wide variety of stimuli, but these abilities become selectively narrowed as a result of exposure to the specific patterns of stimulation in their environment. Thus perceptual narrowing refers, on the one hand, to developmental improvement in perception of often-experienced stimuli, reflecting the strengthening of neural pathways that are consistently stimulated. Perceptual skills are believed to decline, on the other hand, for stimuli the individual is not exposed to, as a result of unused/unstimulated neural pathways becoming less efficient through processes such as synaptic pruning.

The similarity in developmental trajectory in both the speech/spoken word perception and the face perception domains implies to us that a common principle or set of principles is responsible for the development of these and possibly other skilled categorization and discrimination abilities (music perception, for example). We will first consider separately the development of spoken word and phoneme perception, and face perception. We will then outline the development of audio-visual capabilities across face and speech perception. What is apparent in the first year of life is that infants show incredible initial sensory acuity, and that their perceptual skills become tuned by the specific sensory environment they experience and we argue, as well, by the presence of structured variability within that environment.

DEVELOPMENT OF SPOKEN WORD AND PHONEME PERCEPTION

Here we present an overview of development of speech perception and spoken word recognition ability.

Newborns (birth up to 2 months)[1] have a preference for normal speech compared to speech played backwards, filtered or computer-modified, for example, sinewave speech (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Vouloumanos & Werker, 2007a, b), but do not yet prefer speech over animal vocalizations or natural environmental sounds (Shultz & Vouloumanos, 2010; Vouloumanos, Hauser, Werker, & Martin, 2010). Within the speech domain, however, newborns do already prefer infant directed speech (IDS) over adult directed speech (ADS) (Cooper & Aslin, 1990). IDS is found to contain a wider range of variation than ADS along a number of dimensions. This increased yet systematic variation is important to our proposed theoretical framework, and we discuss it in detail below in the Variability in Infant Directed Spoken Interactions Section. Importantly for insights about perceptual narrowing/attunement, newborns also show a preference for their mother's language (Byers-Heinlein, Burns, & Werker, 2010; Mehler et al., 1988; Moon, Cooper, & Fifer, 1993), and their mother's voice (Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld,

[1] "Newborn" in the infant development literature generally refers to the period between birth and up to 6–8 weeks postnatal, as in many ways infants during this period have quite different behavioral, cognitive, and neuropsychological characteristics from those of infants 2 months and older. Many studies of newborn speech perception and listening preferences have focused on the first few minutes, hours, days after birth, but some have included infants between 4 and 6 weeks.

1978). While the preference for normal rather than distorted speech could reflect biological and/or experience-based influences, their very early preferences for maternal language(s) and voice strongly suggest some influence from in utero auditory experience, which is compatible with the fact that the fetal auditory system is functional during at least the final prenatal trimester (Lickliter, 1993).

In keeping with these more global preferences, newborns are able to make a range of finer speech discriminations that appear to set them up for learning about their spoken language environment. Newborns (as we have defined, i.e., birth to 2 months[1]) are able to discriminate most consonant and vowel contrasts found across the languages of the world (e.g., reviews Aslin & Pisoni, 1980; Werker, 1989). They can also detect acoustic cues to word boundaries (Christophe, Dupoux, Bertoncini, & Mehler, 1994) and discriminate words that differ in lexical stress (Sansavini, Bertoncini, & Giovanelli, 1997). They appear to be poised to make the most of the varied input they are exposed to, and to learn most from their primary caregiver, including even prenatal experience with mother's voice and language(s). At this stage of development the newborn can be considered optimally attuned for experiencing the complex variations in speech input, with which their perceptual space is molded to their native language environment.

Between 2 and 6 months of age infants show perceptual patterns and preferences in the speech domain, some of which indicate further attunement to native speech properties. Whereas newborns' preference for natural over artificial audio stimuli is quite broad, extending from speech to rhesus monkey calls and other natural non-speech environmental sounds, these listening preferences narrow down to human speech by 3 months (Shultz & Vouloumanos, 2010; Vouloumanos et al., 2010). Additionally, 4-month-olds, like newborns, can discriminate between spoken passages of languages from different rhythmical classes (e.g., English, a stress-timed language, versus Japanese, a mora-timed pitch accent language, or French, a syllable-timed language that lacks stress contrasts) (Nazzi & Ramus, 2003), but by 5 months of age infants have been shown to discriminate languages from within the same rhythmical class (e.g., English and Dutch, both stress-timed languages) (Nazzi, Jusczyk, & Johnson, 2000) if one of the languages is familiar. Despite showing some tuning to the global prosodic properties of connected speech, however, infants in this age range do not yet demonstrate tuning to the native consonant or vowel contrasts of their own language environment as they appear to still be able to discriminate most consonant and vowel contrasts they have been tested on, whether used in their native language or only in languages they have not experienced (Aslin, Pisoni, Hennessy, & Perey, 1981; Eilers, Gavin, & Wilson, 1979; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Trehub, 1976).

There are some key exceptions to this pattern that are interesting. At 4 months of age discrimination is poor for certain *native* consonant contrasts, such as English /d/-/ð/ (as in *doze-those*) (Polka, Colantonio, & Sundara, 2001; see also Narayan, Werker, & Beddor, 2010, for evidence of early difficulties in discrimination of native nasal consonant contrasts). Conversely, there is also evidence of differences in perception of Kikuyu (Kenya) stop consonant voicing contrasts by native- versus non-native-learning infants (English-learning) at 2 months (Streeter, 1976). These latter findings suggest that not all consonant contrasts are created equal, with some apparently being influenced earlier by experience and others requiring more extended experience. These differences could pose some difficulties for a simple view of perceptual narrowing.

Of particular interest for our hypothesis regarding the importance of experiencing natural systematic variation, 4- to 5-month-olds do show perceptual constancy for native vowel categories and contrasts, specifically recognizing the same vowel when it is spoken by both adult and child speakers of either gender (e.g., Kuhl, 1979, 1983).

Between 6 and 9 months of age additional perceptual narrowing to the finer-grained phonemic categories of native speech occurs. By 6–8 months, infants' discrimination of non-native vowel contrasts is declining (e.g., Polka & Bohn, 1996, 2003; Polka & Werker, 1994), but there is no evidence of a decline in discrimination of non-native consonants until several months later (for a review, see Werker & Tees, 2005; also section on 9–12 months, below). Moreover, by 6 months infants show within-category perceptual differentiation of good versus poor exemplars of native vowels but show no evidence of doing so for non-native vowels (Kuhl et al., 1992). At this same age, infants are also able to segment words from continuous speech, and show a preference for content words as compared to function words (Shi & Werker, 2001), even though specific function words (e.g., THE) occur much more frequently than specific content words (e.g., DOG). This preference for content words may be taken to reflect the infant's apparent disposition at this age for engaging with stimuli from the informationally richest rather than the statistically most frequent categories experienced. Although specific content words are much less numerous in speech than specific function words they tend to be longer, more variable and arguably provide the core meaning of a sentence. By 7–8 months, infants have also developed the ability to recognize familiarized words across a change in amplitude (loudness), but have not been found to generalize this recognition across changes to fundamental frequency, speaker, gender, or affect (Singh, Morgan, & White, 2004; Singh, White, & Morgan, 2008), unless the words were previously highly familiar to them (e.g., MOMMY and DADDY) (Singh, Nestor, & Bortfeld, 2008).

Between 9 and 12 months of age, discrimination of many non-native consonant contrasts shows a dramatic decline (e.g., Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Best & McRoberts, 2003; Werker & Lalonde, 1988; Werker & Tees, 1984; Werker, Yeung, & Yoshida, 2012; Yoshida, Pons, Maye, & Werker, 2010; see reviews by Best, 1994; Werker & Tees, 2005) and discrimination of many non-native vowel contrasts has declined further from the levels seen at 6–9 months (e.g., Polka & Werker, 1994). There are some interesting and informative exceptions to this pattern, however. Discrimination of some non-native consonant contrasts remains good even past 12 months of age despite considerable perceptual narrowing for other contrasts at this age (e.g., for English-learning infants, Tigrinya dental vs. bilabial ejective consonants: Best & McRoberts, 2003; Zulu dental vs. lateral click consonants: Best, McRoberts, & Sithole, 1988; and Nu Chah Nulth velar versus uvular vs. pharyngeal fricatives: Tyler, Best, Goldstein, & Antoniou, 2014). These results for non-native consonants suggest that discrimination is maintained over the 10–12 month period only if the articulator or feature distinctions are used in native consonant contrasts, or if the articulatory properties of the non-native consonants are so highly discrepant from native consonants that adults perceive them as non-speech sounds (outside the native phonological system altogether). There are, conversely, certain *native* contrasts that are *difficult* for younger infants to discriminate, and some of these continue to be poorly discriminated until as late as 4 years of age (such as /d/ vs. voiced TH as in *<u>th</u>ere*) (Polka et al., 2001; Sundara, Polka, & Genesee, 2006; see also Cristià, McGuire, Seidl, & Francis, 2011). However, there is also evidence of perceptual elaboration, or *improved* discrimination for certain other native contrasts (such as English /r/ vs. /l/), discrimination of which shows a decline by this age in infants whose native language does not use these contrasts, for example, Japanese (Kuhl et al., 2006).

Concerning development of perceptual constancy, by 9 months infants can recognize words across discrepancies not only in amplitude but also in fundamental frequency from previously unfamiliar words and non-words they have been familiarized with in the laboratory (Singh, White, et al., 2008). By 10.5 months their ability to recognize such newly familiarized words extends as well across a change in the emotional expression of the speaker, to new speakers and to differences in speaker gender (Singh et al., 2004; Singh, Nestor, et al., 2008).

By 9–12 months, then, infants' perceptual word space is a developing model of both the phonemes and the spoken words of the language environment they will operate within, as opposed to an open model of all possible spoken languages. Not only is perceptual *narrowing* occurring, however, a strategic *perceptual elaboration* is also taking place. Recognizing a word across speakers or affects reflects a skill that results from experience not just of spoken language itself but also from experience with other information in the input such as contextual clues (e.g., facial expression), and dynamic feedback between the infant and the speaker. In this sense the learning, while still being informed by or capturing the variability in the signal, is no longer purely statistical. The kind of perceptual invariance that is emerging is the beginning of being able to abstract away from pure, surface-level environmental statistics toward deriving more abstract rules that will support the formation of categories that include any number of quite dissimilar forms of spoken words.

Although perception of vowel and consonant contrasts have become largely tuned to the native phoneme inventory by 9–12 months, and the ability to segment and recognize familiarized words from connected speech has become fairly robust to variations in speaker, gender, emotion and other superficial speech properties, word learning and word recognition abilities are still not adult-like at the end of the first year. Eleven- to 12-month-olds prefer listening to sets of words that are well-known to children of this age-group, as compared to listening to unfamiliar, low-frequency adult words they have never before heard. However, unlike adults, this familiar word preference extends broadly to mispronunciations of those words, such that they appear to accept non-words that differ by a single consonant from words that they know, as viable variants of the known words, for example, *VABY[2] and *GAIRE are equally preferred as BABY and BEAR (e.g., Hallé & de Boysson-Bardies, 1994, 1996; Mulak & Best, 2013). In short, their perception of words still has further important refinements to undergo (not surprisingly).

Beyond 12 months, in the first half of the second year, there is further perceptual attunement in children's learning and recognition of spoken words, both in

[2] The * indicates that this is not a real word.

terms of recognizing phonemic contrasts that distinguish words and in terms of constancy in recognizing words across variations that do *not* change word identity (see Best, in press; Best, Tyler, Gooding, Orlando, & Quann, 2009; Mulak & Best, 2013). At 14–15 months, several studies suggest that children's ability to distinguish between newly learned words and single-consonant changes to those words is still fairly tenuous (Stager & Werker, 1997; Swingley & Aslin, 2000; Yoshida, Fennell, Swingley, & Werker, 2009), but is somewhat improved over the 11- to 12-month-old's in that they can recognize a change and reject a mispronunciation if the word is either previously very familiar to them, or the task demands are minimized while contextual support for word recognition is optimized (Fennell & Waxman, 2010; Fennell & Werker, 2003). By 18–19 months performance on such word-recognition and word-learning tasks, however, is much more robust and reliable (Swingley, 2003, 2007). In addition, somewhere between 15 and 19 months of age toddlers seem to move from being able to identify familiar words only when spoken in their native accent to also being able to identify the words when spoken in an unfamiliar accent (Best et al., 2009; Mulak, Best, Tyler, Kitamura, & Irwin, 2013). It seems that they become able to abstract their experiences to assess whether a word spoken in an accent they may never have encountered can be related to the phonological form of words they have experienced in their own native accent. As regional accents can change the phonetic details of words dramatically, this is a sophisticated form of perceptual constancy where there may be multiple interpretations of the input. The word the infant is listening to in another accent cannot fit any stored exemplar or be represented by any existing experience-based prototype. This is because the unfamiliar accent has previously not been encountered and it changes the low level phonetic detail of the word substantially. Thus, this kind of perceptual constancy is impossible to describe within a system that represents its input by extracting normalized prototypes or even by encoding an extensive list of experienced exemplars. This kind of constancy would seem more to be supported by coming to recognize that words occupy malleable and dynamic regions along multiple dimensions in a perceptual word space and that in any given situation some of these dimensions will be more important than others in identifying the word. This type of *phonological* constancy for recognizing words must arise from discovering recurring multidimensional patterns within the structured variability of the language spoken in the child's environment.

DEVELOPMENT OF FACE PERCEPTION

Here we present an overview of development of face recognition ability.

Newborns (birth up to 2 months[1]*)* also show surprising visual capabilities that suggest the visual system at birth has biases to attend to basic structural properties of faces and hence to experiences that will allow important distinctions to be made. Newborns show a preference for schematic faces compared to scrambled versions of the same stimulus (Johnson, Dziurawiec, Ellis, & Morton, 1991) that is likely driven by a preference for images that share similar stimulus energy to faces (Kleiner & Banks, 1987) and additionally a similar structure (Kleiner, 1987). Given a stimulus composed of the parts of a face newborns will look preferentially at stimuli that are top heavy (Cassia, Turati, & Simion, 2004) and also at the spatial arrangements of basic shapes that most closely convey a face-like appearance (Cassia, Valenza, Simion, & Leo, 2008). When presented with two different photos of the same face newborns will also preferentially look at the version of the image containing the face that is gazing directly at them (Farroni, Csibra, Simion, & Johnson, 2002). They are also able to discriminate between images of faces on the basis of both external and internal facial features when each is presented in isolation. When hairline is kept constant, newborns who have some experience with faces show a preference for faces that adults rate as attractive (Slater et al., 1998). Mathematically averaged faces are consistently rated as more attractive than the individuals contained within the average (Langlois & Roggman, 1990). However, when the hairline is visible, it appears that it is the preferred cue and precludes processing of the internal features (Turati, Macchi Cassia, Simion, & Leo, 2006). Despite demonstrating sophisticated perceptual abilities, newborns' apparent favoring of the hairline is in line with their lower acuity than older children and adults. It has been suggested that newborns rely on spatial frequencies around .5 cycles per degree of visual angle when recognizing static faces (de Heering et al., 2008), making the hairline a very salient visual cue and highlighting that the developmental progress of the visual system itself is an important factor in determining the information a newborn can extract from a face.

Intriguingly, face recognition also shows early signs of perceptual constancy. Newborns only 1–3 days of age are able to match faces across a rotation of 45°: between a full-face and a 3/4 view (Turati, Bulf, & Simion, 2008). While it is not yet established what cues newborns use to carry out this task it is a skill that is clearly important for forming meaningful perceptual categories about face identity. Related to this kind of perceptual constancy, newborns' ability to recognize a face is enhanced by viewing the face undergoing smooth rigid head motion in the form of a left/right rotation, compared to viewing the same series of video frames presented out of order (Bulf & Turati, 2010). Despite this, it seems that not all motion is beneficial in this way. Neither the rigid (Guellaï, Coulon, & Streri, 2011) nor the non-rigid motion of a speaking face shown without sound is sufficient to promote recognition of a new face at this age (Coulon, Guellai, & Streri, 2011), possibly due to the complexity of the motion used in these studies compared to the rotational motion in the Bulf and Turati (2010) study. The addition of speech in concert with a moving face, however, appears to provide a newborn with the required information to look preferentially at their mother the first time they see her face in person (Sai, 2005) and even to recognize a stranger presented in a photo after audio-visual familiarization (Coulon et al., 2011).

Many of the perceptual capabilities discussed at this age need not be strongly face specific, in that they could reflect basic biases toward key aspects of any visual stimulus that, when found in combination as in a face, make such stimuli very salient for newborns. From 2 months, however, infants begin to show face-specific perceptual effects.

Between 2 and 6 months of age, infants demonstrate quickly changing perceptual effects from their experiences with faces in their environment. The abilities of infants around 2 and 3 months suggest their perceptual space has begun to reflect a foundational structure based on experience. This nevertheless remains extremely sensitive to variations between faces that adults, conversely, show difficulty in discerning. At around 2 months of age infants begin to show an eye movement scanning preference for the eye regio of a face (Hainline, 1978; Haith, Bergman, & Moore, 1977). They have also been shown to prefer scrambled stimuli that retain the phase spectrum of natural faces (spatial frequencies occurring in the image are preserved but their phases are scrambled) and therefore look face-like to adults (Kleiner & Banks, 1987). Infants at 3 months are able to discriminate equally well between faces of their own race and also other races (Kelly et al., 2009; Kelly, Quinn, et al., 2007). At the same age, however, infants have developed preferences for faces similar to those in their most frequently encountered groups (see Sugden, Mohamed-Ali, & Moulson, 2014 for an analysis of an infant's most frequently encountered faces). They have a preference for faces of their primary caregivers' race (Kelly, Liu, et al., 2007; Kelly et al., 2005), and upright (but not inverted) faces that are the same sex as their primary caregiver, whether male or female (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002), as the primary caregiver is an infant's most viewed face (Bushnell, 2001; Sugden et al., 2014). Even more specifically, infants with female caregivers show a preference for female same race faces (Quinn et al., 2008). Interestingly, at 3 months infants' ability to discriminate between faces that are the same sex as their primary care giver is clear, but there is debate as to whether they can discriminate between individuals of the other sex. Thus, the statistics of the environment may indeed be playing an important role at this age and inviting the question of what role a small but significant exposure to a face category plays at this age (Quinn et al., 2002). This suggests, however, that infants at 3 months retain a flexible perceptual face space that has begun to acquire the statistics of their environment.

Also highlighting the importance of the statistics of the environment and in particular the importance of sufficient variability in learning to categorize faces, 3-month-old Caucasian infants do not show evidence of a novelty preference to a new Asian face after habituating to a single Asian face but do show a novelty preference to a new face after habituating to just three Asian identities (Sangrigoli & De Schonen, 2004). That is, at least modest variation among individual faces during the familiarization phase may foster non-Asian infants' ability to show significant discrimination among Asian individuals' faces.

Relatedly, at 3 months of age infants have also been shown to extract the commonality between sets of faces (de Haan, Johnson, Maurer, & Perrett, 2001; Rubenstein, Kalakanis, & Langlois, 1999). This effect is termed *prototype extraction*, where the average of a set of faces is responded to as being equally familiar as any of the experienced exemplars (Rosch, 1978; Rosch, Simpson, & Miller, 1976). Although findings like this are often interpreted to imply that face space is based around these prototypes, our interest in such results is that they also show that infants are sensitive to the statistical structure of their environment.

Despite the beginnings of sophisticated face recognition skills, infants at 3 months are not yet showing all the hallmarks of adult face recognition. In particular, infants' categorical boundaries between faces are as yet quite fuzzy. Four-month-old infants' perception of the identity of morphed faces was tested and it was found that infants treat a morph that contains up to 70% of a face they had never before seen as though it were a familiarized face.

That is, only 30% of a previously seen familiarized face was required in the morph for the face to be treated as familiar (Humphreys & Johnson, 2007).

Between 6 and 9 months of age a perceptual narrowing to categories most often experienced seems to occur for faces, as is observed in the spoken word research. At 6-months infants still show a novelty preference for previously inexperienced, unfamiliar individuals of a race other than their own (as long as it is not too dissimilar: Kelly, Liu, et al., 2007) and even monkey faces (Pascalis et al., 2002). This suggests that as yet they can still differentiate individual members of face-type categories with which they have little or no experience.

At 6 months of age, infants have also been shown to maintain a spontaneous preference for attractive faces (Rubenstein et al., 1999). These findings have typically been interpreted as evidence of prototype extraction. Indeed, when infants are habituated to three equally attractive individuals they show no recovery of habituation when presented with an average of the three faces but will preferentially look at a novel face (Rubenstein et al., 1999). An important observation from this study is that infants will maintain habituation to a prototype face and will actively respond to new faces, demonstrating not only the ability to form a prototype but also the proclivity to explore variability away from it, rather than to rehearse that prototype. Similarly, Heron-Delaney et al. (2011) showed that non-Asian children between 6 and 9 months old and growing up in Australia only needed 1 hr of exposure to individuated Chinese faces for apparent maintenance of the discrimination of other race faces.

At 7 months of age, infants' response to morph stimuli containing mixes of two faces is showing signs of become more sharply tuned. As noted above, 4-month-olds responded to a 70% new face 30% familiarized face mixture as though it were a familiar face. In contrast, 7-month-olds only respond to an up to 50% new face mixture as though it is a familiarized face (Humphreys & Johnson, 2007).

Between 9 and 12 months. In contrast to infants at 6 months, by 9 months of age infants show a reduced novelty preference for previously seen but untrained individual monkey faces (Pascalis et al., 2002) and seem to only differentiate individual humans of their own race (Kelly et al., 2009; Kelly, Quinn, et al., 2007) unless provided with experience of faces from another race (Anzures et al., 2012). Additionally, by 9 months old infants have been shown to demonstrate integration between internal and external facial features only when presented with an upright but not an inverted own-race face (Ferguson, Kulkofsky, Cashon, & Casasola, 2009). They also demonstrate the ability to form categories according to race but to discriminate only among individuals categorized as own-race (Anzures, Quinn, Pascalis, Slater, & Lee, 2010). This indicates tuning based on the available input and the establishment of the basic foundational structure of the perceptual face space that is similar to that found in adult studies. One question is whether the apparent perceptual narrowing represents a time governed or experience governed developmental window (Maurer & Werker, 2014). In consideration of this, the apparent perceptual narrowing can be reversed with meaningful exposure to categories of faces not seen in the environment. For example, Pascalis et al. (2005) showed that training with a small set of individually named macaque faces was sufficient to prevent the loss of discrimination ability for macaque faces seen at 9 months.

The progression of this apparently reversible perceptual narrowing is not yet sufficiently mapped out to understand concretely the process and timeline of perceptual narrowing involving the range of facial judgments considered here (see Maurer & Werker, 2014, for a review). However, it is apparent that judgments involving rarely-to-never experienced categories become more difficult with age in infancy. Just as has been found in speech perception, we anticipate that the perceptual narrowing in face recognition is accompanied by a concomitant perceptual elaboration that can support additional and more advanced perceptual constancies. This elaboration will be guided not only by better understanding of the statistics of the environment but also by cognitively abstracted categories reflecting social and cultural factors that provide feedback about socially relevant categorizations. This is an area that is as yet under-explored, given that studies of face recognition mostly do not use multiple images of the same face to probe whether infants can recognize constancy of a given face across various transformations (i.e., across different emotional expressions, lighting conditions, dynamic changes over time, etc.).

Beyond 12 months, although a direct analogue of the kind of perceptual constancy measured in the spoken language domain (with clear recognition of words across accents and speakers) has not yet been investigated, and thus cannot yet be assumed within face recognition, it may be that studies of infants' recognition of a particular person across changes in makeup, dramatic changes in hairstyle, emotional expressions, spatial perspective, or even changes in lighting conditions could demonstrate the development of quite sophisticated perceptual constancies in this domain between 15 and 19 months of age, as well. As noted above, to date, little or no research has been done on this aspect of face recognition.

MULTI-SENSORY[3] INTEGRATION OF FACE AND SPEECH

From these brief reviews of the developmental trajectories of both face and spoken word processing it is clear that there is a generally common pattern, specifically a move from a very broad yet apparently unstructured ability to match and differentiate a range of features of speech and faces, toward more experience-dependent capabilities. This shift encompasses a perceptual narrowing away from non-experienced aspects, as well as a very likely, yet under-explored, *elaboration* of constancies within commonly experienced aspects, within the first year of life. It is also apparent, however, that the spoken word and face recognition literatures are based overwhelmingly on uni-modal studies—the auditory modality in the case of speech, and the visual modality in the case of faces. Moreover, research on infant perception of words and faces has often focused on the different types of information that can be gained from the stimuli. For example, face recognition research is often focused on the development of the recognition of identity, an indexical and constant aspect of the one face. On the other hand, spoken word/phoneme perception research is often focused not on the indexical aspects of the voice (e.g., who is talking) but on what words that voice is conveying. This makes the similarities striking but it does also make the two literatures difficult to truly compare.

Therefore, a key area where the strength of a common developmental mechanism should be in evidence is when the multi-sensory aspects of face and speech perception are considered in concert, particularly in the context of face-to-face interactions. It cannot be ignored that when interacting with a person an infant's experience is typically audio-visual. This is particularly significant in developmental research because there is a considerable amount of redundancy in the audio-visual stimulus that can be important to the development of the uni-modal perceptual capabilities. Although many studies present faces (visual) and voices (auditory) in isolation (i.e., uni-modally), an infant is more regularly experiencing live, visible + audible people speaking. This means that they experience the combination of multiple modalities, where a multitude of cues to the same information are present. Studies into infants' ability to capitalize on audio-visual cues suggest that the developmental trajectory of face and spoken language perception are closely intertwined, as they would need to be to take advantage of the powerful multi-modal and amodal cues in the natural environment. When presented with a person speaking there are several aspects of both the visual and the auditory stimulus that are shared across the modalities, in particular, onsets and offsets, the duration, tempo, and rhythm of the two modalities of talking faces (Yehia, Rubin, & Vatikiotis-Bateson, 1998). This information is redundant in that the exact same information can be gained in a fairly equivalent manner across the two senses and it therefore provides an unambiguous cue to aid integration across modalities.

Newborns display what appears to be surprising capability in multi-sensory perception, which clearly suggests that the type of information that is redundant in audio-visual speech is a very important cue supporting the development of both face and spoken word perception. For example, newborns at 3 weeks of age have been shown to spontaneously match audio-visual stimuli (a white light and audio white noise) based on the relative intensity of the stimuli. This was measured via the newborn's cardiac response, which differed systematically depending on whether the relative intensity levels were similar or notably different between the audio and the visual stimulus (Lewkowicz & Turkewitz, 1980). They have also been shown to match monkey facial gestures with vocalizations, with the evidence strongly suggesting they do this on the basis of the synchrony of onsets and offsets of the audio vocalization and the facial gesture rather than matching the quality of the complex sound to the shape of the mouth (Lewkowicz, Leo, & Simion, 2010).

These multi-sensory perceptual abilities can be seen to match an infants' physical acuity capabilities across the senses, being driven by basic amodal (non-specific to a given modality) properties present in natural stimuli, such as the direction and speed and start/stop of a moving, sound-making object. Indeed, it has been proposed that the ability to match audio-visual stimuli at this age is due to the young infant's inability to differentiate reliably among the individual sensory modalities of a multi-modal stimulus, rather than reflecting a particular ability to associate across the two modalities of audio-visual stimuli (the Infant as Synaesthete theory; Maurer, 1993; Maurer & Mondloch,

[3] Multi-sensory perception is here used to mean either simultaneous processing and subsequent matching of stimuli across more than one sense OR processing stimuli in an integrated fashion such that modality of delivery information is almost immaterial to categorization or discrimination performance.

2004). To the extent that young infants show no evidence of differentiating among modalities, then experience with the amodal properties of stimuli that newborns experience naturally, such as the onsets and offsets of audio-visual speech, should be extremely important in driving the development of the separate modalities' processing capabilities. The theory that infants do not differentiate the senses at birth also highlights the importance of taking a whole brain/integrated perspective, rather than a modular view of development of independent perceptual modalities within the first year of life. It clearly suggests the importance of an integrated and domain- and modality-neutral set of mechanisms in the development of skilled perception in infants.

The infant synesthesia theory also accords with the suggestion that redundantly specified stimuli, that is, stimuli that specify the same information through multiple modalities, should be strongly attention grabbing/salient for infants (see Bahrick & Lickliter, 2012). To the extent that the redundant aspect of an audio-visual stimulus is dealt with similarly at a neural level (i.e., increased intensity resulting in increased firing), an intrinsic connection between the sensory areas of the brain would ensure that this aspect of the stimulus is activating these two separate sensory areas in concert with each other, making the power/salience of the stimulus greater in effect. It is plausible that from this base the infant is able to begin to experience the features of their multi-modal environment that are not redundantly specified. That is, the patterns that are statistically related across the senses provide a foundation from which to contrastively experience those aspects of a stimulus that are uni-modally specified.

Between 2 and 6 months. From about 2 months of age infants are beginning to respond to audio-visual stimuli based on experience gained within their first months. At 2 months of age infants will respond differentially to multi-modal, moving faces depicting different emotions and in particular they also mirror ("imitate") expressions of joy and sadness presented to them (Haviland & Lelwica, 1987). The multi-modal nature of these stimuli is considered crucial to the discrimination of emotions at this early age (for a review, see Walker-Andrews, 1997). Additionally, 2-month-olds can also match some vowel sounds to the facial motion used to produce the sound (Patterson & Werker, 2003). It has been proposed that infants at this young age are matching the sound and facial gesture on the basis of the full spectral and amplitude properties of the stimulus, as even infants as old as 4.5 months do not show evidence of matching vowels on the basis of only simplified temporal or amplitude changes (Kuhl & Meltzoff, 1984; Kuhl, Williams, & Meltzoff, 1991).

By 3 months of age it has been found that infants are able to associate new people's faces with their voices, looking longer at novel combinations of recently familiarized faces and voices (Brookes et al., 2001). They have also been shown to search visually for a parent's face when they hear the parent's voice unaccompanied by their face (Spelke & Owsley, 1979), suggesting that an association between the identity of a face and voice is established early.

Infants' sensitivity to the correspondences between audio and video (talking face) presentations of specific vowels and consonants is such that 4.5-month-old infants look significantly longer at a face whose articulation matches a synchronously played audio vowel, when two videos of the same face articulating two different vowels are presented synchronously side-by-side (Kuhl & Meltzoff, 1982, 1984). In support of the recognition of vowels across modalities, and as a strong demonstration of the multi-sensory nature of developmental learning, 3- to 4-month-old infants have been shown to imitate facial movements articulating vowels when the vowel sound is paired with the corresponding facial motion but not when the auditory stimulus does not match the visual facial motion (Legerstee, 1990).

Infants also recognize articulatory correspondences between seen and heard speech syllables when the two modalities are presented completely separately rather than simultaneously. Several recent studies assessed infants' recognition of multi-modal consonant correspondences in a task involving familiarization to one of two contrasting audio-only syllables, followed by a test phase in which infants' looking preferences were assessed to silent videos of a speaker producing syllables that matched versus mismatched the preceding audio consonant. Four-month-old infants fixated longer on the face whose articulations corresponded to the preceding audio stimuli, for both native and non-native consonant contrasts (Best, Kroos, & Irwin, 2010, 2011; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009; see also Bristow et al., 2009, for ERP evidence of such multi-sensory sensitivity in 2.5-month olds). At 4 months of age, infants also still match monkey calls to their facial gestures, looking longest at the facial gesture matching the monkey call (Lewkowicz & Ghazanfar, 2006). As yet, however, at 4.5 months infants have not been found to take sex of a face into account when matching vowels across modalities. When two articulating faces are presented side-by-side infants at this age will apparently ignore a mismatch in sex and match according to the corresponding vowel sound (Patterson & Werker, 2003).

When presented with audio-visual IDS (infant-directed speech), from 4 months of age infants are able to detect when changes in the lexical-syntactic content, in a speaker's sex, or in synchrony occurs in any modality (auditory, visual, or audio-visual). Interestingly, when presented with ADS infants at both 6 and 8 months have not shown evidence of detecting the same changes when they occur in the auditory domain only, despite being able to detect them in the visual-only and audio-visual modality (Lewkowicz, 1996). This not only highlights the multi-sensory capabilities of infants but also the importance of the properties of infant directed interactions, such as the presence of systematic co-variation in the input. This aspect will be discussed below (see the Variability in Infant Directed Spoken Interactions Section).

At 5 months of age, infants have been shown to associate the audio with the matching visual component of an audio-visual presentation of a consonant-vowel-consonant-vowel string, preferring to view stimuli that matched in phonemic content and not just synchrony (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). At this age, infants are also able to learn abstract patterns created by systematically paired looming visual objects and auditory syllables. Infants did not learn the pattern when the syllables were presented without a visual stimulus or when the syllables were paired with objects unsystematically. This demonstrates rule learning, which at this age appears to be driven by the systematic relationship between the combined sensory inputs (Frank, Slemmer, Marcus, & Johnson, 2009).

Related to this finding, by 5 months of age infants are able to recognize the correct association between static human versus monkey faces and human speech sounds versus monkey calls, despite not having any particular experience with monkey sounds (Vouloumanos, Druhen, Hauser, & Huizink, 2009). At this age, infants also show evidence of integrating conflicting audio-visual presentations of speech phonemes such that they appear to experience the McGurk effect, in which a synchronously presented visual va/audio ba is heard as a va, just as adults do (Rosenblum, Schmuckler, & Johnson, 1997).

Between 6 and 9 months. Up to 6 months of age, infants have been demonstrating a preference for cross-modal stimulation and an increasing ability to carry out complex perceptual tasks across modalities. Yet infants less than 6 months do not show evidence of a decline in ability to carry out tasks with speech categories they have not experienced in their native language environment. For example 6-month-olds are able to match non-native consonant contrasts (/b/ and /v/ for Spanish learning infants) across separate auditory and visual presentations (Pons et al., 2009).

Yet, as further evidence of beginning to associate aspects of the face with aspects of a voice that are not redundantly specified, at 7 months of age infants can match the emotion of a face and voice across separate presentations of the two modalities (Walker-Andrews, 1986). Moreover, at 8 months infants are able to associate the sex of a face with that of a voice when the voice is articulating the same vowel as the face (Patterson & Werker, 2003).

There is also some evidence of perceptual narrowing of cross-modal perceptual abilities in infants older than 6 months of age. While 6-month-olds are able to match monkey facial gestures with their associated call, when tested at 8 months infants no longer show evidence of making this match (Lewkowicz, Sowinski, & Place, 2008). It is hypothesized that at this age infants are no longer relying on basic/amodal aspects of the stimuli to carry out these kinds of tasks, and that without continued experience with cross species perception the task becomes increasingly difficult (Lewkowicz et al., 2008).

At this age, multi-modally redundant stimuli still appear to capture attention. Crucially, though, such redundant speech streams also appear to aid subsequent recognition of words that occurred in the stream. At around 7.5 months of age, for example, when infants were familiarized with two simultaneous, competing audio speech streams, they subsequently recognized words from one of the streams if the video of that speaker had been presented synchronously with that stream during familiarization (Hollich, Newman, & Jusczyk, 2005).

Between 9 and 12 months there is continued evidence of further perceptual narrowing or experience-based elaboration for multi-modal stimuli. Eleven-month-olds recognize a match between separately presented audio-only followed by visual-only presentations of a native consonant contrast, but they do not show evidence of this for certain non-native consonant contrasts such as ejective stops from the Ethiopian language Tigrinya (Best et al., 2010, 2011; Pons et al., 2009), although they succeed with other crucially different non-native consonant contrasts that are categorized by adults as non-speech sounds, that is, click consonants from the Botswanian language !Xòõ (Best et al., 2010, 2011). The results across that set of studies indicate that by 11 months infants have become perceptually attuned to detect just those multi-modal articulatory correspondences that are relevant to native speech contrasts.

Between 10 and 12 months, infants also demonstrate the emergence of the ability to match the identity of their native language across modalities for connected speech. However, they do not show evidence of doing this with an unfamiliar language (Lewkowicz & Pons, 2013).

In summary, the evidence of a progression toward becoming a skilled perceiver of audio-visual social interactional stimuli suggests that in newborns the multi-modally redundant aspects are extremely important. Given some experience with audio-visual faces, infants begin to extract statistics of the world in relation to commonly co-occurring aspects of the stimuli. As time progresses they become able to recognize and learn associations between increasingly more complex multi-modal patterns. At the same time as these increasingly more sophisticated associations are emerging around 9–12 months, infants' ability to recognize cross-modal matches for rarely encountered classes of audio-visual stimuli that they could/did detect early in the first year, appears to decline just as has been found in the auditory and visual domains separately (see Lewkowicz & Ghazanfar, 2009).

SOME CAVEATS ABOUT COMMON DEVELOPMENTAL PROGRESS ACROSS SENSORY DOMAINS

We propose that the acquisition of perceptual skill may depend on common developmental and representational mechanisms that could be expected to cause developmental milestones to be reached at the same age across domains. Despite the apparently similar developmental trajectory of the perceptual skills we have reviewed here, with perceptual narrowing occurring toward the end of the first year of life in both domains, we do not necessarily expect to find developmental milestones in lock step across domains. Note also that alongside the apparent narrowing by 12 months, we also predict perceptual elaboration by this same age, despite there still being insufficient research on this issue to date. Moreover, the posited common perceptual development mechanism across the two domains will still need to be implemented by, or interact with, the individual neural machinery and the primary input sensory modality(s) of the domain in question. This could lead to different timeframes for the emergence of similar developmental milestones across faces and spoken words. For example, although the retina is not considered to be important in the representation of faces within the visual system, it is nonetheless necessary to visual perception of faces. Its functional maturity will affect the data available to carry out statistical learning about faces. The retina, and the visual system more generally, develop at a much slower rate than the auditory system in the fetus (Gottlieb, 1971; Lickliter, 1993), with the auditory system structurally complete and functional much earlier prenatally. Additionally, the input for spoken language learning is available in the womb, which cannot be said of the visual input necessary for learning to recognize faces (see Lickliter, 1993). Therefore, development of auditory skill may appear to precede analogous visual skill simply because the auditory system itself has been receiving relevant data from the final prenatal trimester whereas the visual system processes little data prior to birth. That is, the input data collected may yet be insufficient in one domain (e.g., vision), while in another (e.g., audition) sufficient data has already prompted the next stage in development.[4]

As can also be seen after reviewing the developmental literature, the myriad of perceptual decisions that can be made when considering the audio and visual aspects of a face, including a talking face, make finding truly analogous perceptual capabilities across faces and spoken words challenging. Moreover, the multi-modal nature of the natural stimuli and the demonstrated importance of multi-modal stimulation to infants strengthen the case for a common mechanism, yet at the same time complicate our ability to design experiments, and to draw conclusions from prior research within single sensory modalities. Without expecting to be able to draw exact milestone comparisons between domains, comparisons of the separate and combined developmental trajectories outlined above is suggestive of a similar development process.

SUMMARY OF DEVELOPMENT

Based on the developmental trajectories of both face and language perception, it appears that at birth infants possess a largely untuned and basic perceptual space capable of differentiating many properties of both faces and speech whether presented in concert (audiovisual talking faces) or separately. Through the first approximately 6 months infants' sensory exploratory behaviors appear to be biased toward collecting data about the statistical structure of the particular stimulus environment within which the infant is immersed. The statistical representation of the basic dimensions of the sensory environment then forms a foundation that is thereafter a base from which perceptual constancies are

[4] Indeed the relevant factor in the developmental trajectory of experiential effects on speech perception does appear to be the combined auditory experience during the prenatal and postnatal periods: preterm infants show a decline in discrimination of nonnative consonant contrasts at the same *gestational* age as full-term infants, rather than at the same *postnatal* age (Peña, Werker, & Dehaene-Lambertz, 2012).

established and crucial abstractions can be perceived. A skilled perceiver is able to both make judgments about very fine scale differences, and to tackle constancy problems that go beyond the basic surface statistics of the detailed input. The transformation from statistical learner to abstraction learner appears to coincide with the onset of perceptual narrowing across both the face and speech domains. However, we propose that rather than losing perceptual capability, the process of perceptual narrowing represents the transition from basic statistical perceiver to an abstraction learner and could more comprehensively be conceived of as a time of perceptual elaboration. Infants become able to deal with more abstract regularities in their environment, such as the constancy of a word's phonological structure and meaning despite a change of emotional affect or speaker or accent, or the facial identity of a person despite a change in hairstyle or makeup or emotional state.

While similarity of the developmental trajectory of face and spoken word perception and the crucial multisensory aspects of these kinds of stimuli is striking, these apparent similarities could be driven by dissimilar developmental principles. However, other evidence can also be brought to bear to bolster the claim for a common developmental principle behind both.

OTHER EVIDENCE FOR A COMMON PRINCIPLE BEHIND PERCEPTUAL DEVELOPMENT FOR FACES AND SPOKEN WORDS

While we have outlined a similar developmental trajectory as evidence for a common underlying mechanism, the evidence used to support the neuronal recycling hypothesis (Dehaene, 2005) as a common neurodevelopmental process for any "human cultural ability" is also compatible with our theory. The neuronal recycling hypothesis proposes that any apparently unique and recent cultural ability that humans exhibit must reflect an incremental use of flexibility already present in the brains of our nearest ancestors. The development of "human abilities" is therefore ultimately constrained by genetically controlled factors such as receptor density and connectivity patterns. It is not the case, in this scheme, that any and all regularities can be learned. Only those regularities that the brain is set up to be able to learn are possible. Dehaene and Cohen (2011) argue, for example, that the visual word form area is an example of a common visual area with a suitable basic visual purpose (preference for high resolution foveal shapes and for line images) that can be co-opted in the human brain to undertake reading and face recognition.

To build on the idea of a common underlying mechanism, we speculate that rather than reinvent totally new systems of learning for each perceptual domain, the same basic neurophysiological processes are recycled throughout the brain and "implemented" when the learner engages with a stimulus that requires the kind of fine grained discrimination and perceptual constancy across variable instantiations that spoken word perception and face recognition demand. This results in a developmental trajectory and a final representational structure that appears similar across different domains. It follows that there would be a range of neuroanatomically distinct regions that appear to operate similarly but manipulate different input.

This view, that one fundamental mechanism or set of mechanisms is responsible for the development of all skilled perceptual behaviors, is also supported by the evidence put forward for the co-development of lateralization of printed word and face responsive areas of the brain (Dundas, Plaut, & Behrmann, 2013). Lateralization of function has been shown to be flexible and to adjust to provide a systematic, and compartmentalized, relationship between differing functions. For example, developmental emergence of lateralization of the visual areas of the cortex that represent faces and printed words has been shown to be inter-related, such that face recognition becomes more strongly lateralized to the right hemisphere as printed word recognition develops and becomes lateralized to the left hemisphere (Dundas et al., 2013). This seemingly paradoxical finding suggests that when a cortical region normally devoted to one function comes under competition from a later-developing function, the brain's response is to increase modularization of the two functions, in this case through increasing the lateralization of the two functions to the opposite hemispheres. Although this example of mutual competition only encompasses the visual modality, we speculate that allocation of resources across many areas of the brain is a closely interwoven process of ongoing organization even outside of visual perception. In particular, this type of mutual competition driving apparent modularization of the brain would be crucially important considering the naturally multi-sensory nature of both speech perception/spoken word recognition and recognition of individuals by their faces and voices.

Indeed we can look at skilled perceptual capabilities through the lens of the information extracted rather than the modality of delivery, and in so doing we begin to identify regions of the brain that are clearly not "modal" (not unimodal). For example, Haxby, Hoffman, and Gobbini (2000) outline a proposal for a distributed neural system responsible for face perception. They suggest that there are two "streams" representing the invariant aspects of faces that facilitate identity recognition versus the changeable aspects of faces that facilitate social communication, respectively. These two "separate" aspects of face perception are equally applicable to voice perception. Indeed, the areas of the brain that have been found to be most responsive to the changeable aspects of faces are located in the superior temporal sulcus (Hoffman & Haxby, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998). These aspects of face perception share a neighborhood with the area most responsive to spoken language, in the superior temporal gyrus (Calvert et al., 1997). Moreover, integral to the distributed system for face perception is the inclusion by Haxby et al. (2000) of areas of the brain considered to subserve "non-face" cognitive functions, particularly where the same information can be gleaned from either the voice or the face. As an example, lip reading is found to elicit activity in areas associated with processing auditory speech (Calvert et al., 1997). The explicit inclusion of "non-face" brain regions (particularly auditory language related areas in our case) within the proposed face perception system acknowledges the multi-modally integrated nature of the stimuli that carry social information and the ultimate efficiency of harnessing a distributed yet integrated system to process these stimuli. It is important to also highlight that we need not be restricted to consideration of the integration of "receptive" senses. Proprioceptive information, or the awareness of how our own face moves, may also be important in the development of both face recognition (Sugita, 2009) and speech perception capabilities (Ito, Tiede, & Ostry, 2009; Skipper, van Wassenhove, Nusbaum, & Small, 2007).

To arrive, as adults, at the distributed system subserving skilled perception that is outlined by Haxby et al. (2000), the parsimonious suggestion would be that the same developmental mechanisms are at play across the senses, shaping the sensory brain to ensure independent functioning of each sense, but also integration between senses, according to the statistics of the environmental (and self-generated) input. Following this line of thought, the existence of a basic, generally available learning mechanism would promote the reallocation of an area of the brain to an unusual role in the absence of a stereotypical sensory diet in early development, as happens in cross-modal sensory plasticity with early impairments in hearing or vision (Wong & Bhattacharjee, 2011; see also Shimojo & Shams, 2001).

Finally, additional support is also found in evidence that abilities we once thought made us unique among animals are most likely an adaptation and elaboration of a more general organizational principal, used to supreme effect in spoken language and face recognition. In particular, other animals display statistical learning (Hauser, Newport, & Aslin, 2001). There are similarities on many levels between birdsong and human speech (Fehér, Wang, Saar, Mitra, & Tchernichovski, 2009; Gardner, Naef, & Nottebohm, 2005) and we are not alone in our ability to recognize individuals via the face (Martin-Malivel & Okada, 2007; Peirce et al., 2001).

By outlining the evidence for a common developmental mechanism/mechanisms that may subserve skilled perceptual capabilities across the two domains, hints emerge regarding which key aspects of the sensory input promote successful development of skilled perception of faces and spoken words and phonemes. Our proposal is that structured variation in natural face and speech input, in particular, is crucial to the development of skilled perception.

WHY IS VARIATION IMPORTANT?

No matter what it is we are trying to categorize or discriminate, there will always be variability in the input. Even in the ideal circumstance where the environment is controlled such that the physical input is unchanged (as in laboratory studies), every time we encounter an instance of a word or face, internal noise (e.g., in background-level neural firing) will ensure that there is variability in the internal representation. Natural variability in the input may, at first glance, appear to be particularly challenging for infants. However, although some variability will be random and uninformative, in many cases the variability will be quite systematic (even if it is also quite substantial), particularly when it comes to identification of a specific word across speakers with differing accents or identification of a particular person across changes in pose. For this reason, it is important for the perceptual systems to become familiar with the natural systematic versus random variability within and between the categories of stimuli that are important (see Best, in press; Bruce, 1994; Burton, 2013; Hay & Drager, 2007). Both the face and spoken word and speech perception literatures acknowledge the utility of organizing perceptual spaces based on variability. For example, principal component analysis (PCA: Jolliffe, 2005) methods have been useful in modeling human face recognition abilities in adults (Furl, Phillips, & O'Toole, 2002). PCA creates a face space by describing a set of faces as dimensions ordered according to explained variance. Despite the utility of these models in describing some aspects of adult face recognition, a standard assumption is that the infant must learn about random and systematic variability in order to discount it, helping to establish a normalized and “invariant” representation of the particular thing being identified (see also, Bruce, 1994). That is, it is tacitly assumed that variability of all kinds is a hindrance to recognition and classification, and that it needs to be filtered out or discarded. That view would suggest that it is optimal to initially present an infant with clean (low-variability) data in order to optimize their ability to establish ideal exemplar traces (or to develop clean category prototypes). In that approach, only then should finer scale variability be introduced to flesh out the representation (Papousek, Papousek, & Bornstein, 1985; Snow & Ferguson, 1977).

One important counter example to the “normalization” view is the Perceptual Learning Theory of Eleanor Gibson (1969) (see also J. J. Gibson & E. J. Gibson, 1955) who proposed that rather than establishing prototypes or ideal exemplars of a category, per se, perceptual learning progresses by establishing *dimensions of difference*. That is, perceptual learning essentially involves coming to recognize the contrastive aspects of stimuli, or elaboration as we have mentioned. E. Gibson (1969) also stressed that learning of differences is boosted when distinctive features are emphasized. Rather than needing a clean, normalized prototype for perceptual learning, the Perceptual Learning Theory view recommends that useful differences should be emphasized. When this is translated to multidimensional stimuli like words and faces, in real life the natural input that supports infant learning should be highly variable along a range of dimensions. This will simultaneously enhance differences that should serve perceptual learning across a range of uses. Indeed, data suggest that in development of both face and spoken language perception, the general pattern observed in caregivers’ behavior toward young infants is that it presents a wider range of systematic variability along multiple stimulus dimensions than is seen in adult-adult communication, rather than a reduced range of variability.

VARIABILITY IN INFANT DIRECTED SPOKEN INTERACTIONS

In the language domain, findings on the audible properties of IDS indicate that a number of crucial acoustic properties of IDS are both exaggerated in range and more variable along a number of important dimensions, relative to ADS (see Best, in press). If variability made initial language learning difficult, then social-cognitive and/or evolutionary principles should push parents to reduce phonetic variability when speaking to their infants (cf. Papousek et al., 1985; Snow & Ferguson, 1977). Instead, caregivers and other people are apparently compelled to expand phonetic variation along multiple dimensions when interacting with babies. IDS, as compared to ADS, displays a larger magnitude and range of excursions in pitch (F0) (e.g., Fernald & Simon, 1984; Kitamura & Burnham, 2003; Kitamura, Thanavisuth, Burnham, & Luksaneeyanawin, 2002), in structured temporal variations (e.g., rhythmic alternation, durational contrasts, speaking rate), and in dynamic adjustments of voice amplitude/intensity, which range from loud to modal to whispered (e.g., Fernald & Mazzie, 1991; Fernald et al., 1989; Grieser & Kuhl, 1988; Kitamura & Lam, 2009; Stern, Spieker, Barnett, & MacKain, 1983). These modifications of pitch, timing and amplitude, in turn, impact on linguistic features such as stress patterning and prosodic modulations that reflect both grammatical structures and pragmatic aspects of discourse (e.g., turn-taking). Variability and range in vowel formant frequencies (F2, F1) is also exaggerated in IDS (Burnham, Kitamura, & Vollmer-Conna, 2002; Kuhl et al., 1997), as is variation in more socially relevant acoustic properties such as emotional affect (Slaney & McRoberts, 2003; Trainor, Austin, & Desjardins, 2000). Moreover, babies prefer and attend more to the increased variation of IDS relative to ADS (Cooper & Aslin, 1990; Fernald, 1985; Fernald & Kuhl, 1987; Kitamura & Lam, 2009). Thus, IDS displays an increased range of variation along multiple acoustic dimensions that are relevant to both the linguistic and social aspects of early language acquisition, and that variability appears to capture infants’ attention rather than overwhelming them.

Recent research has provided evidence consistent with our reasoning that increased acoustic variation in speech helps rather than hinders infants’ learning of speech distinctions and spoken words. Infants discriminate vowels better if the stimuli are presented in a variety of pitches than if they are presented in only a single high pitch (Trainor & Desjardins, 2002). Infants of 14 months can learn a novel minimal-pair word distinction (/buk/-/puk/) if the training tokens are produced by multiple speakers, but not if they are produced by just a single speaker (Rost & McMurray, 2009, 2010), and toddlers of 21 months can learn new words presented in IDS but do not show evidence of this with ADS (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011). Moreover, even infants as young as 7.5 months recognize familiarized words better if they were originally presented in IDS than in ADS (Singh, Nestor, Parikh, & Yull, 2009). They also recognize familiarized words better if they were originally presented in multiple emotional affects (happy, sad, neutral) than in a single affect (Singh, Nestor, et al., 2008). Likewise, 7-month-olds are better able to segment familiarized words from sentences if the words were originally presented in IDS than in ADS (Thiessen, Hill, & Saffran, 2005). Infants whose mothers use wider acoustic variations in their vowels in IDS perform better on discriminating native speech contrasts than do infants of mothers who display less variation in their IDS vowels (Liu, Kuhl, & Tsao, 2003). Crucially, infants' speech discrimination performance predicts later word-learning: *better* discrimination of native speech contrasts and *poorer* discrimination of non-native speech contrasts at 7 months both predict larger vocabulary size at 14–30 months (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Tsao, Liu, & Kuhl, 2004). In complement to these observations, studies of the acoustic variations in IDS vowels indicate that caregivers provide systematic distributions of those variations that help to distinguish between relevant vowel contrasts in their language (Werker et al., 2007).

VARIABILITY IN INFANT DIRECTED FACIAL INTERACTIONS

But do these beneficial effects of input variation also apply to the *visible* motions of a speaker's face when she is interacting with infants as compared to adults? Is there also a greater range and more variation in adults' facial motions during infant-directed than adult-directed interactions? If so, does this aid infants' perceptual learning about faces? Although informal observation and general belief would suggest that this is surely the case (e.g., Werker, Pegg, & McLeod, 1994), there has been remarkably little research addressing these questions. What little evidence we could find, nevertheless, does indeed indicate that there is more extensive (systematically increased variation in) speech-related facial motion in infant-directed interactions than in adult-directed ones. Infants also seem attracted to and/or benefit when they are exposed to variable facial patterns, including dynamic variation (e.g., videos rather than still pictures). Infant directed facial speech exhibits more extensive lip movements for vowels than does adult-directed facial speech (Green, Nip, Wilson, Mefferd, & Yunusova, 2010; Kim, Davis, & Kitamura, 2012). Moreover, three-dimensional measures of head, eyebrow, jaw, mouth, and lip motion more generally support the same pattern of greater motion in IDS than ADS (Chong, Werker, Russell, & Carroll, 2003; Kim et al., 2012). Relatedly, caregivers' manual gestures and motions involving objects are more extensive and varied in infant-directed than adult-directed verbal interactions about those objects (Brand, Baldwin, & Ashburn, 2009). And with respect to early development of face perception, research suggests that dynamic stimuli (e.g., videos of facial expressions) benefit infants' ability to discriminate facial emotional expressions (Caron, Caron, & MacLean, 1988; Walker-Andrews, 1986; see also Walker-Andrews, 1997) and may better support their learning and recognition of familiar faces (Cecchini, Baroni, Di Vito, Piccolo, & Lai, 2011; Layton & Rochat, 2007).

Infant sensitivity to time-varying audio-visual correspondences in IDS versus ADS would provide further support for our premise that systematic variability is crucial to perceptual learning. Despite stimulus variability being essentially doubled across the two modalities of audio-visual speech, relative to uni-modal audio speech, multi-modal studies have found that infants' preference for IDS over ADS speech is robust when the speech signal is synchronously presented with the video of the speakers. However, the IDS preference reduces or disappears if synched with a video of the speaker simply nodding or producing ADS (Werker & McLeod, 1989; see also Lewkowicz, 1996). This pattern holds up regardless of whether the audio stimulus materials are native or non-native speech (Werker et al., 1994). Together these results suggest that the dynamic correspondences between the increased variation in each modality of IDS guides the infant's attention to multi-modally informative aspects of speech.

According to the results outlined, increased yet systematic variability is pervasive along a number of stimulus dimensions in infant directed social interaction. The evidence also suggests that stimulus variation is crucial to the development of speech and word perception skills. Indirect evidence leads us to suggest that the same is true of face recognition. Given the task of categorizing or distinguishing faces or words across different instances, we suspect that the increased variability of infant directed interaction is central to supporting the rich categorical knowledge the infant must acquire for faces and spoken words. Moreover, we argue that knowledge of informative variability (i.e., systematically organized) is maintained as crucial information about the complex statistical regularities of the natural input in these two domains. That is to say, the key representational strategy for the perceptual spaces the infant is developing is not the extraction of constrained and clean prototypes or averaged representations for facial or language categories. Rather it is the extraction of the acceptable variance within and relationships between categories (Best, in press).

Complementary evidence supporting the role of variability in achieving skilled recognition comes from a modeling study looking at development of the other race effect (Balas, 2012). The other race effect is the diminished recognition memory we experience for individual faces from races other than those in our own environment (see above for the development of this perceptual narrowing). The study used a Bayesian estimation of recognition performance after a PCA based model had been trained on faces from one race or two. One key aspect of the model is that it was trained using "difference images" rather than images of individuals. A difference image was constructed by calculating the pixel-by-pixel difference between images of individuals. What this means is the model was trained using variability as the input, not using exemplars. As the number of faces that were used to create the training images increased, the face discrimination performance of the model increased, showing that increased variability can support better discrimination performance. The key manipulation was whether the model was trained using difference images created between different races or not. The model trained with difference images created across race boundaries (other race training) developed an ability to discriminate individuals within both the minority and majority race faces. The model trained without cross race difference images produced less discrimination between other race faces. While acknowledging that infants and the model start from different baseline perceptual spaces and with little expectation that we should find this particular model implemented within the brain, the model makes an important contribution: namely that the kind of variability in the training images shapes the performance of the model. Moreover, where increased variability was learned, overall superior performance resulted. We suggest this is also important to consider when looking at the perceptual development of infants in the first year of life.

BILINGUAL DEVELOPMENT: IS INCREASED VARIABILITY HARMFUL?

A key question, then, is whether there can be too much variability. One way to address this is research into a population who experience relatively more variability, such as infants who are born into a bilingual environment. Bilingual-learning infants receive the added variability of regularly encountering two (or more) languages in their input. Thus, it is important to ask whether bilingual-to-be infants' speech perception and spoken word recognition skills are impeded, unaffected, or enhanced—the latter as our proposition would predict—by this increased yet linguistically systematic variation. A recent burgeoning of research on bilingual versus monolingual infants makes it possible to begin answering that question. Depending on the task used, on the abilities tested, and possibly on aspects of their language environments, all three patterns of difference between bilingual and monolingual infants' performance have been reported. Nonetheless, consideration of the full array of findings suggests that bilingual infants are well able to accommodate the extra cross-language variability in their speech input, and even show benefits in speech perception and in certain non-linguistic perceptual-cognitive skills, relative to their monolingual peers. For example, newborns of bilingual mothers show listening preferences for *both* of her languages as compared to non-native languages, and can also discriminate between the two maternal languages (Byers-Heinlein et al., 2010) even if they are from the same rhythmic class, unlike monolingual-to-be newborns. As for attunement to the phoneme contrasts of their two native languages, early findings indicated a modest temporary decline around 8 months in bilingual infants' audio-only discrimination of speech contrasts in both their native languages, in contrast to the good native speech contrast discrimination observed in their monolingual peers of each language. However, this decline was temporary. The bilinguals regained good discrimination of contrasts in both of their languages within the following month or two, by 10–14 months of age (e.g., Bosch & Sebastián-Gallés, 2003).

Importantly, though, more recent evidence from a range of studies indicates that if more sensitive testing techniques are used, bilingual infants *do* discriminate the contrasts of both languages across the full age range including the 8- to 12-month period (Albareda-Castellot, Pons, & Sebastián-Gallés, 2011; Burns, Yoshida, Hill, & Werker, 2007), as well as outperforming monolingual infants in discriminating consonant differences between their languages (Sundara, Polka, & Molnar, 2008). They also outperform monolingual peers at 8 months in discriminating between their two languages when presented with only silent-video talking faces (Weikum et al., 2007). Even more important for our hypothesis about the role of systematic variation in perceptual attunement, 8-month-old bilingual infants may show additional benefits over monolinguals even for discrimination of unfamiliar non-native speech contrasts, whether the speech is presented in audio (Petitto et al., 2012) or visual-only form (Sebastiàn-Gallès, Albareda, Weikum, & Werker, 2012).

And beyond the influence of bilingual exposure on speech and word perception, several recent studies reveal cognitive benefits as well. Bilingual 7-month-olds outperform their monolingual peers in non-language tasks that involve learning multiple rules (Frank et al., 2009; Kovács & Mehler, 2009; Kovács, Mehler, & Carey, 2009) or require delayed recall of a series of actions when they must generalize across multiple dimensions of stimulus variation (Brito & Barr, 2014). Altogether, the findings on bilingual-experienced infants support the idea that they not only can and do sort out, but in fact take advantage of, the multi-lingual (and multi-dimensional) variation they are exposed to. They apply that knowledge both to recognition of speech contrasts in their two native languages as well as in unfamiliar languages. Moreover, their ability to detect systematicity in variation along multiple stimulus dimensions extends even beyond spoken language, supporting their ability to categorize and remember multiple dimensions of variation across non-linguistic events and objects.

HOW IS VARIABILITY USED IN DEVELOPMENT?

As outlined above, it appears that enhanced yet systematic variation is important to the development of the perceptual spaces representing faces and spoken words. The suggestion from the above reviews of the developmental trajectory of face and spoken word and phoneme perception is that the first stage of development involves establishing the basic/foundational dimensions of the sensory space specific to an individual infant's sensory environment. This is proposed to be created via the interaction of young infants' early perceptual biases to attend to certain types of stimuli and statistical learning. Perceptual biases would act to constrain the general focus of statistical learning to stimuli most relevant for development of a robust perceptual space. We have discussed some "static" face- and language-specific perceptual biases in infants during the first few months of life (e.g., visual objects that have a face-like energy profile or a top heavy arrangement of elements) and there are, additionally, indications in both the speech perception/word recognition and face perception literature that statistical properties of the input are important in the subsequent developmental organization of the child's internal word space and face space (O'Toole, Abdi, Deffenbacher, & Valentin, 1993; Saffran, Aslin, & Newport, 1996; see also, Gervain & Mehler, 2010; O'Toole, 2011). Beyond this stage it is suggested that there is a progression from statistical learning to a more domain specific, referent-based, abstract and socially or culturally influenced learning in older infants, as discussed within the language acquisition literature (Gervain & Mehler, 2010).

The key question, then, is: Which aspects of variability are most important to supporting optimal development of an infants' perceptual space? It would have to be admitted that what is important likely changes with developmental progression as well as situationally, depending on the motivations of the infant and perhaps even the complexity of the environment itself. To answer this question, though, we suggest that the aspects of the sensory environment that are most beneficial to infants can be revealed by observing which aspects they preferentially interact with and which aspects they actively disengage from. Just as we can observe that infants will engage preferentially with a static object that contains properties that make it more face like, we can observe the dynamics of a certain set of stimuli that will preferentially engage an infant's attention. There is a hypothesis, endearingly termed "the Goldilocks effect" (Kidd, Piantadosi, & Aslin, 2012), that infants will engage most with stimuli that are optimal in complexity for their developmental stage, with stimuli outside of the optimal complexity range either failing to attract attention or simply eluding the cognitive or sensory capacity of an infant (what we might call a "dynamic bias"). For example, when presented with checkerboard stimuli with different numbers of checks, infants at 3, 8, and 14 weeks will, respectively, look preferentially at increasingly complex checkerboards (Brennan, Ames, & Moore, 1966). Importantly, with these stimuli the preference is thought to not be linked to acuity or accommodation ability (i.e., the infant is able to resolve the more complex but not preferred stimuli). Rather, their preference is proposed to be due to the level of complexity in the visual stimulus itself, likely related to a more nuanced set of developmental factors within the perceptual system itself. Newborns have also been shown to recognize sequences of pairs of objects when the sequence consisted of only two pairs presented in a random order, but they were not found to recognize the same pairs when the sequences were increased to three pairs of objects (Bulf et al., 2011). In this study the sequences consisted of pairs of simple shapes, such as triangles and squares, which were always presented together in the same order; however, the pairs themselves could be presented in any order relative to each other. Learning was tested using a habituation procedure, which revealed that infants look preferentially at a post-habituation violation of the statistics of the sequence of the objects. This shows that newborns only several days old can learn information about the probabilities of occurrence of related visual stimuli. However, this learning was shown to be confined to two pairs of objects without being apparent when three pairs were presented. Learning, and therefore meaningful engagement with a complex statistical sequence, appears to be constrained by cognitive capacity, suggesting a natural engagement with an optimal level of variability for a newborn.

At 5 months of age, infants have been shown to look longer at a random sequence of looming objects than at a sequence composed of repeating pairs of stimuli (a sequence newborns can already learn). They were also found to disengage attention to sequences structured into pairs or triplets mainly at points in the sequence where the transition between shapes is locally repetitive rather than according to the global sequence pattern (Addyman & Mareschal, 2013). This suggests that at 5 months of age infants' looking preferences remain attuned to a certain level of complexity in the stimulation, given that they disengage from a locally repetitive sequence. As the local rather than global repetition appears to govern disengagement, however, it is likely that an infant's mode of engagement with a stimulus containing many types of complexity is also constrained by their current cognitive capacity.

At 8 months of age, Kidd et al. (2012) have measured infants' likelihood of looking away from a visual scene composed of objects appearing from behind occluders with varying probabilities. They found that disengagement was related to the level of information contained in the scene. Infants engaged with an intermediate level of complexity; too much or too little and the infant was shown to disengage from the scene.

Despite showing clear signs of developing an experience constrained perceptual space, even at 11 months of age infants have been shown to combine predictive cues of different strengths in a straightforward fashion to learn a regular pattern. This is unlike the adult participants in this study, who combined these cues in a way that is less than optimal, apparently favoring an overly complex interpretation of the pattern of the cues (Yurovsky, Boyer, Smith, & Yu, 2013). The pattern suggests that even at 11 months infants demonstrate different cognitive capabilities or strategies than adults. This will have a significant impact on the shaping of their developing perceptual spaces, bestowing an advantage in spite of environmental stimulation that appears suboptimal for an adult.

Despite the influence of "The Goldilocks Effect," if we were to construct an optimal artificial environment for a developing infant, to truly establish what is important to provide at each age, then different aspects of the possible stimuli should be pitted against each other, moreso than investigating what an infant is able to perceive in an isolated cue situation. Using an artificial language created to present infants with a set of useable cues, Saffran and Thiessen (2003) found that at 7.5–8 months of age infants follow statistical information in the form of transitional probabilities of syllables within and between words, rather than relying on stress cues that indicate the start of a word. In contrast, at ~9 months of age infants were shown to use stress cues over statistical information (Johnson & Jusczyk, 2001). This would imply that if we were to construct an optimal artificial language to aid perception, for infants at 7 months it should focus on conveying the statistical probabilities of the language they are learning, perhaps with exaggeration of permissible transitional probabilities conveyed via repetitions of a range of common, key transitions. By 9 months, however, it would seem that provision of an enhanced range of the stress patterns of words would be more important.

In the face recognition literature, although it is difficult to enhance variability in the same way when it comes to identifying individuals, while young infants appear to be sensitive mainly to the individual features of a face they should be best served by seeing many individuals with many different features. From ~8 months of age, when they become more sensitive to the configuration of a face (Ferguson et al., 2009) they should then be exposed to many faces with different configurations. While this may seem impossible to manipulate naturally (by one person) in the same way as spoken words can be manipulated, it should be acknowledged that at a basic level, a dynamic face presents an infant with a set of continuously changing (yet constrained) features and configurations as the face moves and creates new facial expressions. Beyond this, however, it may be that the exposure infants of this age experience to multiple faces of differing shapes, and multiple speakers with differing voice qualities or accents, may be optimal for them to learn the critical properties of face configurations and words.

We should point out that this is not to say we should create an artificial environment that would be optimal for infants. As we have seen, infant directed interactions already naturally provide an optimal environment to infants, which appears to be best suited to infant perceptual development. Moreover, infants themselves appear to tailor their interactions, engaging primarily with aspects of the world that appear to suit their stage of development best. Further studies investigating perceptual constancies pitted directly against discrimination ability within the same sets of stimuli should elaborate the cues that are most important at each age across the many aspects of face and spoken language/word perception.

NEW DIRECTIONS FOR EXPERIMENTATION

The proposals here suggest several new directions in experimentation. In particular, several areas have been highlighted where questions that are common and central in one domain can be translated meaningfully into the other. The new directions can be grouped into studies of the importance of variability in faces and speech during infant directed interaction, studies into perceptual constancy, and studies into perception among multiple dimensions of stimulus variation.

The importance of variation in stimuli for infant learning has been highlighted in studies addressing how infants acquire speech perception and spoken word recognition skills, particularly with respect to the increased variability in infant-directed speech input. The further development of cross-language speech perception in bilingual-learning infants is a particularly useful context to investigate the principles we have been discussing. We propose that structured variability is crucial to development of skilled perceptual capabilities. This implies that across the environmental input there must be statistical regularities that structure the increased variability. Within the population of bilingual-learning infants there likely exists great diversity in the structure of infant's interactions with both languages. Investigating bilingual infants' interactions in their two languages could inform us further about which aspects of the structure of variability are important. For example, infants learning two quite similar languages may develop some capabilities at a younger age if the languages are separated by some other context, such as the identity of the speaker (e.g., mum speaks French and dad speaks English). Similarly, to the extent that infants are regularly interacting with adults who mix languages within utterances, we may discover the important limits of an infant's ability to thrive on variability.

Conversely, there is also a natural circumstance whereby an infant receives reduced variation in IDS. It has been shown that mothers with depression tend to exhibit flat vocal affect (Bettes, 1988) and their IDS contains significantly less modulation in fundamental frequency (Kaplan, Bachorowski, Smoski, & Zinser, 2001). A careful study of learning of speech distinctions and spoken words by infants with a mother with depression should also highlight the crucial aspects of increased variability in IDS.

As yet there have been few studies addressing the importance of increased variability in face recognition. According to the hypothesis that face and language recognition are mediated by the same underlying principle, we would predict that just as in IDS, caregivers' facial motion is more exaggerated when interacting with infants than when interacting with adults in the context of all interactions, not just those outlined above (Brand et al., 2009; Chong et al., 2003; Green et al., 2010; Kim et al., 2012). This exaggeration should be found to be preferred by infants, and also to be crucial to an infant's development of face perception capabilities, even to recognition of individual identities. For example, further studies pitting recognition of individual faces displaying adult directed facial expressions (ADFE) versus infant directed facial expressions (IDFE) should demonstrate that IDFE support a range of judgments over and above ADFE. IDFE may also provide advantages in discrimination of stimuli that are currently thought to be undifferentiated, such as attention to and memory for internal features of a face earlier in infancy.

Additionally, we have outlined evidence that suggests that infants actively engage with stimuli of just the right complexity for their developmental stage (Addyman & Mareschal, 2013; Bulf et al., 2011; Kidd et al., 2012). As infant and caregiver interaction is a two way process (see Ainsworth, 1979; Bowlby, 1958), this might suggest that infants are able to influence a parents' infant directed interaction to subtly adjust which aspects of the social stimulus (face or voice) are being enhanced. We would predict that careful measurement of the aspects of caregivers' facial and vocal interactions with infants at each age should reveal changes that will reflect the dimensions of the environmental space that infants are most sensitive to at each stage of perceptual development.

To fully understand the transition from purely statistical learning to more nuanced elaboration of the perceptual space capable of supporting complex perceptual constancy judgments, investigation is needed into perceptual constancy in face recognition in particular, but also spoken phoneme and word perception. We would predict that more complex constancies unfold as an infant moves from a statistical learning regimen to the more elaborated and domain-specific referents regimen. The constancy judgments infants are able to achieve should therefore be influenced by experience. As such, we would predict that there may also be a perceptual narrowing found for constancy judgments. Infants past 6 months of age should display perceptual constancy only within experienced classes of stimuli. For example, recognition of individuals across viewpoints should decline for faces within categories that are not experienced, such as other race faces. Similarly, recognition of vowels across speakers should also decline when the vowel is from a non-native language and does not occur in the native language. By building variability into the stimulus design of experiments we can understand the developmental trajectory of not only discrimination but also perceptual constancy.

To understand which particular dimensions of the perceptual space are being formed at each developmental stage we propose that it is important to construct stimuli that are able to pit different aspects of the environmental space against each other directly. For example, by constructing an artificial diet of faces that vary to differing extents along dimensions suggested by using an image description system such as PCA, it may be possible to elucidate the aspects of faces that infants are most sensitive to at each stage of perceptual development. Faces are particularly challenging to study, as the units of a face that are nameable (e.g., eyes and nose) do not necessarily correspond to the perceptual units important for perception of face identity. Therefore, it will be extremely important to use statistical and modeling techniques that move away from a language-based face specification manipulation to create stimuli.

CONCLUSION

In summary, we have proposed that the development of perceptual skill in face and spoken word recognition follow a similar trajectory due to a set of common, centrally important learning mechanisms. These mechanisms capitalize on the structured variability in infant directed communication, which supports development of the perceptual spaces representing the organization of the sensory information relating to all aspects of faces and spoken words. The perceptual space is initially elaborated according to physical statistical properties in the infant's environment that lead to the apparent narrowing in discrimination of non-experienced stimuli during the first year of life for both faces and spoken words. Thereafter, a combined environmentally, socially and culturally driven strategy supports the development of elaborated representations that can produce more sophisticated perceptual constancies. We anticipate that these domain-general developmental mechanisms and the subsequent description of the perceptual space would apply to any stimulus for which we become such exquisitely tuned skilled perceivers so early in life. However, the perception of spoken words and faces occupies such an important role in infant development as well as mature perceptual functioning, that these may be the domains in which this is seen most clearly.

REFERENCES

Addyman, C., & Mareschal, D. (2013). Local redundancy governs infants' spontaneous orienting to visual-temporal sequences. Child Development, 84(4), 1137–1144.

Ainsworth, M. D. (1979). Infant–mother attachment. The American Psychologist, 34(10), 932–937. doi: 10.1037/0003-066x.34.10.932

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. Developmental Science, 14(2), 395–401. doi: 10.1111/j.1467-7687.2010.00989.x

Anzures, G., Quinn, P.C., Pascalis, O., Slater, A. M., & Lee, K. (2010). Categorization, categorical perception, and asymmetry in infants' representation of face race. Developmental science, 13(4), 553–564. doi: 10.1111/j.1467-7687.2009.00900.x

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. doi: 10.1016/j.jecp.2012.04.005

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. Yeni Komshian, J. F. Kavanagh, & C. A. Ferguson, (Eds.), Child phonology: Perception and production (pp. 67–96). New York: Academic Press.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. Child Development, 52(4), 1135–1145. doi: 10.2307/1129499

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), Multisensory development (pp. 183–205). Oxford, England: Oxford University Press.

Balas, B. (2012). Bayesian face recognition and perceptual narrowing in face-space. Developmental Science, 15(4), 579–588. doi: 10.1111/j.1467-7687.2012.01154.x

Best, C. T. (1994). Learning to perceive the sound pattern of English. In C. Rovee-Collier & L. P. Lipsitt (Eds.), Advances in infancy research (Vol. 9, pp. 217–304). Norwood, NJ: Norwood.

Best, C. T. (in press). Devil or angel in the details? Complementary principles of phonetic variation provide the key to phonological structure. In J. Romero & M. Riera (Eds.), Sounds, representations and methodologies: Essays on the phonetics-phonology interface. Current issues in linguistic theory. Amsterdam: John Benjamins.

Best, C. T., Kroos, C., & Irwin, J. (2010). *I can see what you said: Infant sensitivity to articulator congruency between audio-only and silent-video presentations of native and nonnative consonants*. Paper presented at the Auditory-Visual Speech Processing 2010.

Best, C. T., Kroos, C., & Irwin, J. (2011). *Do infants detect AV articulator congruency for non-native click consonants?* Paper presented at the Auditory-Visual Speech Processing 2011.

Best C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. Language and Speech, 46(2–3), 183–216.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. Infant Behavior and Development, 18(3), 339–350.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. Journal of Experimental Psychology. Human Perception and Performance, 14(3), 345–360. doi: 10.1037/0096-1523.14.3.345

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. Psychological Science, 20(5), 539–542. doi: 10.1111/j.1467-9280.2009.02327.x

Bettes, B. (1988). Maternal depression and motherese: Temporal and intonational features. Child Development, 59, 1089–1096.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. Language and Speech, 46(2–3), 217–243.

Bowlby, J. (1958). The nature of the child's tie to his mother. International Journal of Psycho-Analysis, 39, 350–373.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2009). Evidence for 'motionese': Modifications in mothers' infant-directed action. Developmental Science, 5(1), 72–83.

Brennan, W. M., Ames, E. W., & Moore, R. W. (1966). Age differences in infants' attention to patterns of different complexities. Science, 151(3708), 354–356. doi: 10.1126/science.151.3708.354

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., & Baillet, S. (2009). Hearing faces: How the infant brain matches the face it sees with the speech it hears. Journal of Cognitive Neuroscience, 21(5), 905–921.

Brito, N., & Barr, R. (2014). Flexible memory retrieval in bilingual 6-month-old infants. Developmental psychobiology, 56(5), 1156–1163. doi: 10.1002/dev.21188.

Brookes, H., Slater, A., Quinn, P. C., Lewkowicz, D. J., Hayes, R., & Brown, E. (2001). Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. Infant and Child Development, 10(1–2), 75–82. doi: 10.1002/icd.249

Bruce, V. (1994). Stability from variation: The case of face recognition. The MD Vernon memorial lecture. The Quarterly Journal of Experimental Psychology, 47(1), 5–28.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. Cognition, 121(1), 127–132. doi: 10.1016/j.cognition.2011.06.010

Bulf, H., & Turati, C. (2010). The role of rigid motion in newborns' face recognition. Visual Cognition, 18(4), 504–512. doi: 10.1080/13506280903272037

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. Science, 296(5572), 1435–1435. doi: 10.1126/science.1069587

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. Applied Psycholinguistics, 28(3), 455–474. doi: 10.1017/s0142716407070257

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. Quarterly Journal of Experimental Psychology, 66(8), 1467–1485.

Bushnell, I. (2001). Mother's face recognition in newborn infants: Learning and memory. Infant and Child Development, 10(1–2), 67–74.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. Psychological Science, 21(3), 343–348. doi: 10.1177/0956797609360758

Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? Psychological Science, 15, 379–383. doi: 10.1111/j.0956-7976.2004.00688.x

Cassia, V. M., Valenza, E., Simion, F., & Leo, I. (2008). Congruency as a nonspecific perceptual property contributing to newborns face preference. Child Development, 79, 807–820. doi: 10.1111/j.1467-8624.2008.01160.x

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., & McGuire, P. K. (1997). Activation of auditory cortex during silent lipreading. Science, 276(5312), 593–596. doi: 10.1126/science.276.5312.593

Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional expressions: The role of face and voice. Child Development, 59(3), 604–616. doi: 10.2307/1130560

Cecchini, M., Baroni, E., Di Vito, C., Piccolo, F., & Lai, C. (2011). Newborn preference for a new face vs. a previously seen communicative or motionless face. Infant Behavior & Development, 34(3), 424–433. doi: 10.1016/j.infbeh.2011.04.002

Chomsky, N. (2006). Language and mind. Cambridge, UK: Cambridge University Press.

Chong, S., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. Infant and Child Development, 12(3), 211–232.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. The Journal of the Acoustical Society of America, 95, 1570–1580.

Coltheart, M. (1999). Modularity and cognition. Trends in Cognitive Sciences, 3(3), 115–120.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. Child Development, 61(5), 1584–1595. doi: 10.1111/j.1467-8624.1990.tb02885.x

Coulon, M., Guellai, B., & Streri, A. (2011). Recognition of unfamiliar talking faces at birth. International Journal of Behavioral Development, 35(3), 282–287. doi: 10.1177/0165025410396765

Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. Journal of Phonetics, 39(3), 388–402. doi: 10.1016/j.wocn.2011.02.004

de Haan, M., Johnson, M. H., Maurer, D., & Perrett, D. I. (2001). Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. Cognitive Development, 16(2), 659–678. doi: 10.1016/s0885-2014(01)00051-x

de Heering, A., Turati, C., Rossion, B., Bulf, H., Goffaux, V., & Simion, F. (2008). Newborns' face recognition is based on spatial frequencies below 0.5 cycles per degree. Cognition, 106(1), 444–454. doi: 10.1016/j.cognition.2006.12.012

de Schonen, S., & Mathivet, E. (1989). First come, first served: A scenario about the development of hemispheric specialization in face recognition during infancy. European Bulletin of Cognitive Psychology, 9, 3–44.

Deacon, T. W. (1997). The symbolic species: The co-evolution of language and the brain. New York: WW Norton & Company.

Dehaene, S. (2005). From monkey brain to human brain: A Fyssen Foundation symposium. Cambridge, MA: MIT Press.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. Trends in Cognitive Sciences, 15(6), 254–262. doi: 10.1016/j.tics.2011.04.003

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. Science, 298(5600), 2013–2015.

Dundas, E. M., Plaut, D. C., & Behrmann, M. (2013). The joint development of hemispheric lateralization for words and faces. Journal of Experimental Psychology. General, 142(2), 348–358.

Eilers, R. E., Gavin, W., & Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: A cross-linguistic study. Child Development, 50(1), 14–18. doi: 10.1111/j.1467-8624.1979.tb02973.x

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. Science (New York, N.Y.), 171(3968), 303–306. doi: 10.1126/science.171.3968.303

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. Proceedings of the National Academy of Sciences of the United States of America, 99, 9602–9605. doi: 10.1073/pnas.152159999

Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. Nature, 459(7246), 564–568.

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. Child Development, 81(5), 1376–1383. doi: 10.1111/j.1467-8624.2010.01479.x

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. Language and Speech, 46(Pt 2–3), 245–264. doi: 10.1177/00238309030460020901

Ferguson, K. T., Kulkofsky, S., Cashon, C. H., & Casasola, M. (2009). The development of specialized processing of own-race faces in infancy. Infancy, 14(3), 263–284. doi: 10.1080/15250000902839369

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. Infant Behavior and Development, 8(2), 181–195. doi: 10.1016/s0163-6383(85)80005-9

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. Infant Behavior and Development, 10(3), 279–293. doi: 10.1016/0163-6383(87)90017-8

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. Developmental Psychology, 27(2), 209–221. doi: 10.1037/0012-1649.27.2.209

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. Developmental Psychology, 20(1), 104–113. doi: 10.1037/0012-1649.20.1.104

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. Journal of Child Language, 16(3), 477–501. doi: 10.1017/s0305000900010679

Fodor, J. A. (1983). The modularity of mind. Cambridge, Mass: MIT Press.

Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. Developmental Science, 12(4), 504–509. doi: 10.1111/j.1467-7687.2008.00794.x

Furl, N., Phillips, P. J., & O'Toole, A. J. (2002). Face recognition algorithms and the other-race effect: Computational mechanisms for a developmental contact hypothesis. Cognitive Science, 26(6), 797–815. doi: 10.1207/s15516709cog2606_4

Gardner, T. J., Naef, F., & Nottebohm, F. (2005). Freedom and rules: The acquisition and reprogramming of a bird's learned song. Science, 308(5724), 1046–1049.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. Annual Review of Psychology, 61(1), 191–218. doi: 10.1146/annurev.psych.093008.100408

Gibson, E. J. (1969). Principles of perceptual learning and development. Englewood Cliffs, NJ: Prentice-Hall.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning; differentiation or enrichment? Psychological Review, 62(1), 32–41. doi: 10.1037/h0048826

Gottlieb, G. (1971). Ontogenesis of sensory function in birds and mammals. In E. Tobach, L. R. Aronson, & E. Shaw (Eds.), The biopsychology of development (pp. 67–128). New York: Academic Press.

Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. Journal of Speech, Language and Hearing Research, 53(6), 1529–1542.

Grieser, D. A. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. Developmental Psychology, 24(1), 14–20.

Guellaï, B., Coulon, M., & Streri, A. (2011). The role of motion and speech in face recognition at birth. Visual Cognition, 19(9), 1212–1233.

Hainline, L. (1978). Developmental changes in visual scanning of face and nonface patterns by infants. Journal of Experimental Child Psychology, 25(1), 90–115. doi: 10.1016/0022-0965(78)90041-3

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. Science, 198(4319), 853–855. doi: 10.1126/science.918670

Hallé, P. A., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. Infant Behavior and Development, 17(2), 119–129. doi: 10.1016/0163-6383(94)90047-7

Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. Infant Behavior and Development, 19(4), 463–481. doi: 10.1016/s0163-6383(96)90007-7

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. Cognition, 78(3), B53–B64.

Haviland, J. M., & Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. Developmental Psychology, 23(1), 97–104. doi: 10.1037/0012-1649.23.1.97

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. Trends in Cognitive Sciences, 4(6), 223–233. doi: 10.1016/s1364-6613(00)01482-0

Hay, J., & Drager, K. (2007). Sociophonetics. Annual Review of Anthropology, 36, 89–103.

Heron-Delaney, M., Anzures, G., Herbert, J. S., Quinn, P. C., Slater, A. M., & Tanaka, J. W. (2011). Perceptual training prevents the emergence of the other race effect during infancy. PLoS ONE, 6(5), e19858. doi: 10.1371/journal.pone.0019858

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nature Neuroscience, 3(1), 80–84. doi: 10.1038/71152

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. Child Development, 76(3), 598–613. doi: 10.1111/j.1467-8624.2005.00866.x

Humphreys, K., & Johnson, M. H. (2007). The development of "face-space" in infancy. Visual Cognition, 15(5), 578–598. doi: 10.1080/13506280600943518

Ito, T., Tiede, M., & Ostry, D. (2009). Somatosensory function in speech perception. Proceedings of the National Academy of Sciences of the United States of America, 106, 1245–1248. doi: 10.1073/pnas.0810063106

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. Journal of Memory and Language, 44(4), 548–548. doi: 10.1006/jmla.2000.2755

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition, 40(1), 1–19. doi: 10.1016/0010-0277(91)90045-6

Jolliffe, I. (2005). Principal component analysis (2nd ed.). New York, NY: Springer.

Kaplan, P. S., Bachorowski, J. A., Smoski, M. J., & Zinser, M. (2001). Role of clinical diagnosis and medication use in effects of maternal depression on infant-directed speech. Infancy, 2, 537–548. doi: 10.1207/S15327078IN0204_08

Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., & Lee, K. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old Infants. Infancy, 11(1), 87–95.

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., & Slater, A. M. (2009). Development of the other-race effect during infancy: Evidence toward universality? Journal of Experimental Child Psychology, 104(1), 105–114. doi: 10.1016/j.jecp.2009.01.006

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(12), 1084–1089. doi: 10.1111/j.1467-9280.2007.02029.x

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., & Smith, M. (2005). Three-month-olds, but not newborns, prefer own-race faces. Developmental Science, 8(6), F31–F36. doi: 10.1111/j.1467-7687.2005.0434a.x

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. PLoS ONE, 7(5), e36399. doi: 10.1371/journal.pone.0036399

Kim, J., Davis, C., & Kitamura, C. (2012). *Auditory-visual speech to infants and adults: Signals and correlations*. Paper presented at the INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association, Portland, Oregon, USA.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. Infancy, 4(1), 85–110.

Kitamura, C., & Lam, C. (2009). Age-specific preferences for infant-directed affective intent. Infancy, 14(1), 77–100.

Kitamura, C., Thanavisuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universal pitch modifications in infant directed speech: A prelinguistic longitudinal study in a tonal and non-tonal language. Infant Behavior and Development, 24(4), 372–392.

Kleiner, K. A. (1987). Amplitude and phase spectra as indices of infants' pattern preferences. Infant Behavior and

Development 10, 49–59. doi: 10.1016/0163-6383(87)90006-3

Kleiner, K. A., & Banks, M. S. (1987). Stimulus energy does not account for 2-month-olds' face preferences. Journal of Experimental Psychology. Human Perception and Performance, 13(4), 594–600. doi: 10.1037/0096-1523.13.4.594

Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. Science, 325(5940), 611–612. doi: 10.1126/science.1173947

Kovács, Á. M., Mehler, J., & Carey, S. E. (2009). Cognitive gains in 7-month-old bilingual infants. Proceedings of the National Academy of Sciences of the United States of America, 106(16), 6556–6560. doi: 10.1073/pnas.0811323106

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66(6), 1668–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. Infant Behavior and Development, 6(2), 263–285. doi: 10.1016/s0163-6383(83)80036-8

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., & Ryskina, V. L. (1997). Cross-language analysis of phonetic units in language addressed to infants. Science, 277(5326), 684–686.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "Critical Period". Language Learning and Development, 1(3–4), 237–264.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. Science, 218(4577), 1138–1141. doi: 10.1126/science.7146899

Kuhl, P. K., & Meltzoff, A. N. (1984). Intermodal speech perception. Infant Behavior and Development, 7, 361–381.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental Science, 9(2), F13–F21. doi: 10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. Science, 255(5044), 606–608. doi: 10.1126/science.1736364

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(3), 829–840. doi: 10.1037/0096-1523.17.3.829

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. Psychological Science, 1(2), 115–121. doi: 10.1111/j.1467-9280.1990.tb00079.x

Layton, D., & Rochat, P. (2007). Contribution of motion information to maternal face discrimination in infancy. Infancy, 12(3), 257–271.

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. Infant Behavior and Development, 13(3), 343–354. doi: 10.1016/0163-6383(90)90039-b

Lewkowicz, D. J. (1996). Infants' response to the audible and visible properties of the human face: Role of lexical-syntactic content, temporal synchrony, gender, and manner of speech. Developmental Psychology, 32(2), 347–366. doi: 10.1037/0012-1649.32.2.347

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. Proceedings of the National Academy of Sciences of the United States of America, 103(17), 6771–6774. doi: 10.1073/pnas.0602027103

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. Trends in Cognitive Sciences, 13(11), 470–478. doi: 10.1016/j.tics.2009.08.004

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match nonhuman primate faces and voices. Infancy, 15(1), 46–60.

Lewkowicz, D. J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. International Journal of Behavioral Development, 37(2), 90–94. doi: 10.1177/0165025412467582

Lewkowicz, D. J., Sowinski, R., & Place, S. (2008). The decline of cross-species intersensory perception in human infants: Underlying mechanisms and its developmental persistence. Brain Research, 1242(Journal Article), 291–302. doi: 10.1016/j.brainres.2008.03.084

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. Developmental Psychology, 16(6), 597–607. doi: 10.1037/0012-1649.16.6.597

Lickliter, R. (1993). Timing and the development of perinatal perceptual organization. In G. Turkewitz & D. A. Devenny (Eds.), Developmental time and timing (pp. 105–123). Hillsdale, NJ: Lawrence Erlbaum Associates.

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. Developmental Science, 6(3), F1–F10.

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. Language Learning and Development, 7(3), 185–201.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. Science, 219(4590), 1347–1349. doi: 10.1126/science.6828865

Martin-Malivel, J., & Okada, K. (2007). Human and chimpanzee face recognition in chimpanzees (*Pan troglodytes*): role of exposure and impact on categorical perception. Behavioral Neuroscience, 121(6), 1145–1155. doi: 10.1037/0735-7044.121.6.1145

Maurer, D. (1993). Neonatal synesthesia: Implications for the processing of speech and faces. In B. de Boysson-Bardies S. de Schonen P. Jusczyk P. McNeilage & J. Morton (Eds.), Developmental neurocognition: Speech and face processing in the first year of life (pp. 109–124). Doetinchem, Netherlands: Springer.

Maurer, D., & Mondloch, C. (2004). Neonatal synesthesia: A re-evaluation. In L. C. Robertson & N. Sagiv (Eds.), Synesthesia: Perspectives from cognitive neuroscience (pp. 193–123). Oxford, UK: Oxford University Press.

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. Developmental Psychobiology, 56, 154–178. doi: 10.1002/dev.21177

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. Perception, 7(5), 491–497.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. Cognition, 29(2), 143–178. doi: 10.1016/0010-0277(88)90035-2

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. Infant Behavior and Development, 16(4), 495–500. doi: 10.1016/0163-6383(93)80007-u

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. Psychological Review, 98(2), 164–181.

Mulak, K. E., & Best, C. T. (2013). Development of word recognition across speakers and accents. In L. Gogate & G. Hollich (Eds.), Theoretical and computational models of word learning: Trends in psychology and artificial intelligence (pp. 242–269). Hershey, PA: IGI Global-Robotics.

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. Child Development, 84(6), 2064–2078. doi: 10.1111/cdev.12087

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. Developmental Science, 13(3), 407–420. doi: 10.1111/j.1467-7687.2009.00898.x

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. Journal of Memory and Language, 43(1), 1–19. doi: 10.1006/jmla.2000.2698

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. Speech Communication, 41(1), 233–243. doi: 10.1016/s0167-6393(02)00106-1

O'Toole, A. J. (2011). Cognitive and computational approaches to face recognition. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), Oxford handbook of face perception (pp. 15–30). Oxford: Oxford University Press.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. Journal of the Optical Society of America A. Optics, Image Science, and Vision, 10(3), 405–411.

Papousek, M., Papousek, H., & Bornstein, M. (1985). The naturalistic vocal environment of young infants: On the significance of homogeneity and variability in parental speech. In T. Field & N. Fox (Eds.), Social perception in infants (pp. 269–297). Norwood, NJ: Ablex.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? Science, 296(5571), 1321–1323.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., & Coleman, M. (2005). Plasticity of face processing in infancy. Proceedings of the National Academy of Sciences of the United States of America, 102(14), 5297–5300. doi: 10.1073/pnas.0406627102

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. Developmental Science, 6(2), 191–196. doi: 10.1111/1467-7687.00271

Peirce, J. W., Leigh, A. E., daCosta, A. P. C., & Kendrick, K. M. (2001). Human face recognition in sheep: Lack of configurational coding and right hemisphere advantage. Behavioural Processes, 55(1), 13–26. doi: 10.1016/s0376-6357(01)00158-9

Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. Journal of Neuroscience, 32, 11159–11163. doi: 10.1523/JNEUROSCI6516-11.2012

Petitto, L. A., Berens, M. S., Kovelman, I., Dubins, M. H., Jasinska, K., & Shalinsky, M. (2012). The "Perceptual Wedge Hypothesis" as the basis for bilingual babies' phonetic processing advantage: New insights from fNIRS brain imaging. Brain and Language, 121(2), 130–143. doi: 10.1016/j.bandl.2011.05.003

Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. Journal of the Acoustical Society of America, 100(1), 577–592.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. Speech Communication, 41(1), 221–231. doi: 10.1016/s0167-6393(02)00105-x

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/th/ perception: Evidence for a new developmental pattern. The Journal of the Acoustical Society of America, 109(5 Pt 1), 2190–2201.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20(2), 421–435. doi: 10.1037/0096-1523.20.2.421

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. Proceedings of the National Academy of Sciences of the United States of America, 106(26), 10598–10602. doi: 10.1073/pnas.0904134106

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. Journal of Neuroscience, 18(6), 2188–2199.

Quinn, P. C., Uttley, L., Lee, K., Gibson, A., Smith, M., & Slater, A. M. (2008). Infant preference for female faces occurs for same- but not other-race faces. Journal of Neuropsychology, 2(Pt 1), 15–26.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. Perception, 31(9), 1109–1121.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), Cognition and categorization (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 2(4), 491–502. doi: 10.1037/0096-1523.2.4.491

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. Perception and Psychophysics, 59(3), 347–357.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. Developmental Science, 12(2), 339–349.

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. Infancy, 15(6), 608–635.

Rubenstein, A. J., Kalakanis, L., & Langlois, J. H. (1999). Infant preferences for attractive faces: A cognitive explanation. Developmental Psychology, 35(3), 848–855. doi: 10.1037/0012-1649.35.3.848

Saffran, J. R., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. Science, 274(5294), 1926–1928.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. Developmental Psychology, 39(3), 484–494. doi: 10.1037/0012-1649.39.3.484

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. Infant and Child Development, 14(1), 29–50. doi: 10.1002/icd.376

Sangrigoli, S., & De Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. Journal of Child Psychology and Psychiatry, 45(7), 1219–1227. doi: 10.1111/j.1469-7610.2004.00319.x

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. Developmental Psychology, 33(1), 3–11. doi: 10.1037/0012-1649.33.1.3

Scott, L. S., & Monesson, A. (2009). The origin of biases in face perception. Psychological Science, 20(6), 676–680. doi: 10.1111/j.1467-9280.2009.02348.x

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(4), 197–201. doi: 10.1111/j.1467-8721.2007.00503.x

Sebastiàn-Gallès, N., Albareda, B., Weikum, W., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. Psychological Science, 23(9), 994–999. doi: 10.1177/0956797612436817

Shi, R., & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. Psychological Science, 12(1), 70–75. doi: 10.1111/1467-9280.00312

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. Current Opinion in Neurobiology, 11(4), 505–509. doi: 10.1016/s0959-4388(00)00241-5

Shultz, S., & Vouloumanos, A. (2010). Three-month-olds prefer speech to other naturally occurring signals. Language Learning and Development, 6(4), 241–257.

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. Journal of Memory and Language, 51(2), 173–189. doi: 10.1016/j.jml.2004.04.004

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. Infancy, 14(6), 654–666.

Singh, L., Nestor, S. S., & Bortfeld, H. (2008). Overcoming the effects of variation in infant speech segmentation: Influences of word familiarity. Infancy, 13(1), 57–74.

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. Language Learning & Development, 4, 157–178.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. Cerebral Cortex, 17, 2387–2399. doi: 10.1093/cercor/bhl147

Slaney, M., & McRoberts, G. (2003). BabyEars: A recognition system for affective vocalizations. Speech Communication, 39(3), 367–384.

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., & Parsons, S. (1998). Newborn infants prefer attractive faces. Infant Behavior and Development, 21(2), 345–354. doi: 10.1016/s0163-6383(98)90011-x

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(2), 249–261. doi: 10.1002/dev.21183

Snow, C. E., & Ferguson, C. A. (1977). *Talking to children: Language input and acquisition*. Papers from a conference sponsored by the Committee on Sociolinguistics of the Social Science Research Council (USA): Cambridge University Press.

Spelke, E. S., & Owsley, C. J. (1979). Intermodal exploration and knowledge in infancy. Infant Behavior and Development, 2(Journal Article), 13–27. doi: 10.1016/s0163-6383(79)80004-1

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature, 388(6640), 381–382. doi: 10.1038/41102

Stern, D. N., Spieker, S., Barnett, R., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. Journal of Child Language, 10(1), 1–15.

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. Nature, 259(5538), 39–41. doi: 10.1038/259039a0

Sugita, Y. (2009). Innate face processing. Current Opinion in Neurobiology, 19(1), 39–44. doi: 10.1016/j.conb.2009.03.001

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-/ in monolingual and bilingual acquisition of English. Cognition, 100(2), 369–388. doi: 10.1016/j.cognition.2005.04.007

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. Cognition, 108(1), 232–242. doi: 10.1016/j.cognition.2007.12.013

Swingley, D. (2003). Phonetic detail in the developing lexicon. Language and Speech, 46(Pt 2–3), 265–294. doi: 10.1177/00238309030460021001

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. Developmental Psychology, 43(2), 454–464. doi: 10.1037/0012-1649.43.2.454

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. Cognition, 76(2), 147–166. doi: 10.1016/s0010-0277(00)00081-0

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. Infancy, 7(1), 53–71.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? Psychological Science, 11(3), 188–195.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. Psychonomic Bulletin & Review, 9(2), 335–340. doi: 10.3758/bf03196290

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. Child Development, 47(2), 466–472. doi: 10.2307/1128803

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. Child Development, 75(4), 1067–1084.

Turati, C., Bulf, H., & Simion, F. (2008). Newborns' face recognition over changes in viewpoint. Cognition, 106(3), 1300–1321. doi: 10.1016/j.cognition.2007.06.005

Turati, C., Macchi Cassia, V., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. Child Development, 77(2), 297–311.

Tyler, M. D., Best, C. T., Goldstein, L. M., & Antoniou, M. (2014). Investigating the role of articulatory organs and perceptual assimilation in infants' discrimination of native and non-native fricative place contrasts. Developmental Psychobiology, 56, 210–227. doi: 10.1002/dev.21195

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 43(2), 161–204.

Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old infants' identification of the sources of vocalizations. Proceedings of the National Academy of Sciences of the United States of America, 106(44), 18867–18872. doi: 10.1073/pnas.0906049106

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates preference for speech. Child Development, 81(2), 517–527. doi: 10.1111/j.1467-8624.2009.01412.x

Vouloumanos, A., & Werker, J. F. (2007a). Listening to language at birth: Evidence for a bias for speech in neonates. Developmental Science, 10(2), 159–164. doi: 10.1111/j.1467-7687.2007.00549.x

Vouloumanos, A., & Werker, J. F. (2007b). Why voice melody alone cannot explain neonates' preference for speech. Developmental Science, 10(2), 169–171. doi: 10.1111/j.1467-7687.2007.00551.x

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? Developmental Psychology, 22(3), 373–377.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. Psychological Bulletin, 121(3), 437–456. doi: 10.1037/0033-2909.121.3.437

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. Science (New York, N.Y.), 316(5828), 1159. doi: 10.1126/science.1137686

Werker, J. F. (1989). Becoming a native listener. American Scientist, 77(1), 54–59.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. Developmental Psychology, 24(5), 672–683. doi: 10.1037/0012-1649.24.5.672

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. Canadian Journal of Psychology/Revue Canadienne de Psychologie, 43(2), 230–246.

Werker, J. F., Pegg, J. E., & McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. Infant Behavior and Development, 17(3), 323–333.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. Cognition, 103(1), 147–162. doi: 10.1016/j.cognition.2006.03.006

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7(1), 49–63.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. Developmental Psychobiology, 46(3), 233–251. doi: 10.1002/dev.20060

Werker, J. F., Yeung, A. C., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? Current Directions in Psychological Science, 21(4), 221–226. doi: 10.1177/0963721412449459

Wong, M., & Bhattacharjee, A. (2011). How does the visual cortex of the blind acquire auditory responsiveness? Frontiers in Neuroanatomy, 5, 52. doi: 10.3389/fnana.2011.00052

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998) Quantitative association of vocal-tract and facial behavior. Speech Communication, 26(1), 23–43. doi: 10.1016/S0167-6393(98)00048-X