The emergence of multisensory systems through perceptual narrowing

David J. Lewkowicz[1] and Asif A. Ghazanfar[2]

[1] Department of Psychology, Florida Atlantic University, Boca Raton, FL 33431, USA
[2] Neuroscience Institute, Departments of Psychology and Ecology & Evolutionary Biology, Princeton University, Princeton, NJ 08540, USA

According to conventional wisdom, multisensory development is a progressive process that results in the growth and proliferation of perceptual skills. We review new findings indicating that a regressive process – perceptual narrowing – also contributes in critical ways to perceptual development. These new data reveal that young infants are able to integrate non-native faces and vocalizations, that this broad multisensory perceptual tuning is present at birth, and that this tuning narrows by the end of the first year of life, leaving infants with the ability to integrate only socio-ecologically-relevant multisensory signals. This narrowing process forces us to reconsider the traditional progressive theories of multisensory development and opens up several new evolutionary questions as well.

The world is multisensory

The objects and events that make up our everyday experience provide us with a constant flow of sensory signals in multiple modalities. Although such inputs can potentially create confusion, our ability to integrate multisensory information enables us to have coherent and meaningful perceptual experiences. Talking faces, for example, are typically specified by various spatiotemporally congruent and modality-specific attributes as well as a host of invariant amodal attributes. Modality-specific attributes include facial configuration cues, skin color, facial hair, and the pitch and timbre of the voice. Amodal attributes provide information about the relations between visible and audible articulator actions and include such attributes as intensity, duration, tempo and rhythm [1,2]. Our ability to integrate the diverse multisensory perceptual attributes representing talking faces is crucial to our ability to extract coherent meanings from such ubiquitous communicative signals. Indeed, multisensory integration is central to adaptive behavior because it allows us to perceive a world of coherent perceptual entities and enables us to take advantage of the increased salience created by multisensory redundancy [3,4].

The central role of multisensory perception in behavior naturally raises questions about its developmental and evolutionary origins and the relationship between these two processes [5–7]. Here, we review the results from recent studies of multisensory perceptual development that challenge extant theories of multisensory perceptual

development and the conventional view that development is a progressive process. They show that multisensory perceptual tuning to vocalizing faces is initially broad, enabling young infants to integrate their native and non-native perceptual attributes, but that this tuning narrows as a function of selective experience with native perceptual attributes and that this results in a decline in the integration of non-native attributes.

The developmental problem and the progressive framework

Because the world is multisensory, a developing infant's task is to discover the multisensory coherence of the objects and events that constitute the infant's normal ecology. This task is difficult because the infant has an immature nervous system and is perceptually inexperienced. Nonetheless, infants gradually overcome these limitations and become capable of detecting multisensory coherence [8,9]. They are able to overcome these limitations for two reasons. First, multisensory coherence is easy for infants to discover because a great deal of the multisensory perceptual array consists of invariant amodal attributes [10]. Second, the early developmental limitations facilitate the emergence of multisensory perceptual skills by reducing the number of potential concurrent multisensory interactions and thus promoting the orderly integration of sensory modalities in a nervous system that might otherwise get easily overwhelmed [11].

Most studies yielding evidence of multisensory perception in infancy have been driven either explicitly or implicitly by one of two theoretical views: (a) basic multisensory perceptual abilities are not present at birth and emerge gradually during the first years of life as a result of the child's active exploration of the world and experience [12,13] (Figure 1a), or (b) they are present at birth and become increasingly differentiated and refined with experience [14] (Figure 1b). Importantly, in both theoretical views, multisensory development is thought of as a progressive process that results in the improvement of early emerging multisensory perceptual abilities and the proliferation of new ones with development and increasing experience.

On its surface, most of the empirical evidence to date is consistent with the differentiation view in showing that basic multisensory perceptual abilities are present in infancy, and that these abilities change and improve as infants grow. For example, infants between 2 and 12
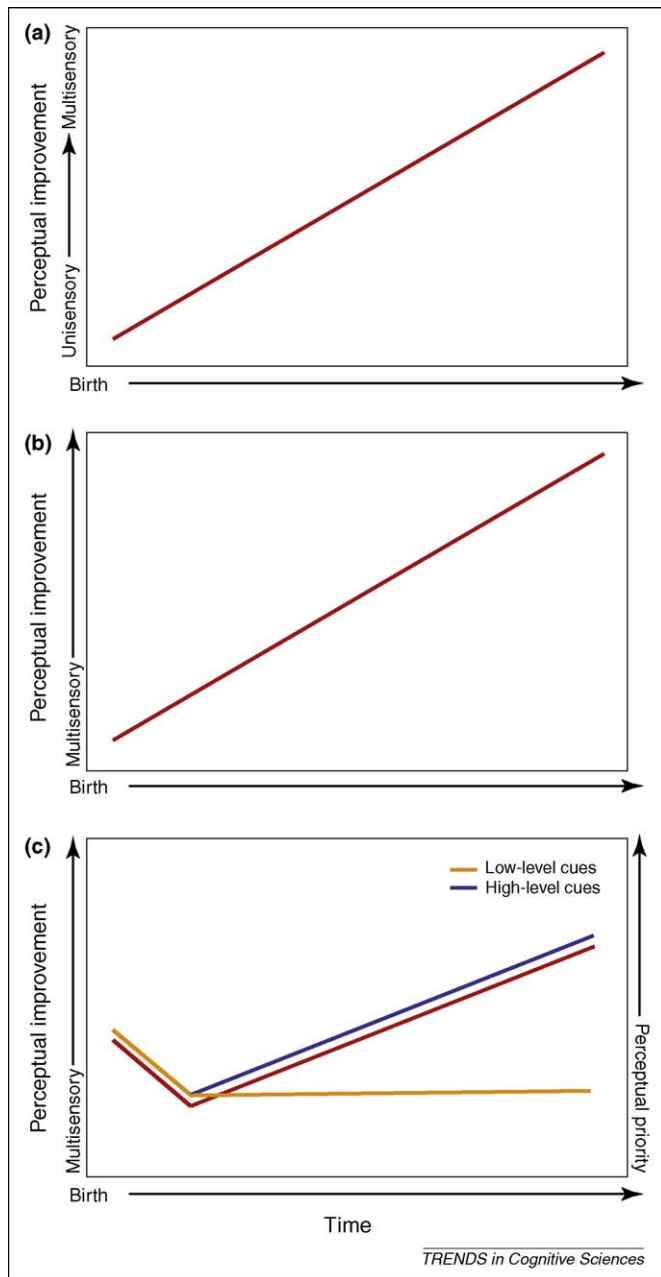
Figure 1. Three possible theoretical scenarios for the emergence of multisensory perception in early development. **(a)** Multisensory perceptual abilities are absent at birth and only emerge gradually during the first years of life as a function of the child's experience with the external world. **(b)** Basic, narrowly tuned, multisensory perceptual abilities are present at birth and, as development proceeds, these abilities become increasingly differentiated and refined. **(c)** Broadly tuned basic multisensory perceptual abilities are present at birth and as development proceeds, multisensory perceptual tuning narrows within the first year of life to match the organism's typical environment/ecology. After this point, multisensory perceptual abilities differentiate and improve as development proceeds, ending up with an expert multisensory system that is more narrowly tuned than in **(b)**. The right y-axis shows the magnitude of perceptual priority given to low-level cues (orange line) and high-level cues (blue line) as a function of developmental time.

months of age can match faces and voices based on simple intersensory cues, using mechanisms that are not likely to be specific to sensory signals in the social domain [15–17]. As infants get older, however, and acquire greater experience with conspecific (i.e. the same species) faces and voices, they become increasingly better at perceiving higher-level cues inherent in audiovisual facial expressions such as affect and gender [18,19]. Moreover,

older infants no longer rely on low-level attributes such as synchrony to perceive higher-level cues [20]. Overall, the findings suggest a progressive developmental scenario where low-level, relatively crude, multisensory abilities emerge first and are then gradually replaced by higher-level multisensory abilities. Critically, the transition from reliance on lower-level to higher-level cues for perception is vitally dependent on appropriate early sensory input [21] and, in general, a high degree of plasticity characterizes multisensory development [22–24].

The progressive framework is not the whole story: perceptual narrowing

Although the evidence consistent with the progressive theoretical framework is unquestionable, other evidence suggests that various human perceptual functions undergo developmental narrowing in early life and that this is crucial for the eventual development of species-specific patterns of perceptual expertise (see Box 1 for historical note). This evidence comes from studies of speech, face and music perception, and shows that, initially, perceptual tuning is so broad that it allows young infants to respond to native as well as non-native attributes. As development proceeds, and as infants are selectively exposed to native perceptual attributes, this tuning narrows in scope leaving older infants with a perceptual insensitivity to non-native attributes. We briefly describe this unisensory evidence and then move on to the most recent evidence for perceptual narrowing in multisensory perception.

Speech perception
In the best-known research on perceptual narrowing, Werker and Tees [25] found that 6-8 month-old English-learning infants can discriminate non-native consonants – the Hindi retroflex /Da/ versus the dental /da/ as well as the Thompson glottalized velar /k'i/ versus the uvular /q'i/ – but that 10-12 month-old infants do not. Werker and Tees concluded that the decline in responsiveness to non-native phonetic contrasts is due to language-specific experience that provides infants with continuing exposure to native consonant contrasts and no exposure to non-native ones. Subsequent cross-linguistic consonant and vowel discrimination studies have provided additional evidence of this type of narrowing [26–28]. Finally, recent work has directly adduced that narrowing of speech perception is due to experience-dependent processes, showing that English-learning infants who were exposed to natural Mandarin Chinese during play sessions between 9 and 10 months of age were better able than infants in a control group to discriminate a Mandarin Chinese phonetic contrast that does not occur in English [29].

Face perception
As in speech perception, younger infants are better at recognizing and discriminating non-native faces than are older infants [30]. Thus, 6-month-old infants can discriminate both human and monkey faces whereas 9-month-old infants can only discriminate human faces. Similar again to speech perception, the decline in non-native face discrimination is due to selective perceptual experience with human faces. For example, infants who are exposed to

monkey faces at home between 6 and 9 months of age, when sensitivity to non-native faces normally declines, can discriminate monkey faces at 9 months [31]. A similar pattern of perceptual narrowing is thought to underlie the 'other race effect' (ORE). This effect is characterized by adults' poorer discrimination of faces of people from races other than one's own [32] and it is independent of culture [33]. The ORE emerges gradually in infancy between 3 and 9 months [34] and, similar to the perception of non-native speech contrasts and the faces of other species, depends on experience [35]. For example, whereas new-born Caucasian infants do not demonstrate a preference for female over male faces when looking at Caucasian faces, 3-month-old Caucasian infants do when the faces are Caucasian but not when the faces are Asian [36]. Similar effects of experience have been found in infant perception of visual speech: 4- and 6-month-old infants can visually discriminate native from non-native articulations but 8-month-old infants are no longer able to do so [37].

Music perception
Simple metrical structure, defined by simple duration ratios of inter-onset intervals of sounds (e.g. 2:1), predominates in North American music, whereas complex metrical structure, defined by complex duration ratios (e.g. 3:2), predominates in many other musical cultures (e.g. in the Balkans). North American adults can detect differences in melodies based on alterations of simple meters, but not

when the differences are based on alterations of complex meters characteristic of Balkan music [38]. Adults of Bulgarian or Macedonian origin detect melodic differences in both simple and complex metrical structure. Similar to the adults of Bulgarian or Macedonian origin, 6-month-old, but not 12-month-old, North American infants detect simple and complex metrical structure but, unlike the adults, 12-month-olds become capable of discriminating them after a 2-week exposure to the complex meters [38,39].

Multisensory perceptual narrowing
If multisensory perception is the default mode of perception [40,41] then perceptual narrowing might be a pan-sensory process. If it is, then the progressive developmental theoretical framework does not adequately account for the development of multisensory perception. The first hint that this might be so came from a study that pre-dates the various findings of narrowing in the unisensory domain. This study showed that whereas 3-week-old infants made spontaneous intensity-based audiovisual (A-V) matches, adults did not unless they were explicitly asked to do so [42]. Importantly, the adults reported that they found the task 'bizarre'. In hindsight, the adults' reaction probably reflected a decline in a system that initially integrates audiovisual information regardless of its specific nature but that due to *multisensory* perceptual narrowing no longer does so in adulthood.

Several recent studies provide direct support for multisensory perceptual narrowing. The first of these investigated whether infant sensitivity to multisensory relations is initially broad and then narrows with development [43]. To test this prediction, 4-, 6-, 8- and 10-month-old infants were presented with side-by-side movies of the faces of a rhesus monkey producing a coo call on one side and a grunt call on the other side, and their preferences were measured for each of these visual calls in silence and then in the presence of one of the corresponding audible vocalizations (Figure 2a). Consistent with perceptual narrowing, 4- and 6-month-old infants matched the visual and audible calls by looking longer at the visible call in the presence of the corresponding vocalization than in its absence (Figure 2b). In contrast, 8- and 10-month-old infants exhibited no evidence of multisensory matching. Because the onsets of the facial gestures and the corresponding audible calls were presented in temporal synchrony, it was concluded that successful matching in the younger infants was mediated by synchrony. Indeed, a follow-up study revealed that when synchrony is disrupted, infants no longer make the multisensory matches [44]. In addition, the failure of the older infants to match was not due to unisensory deficits because the older infants easily discriminated between the visual call gestures and between the audible calls. This decline in multisensory matching of non-native faces and vocalizations was found to persist up to 18 months of age (Figure 2b) [44].

The developmental pattern of initial multisensory matching followed by its decline contrasts with all previous findings on face–voice matching which show that infants as young as two months of age and as old as 18 months of age can perform face–voice matching when the faces and voices
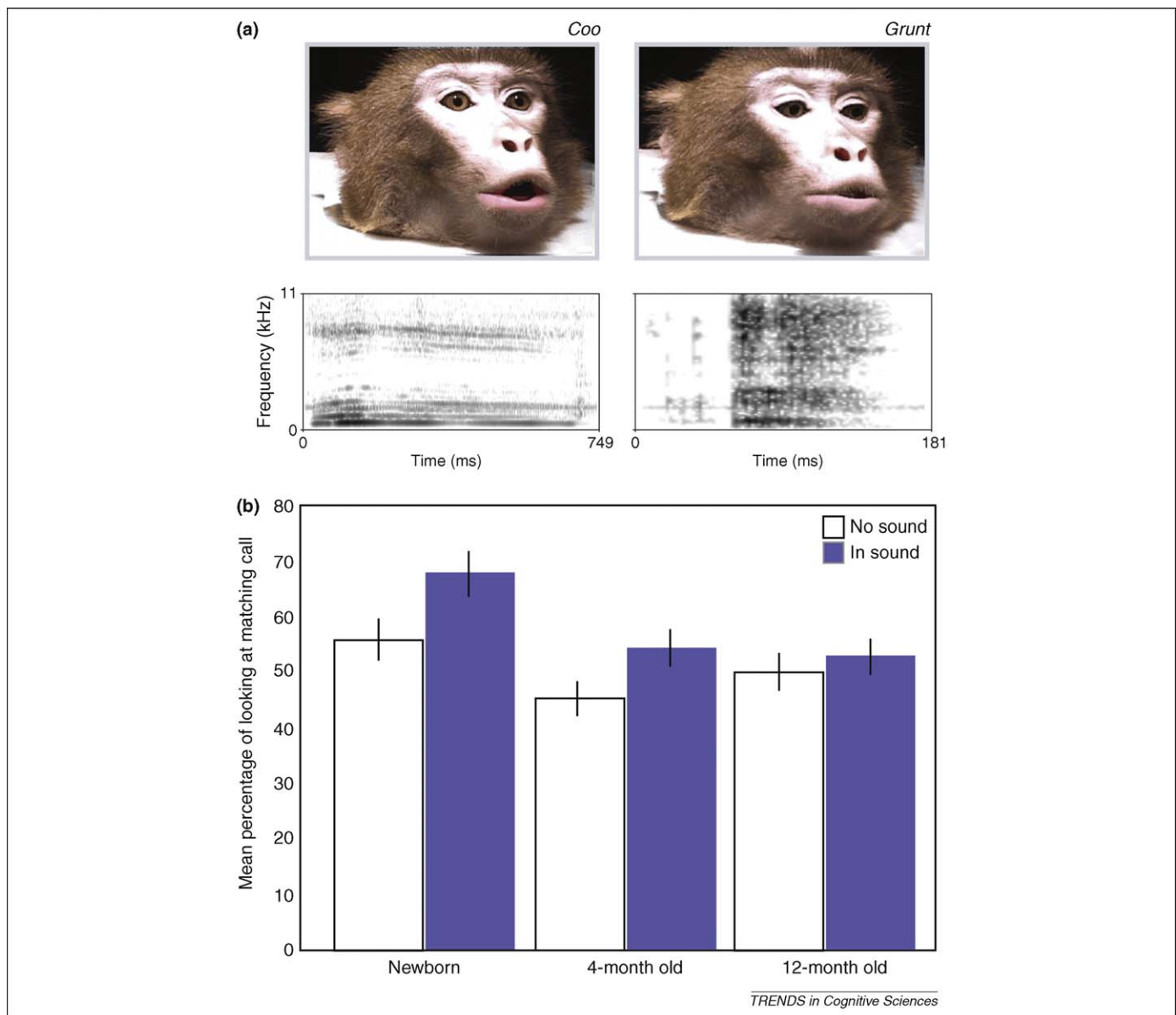
Figure 2. Cross-species multisensory integration in human infants. **(a)** Top two panels depict still images of the facial gestures corresponding to the coo and grunt calls at the point of maximum mouth opening and the two bottom panels depict the spectrograms of the respective vocalizations. **(b)** The mean percentage of looking time at the matching call out of the total amount of looking time directed at both facial gestures in the no-sound condition (i.e. when the videos of the two visual gestures were presented side-by-side in silence) compared to the mean percentage of looking time at the matching call in the in-sound condition (i.e. when the vocalization accompanied the presentation of the videos of the two facial gestures). Shown are the results from studies with newborn [46], 4-month-old [43], and 12-month-old [44] infants showing that the two younger groups of infants exhibited a significant preference for the matching facial gesture when listening to the vocalization but that the older infants did not.

are human [15,18,19,45]. Overall, the findings on infant multisensory responses to native and non-native faces and vocalizations indicate that whereas infants' ability to perceive various types of native face–voice relations improves with age, their ability to perceive non-native face–voice relations declines.

That infants undergo a period of multisensory perceptual narrowing raises at least three questions. First, does the broad multisensory perceptual tuning found at four months of age and the ability to integrate the social signals of another species represent the initial developmental condition in humans (i.e. is it present at birth)? Second, does this broad multisensory perceptual tuning extend to other domains (e.g. speech)? Finally, does multisensory perceptual narrowing occur in other primate species? The answer to the first question is affirmative. A recent

study showed that newborn infants also match monkey facial gestures and vocalizations (Figure 2b) [46], and that newborns can make such matches even when the normally available dynamic correlations [1] between facial gesture cues and the acoustic envelope and spectral (formant) structure is eliminated. The latter finding suggests that newborns perform such multisensory matching on the basis of stimulus energy onsets and offsets and that, consistent with previous work [43,44], this broad multisensory perceptual tuning is based on sensitivity to low-level audio-visual temporal synchrony cues (Figure 1c, orange line).

If the decline in multisensory matching of cross-species faces and voices reflects a general feature of multisensory perceptual development, then it should also be observed in the development of audiovisual speech perception. To test

---

Box 2. Putative neural mechanisms of perceptual narrowing

The kinds of narrowing effects found in the development of speech, face and music perception as well as the multisensory perception of face–voice raise the obvious question of what putative neural mechanisms might underlie perceptual narrowing effects. Naturally, it is tempting to link 'selectionist' or regressive theories of neural development [74,75] with the regressive nature of perceptual narrowing. These theories postulate that neural development occurs in two stages, the first of which is the construction of neuronal networks that are initially diffuse and somewhat global in nature. This first stage is constructed through genetic and epigenetic factors and sets up what will ultimately be considered 'exuberant' connections. The second stage involves the selective elimination of some of the connections in this initial network, leading to a more modularized network that is better adapted to mediate mature perceptual and motor skills needed in the current species-typical environment. In this stage, the reshaping (or 'pruning') of the network occurs through the competitive stabilization of some synapses versus others. The competition is decided through experience. This neuro-developmental scheme fits perfectly with the phenomena related to perceptual narrowing: the initially diffuse network mediates the broad tuning of early infant perception and experience and subsequently sculpts the network to generate more finely tuned perceptual capacities. Indeed, a recent review of unisensory perceptual narrowing concludes that at the neural level, narrowing is due to the pruning of exuberant synaptic connections [76].

Although conceptually elegant, there are many problems with the selectionist scheme [77,78]. First, the basic premise of the theory is that there are extra synapses in the initial developmental state, resulting in extra-exuberant axonal and dendritic arbors. Although this might be true for a few brain regions (e.g. the transient connections between the visual cortex and the spinal cord in the developing rodent brain), it is not true for many others (e.g. the axonal arborizations of thalamocortical neurons in layer 4 of the rodent somatosensory cortex [79] and ferret visual cortex [80]. The fact that extra synapses are not widespread throughout the developing brain throws into question the relevance of selectionist theories to cognitive development [78]. The most damning evidence against them, however, is simply that as the primate brain matures it grows in size for a long time after birth [77]. This growth is attributable to neurons increasing their morphological complexity through the elaboration of axonal and dendritic processes. For example, there is an explosive rise in the number of synapses in the perinatal rhesus monkey brain, followed by a long period during which a steady number of synapses is present [81]. In other words, *there is a net gain in synapses over the course of development*. As a result, the narrowing that is observed at the functional level is most probably due to the formation of new neural connections rather than to the loss of neurons and/or their connections through a Darwinian-like process of selective pruning.

What does this all mean for the neural basis of perceptual narrowing? The neural developmental data suggest that perceptual narrowing is probably the result of a *selective elaboration* of synapses, whose relevance is determined by postnatal experience, rather than the selective pruning of irrelevant synapses.

---

this prediction, 6- and 11-month-old Spanish-learning and English-learning infants were presented with facial gestures and speech sounds representing the /ba/ and /va/ phonetic distinction [47]. Because the /v/ sound is not phonetically relevant in Spanish, it was predicted that younger Spanish-learning infants would match the facial gestures representing a /ba/ and a /va/ syllable with their corresponding vocalizations, but that older infants would not. In contrast, it was expected that English-learning infants would match at both ages. In the experiment, infants were familiarized with one of the two audible syllables and then were given preference trials during which they saw the same person mouthing the /ba/ syllable on one screen and the /va/ syllable on the other. As predicted, the 6-month-old Spanish-learning infants matched the visible and audible syllables, whereas the older infants did not. In contrast, the English-learning infants matched at both ages. The decline was found to persist into adulthood. When adults were asked to indicate in a forced choice task which of the two facial gestures representing the /ba/ and /va/ syllables corresponded to an immediately preceding presentation of one or the other audible syllable, Spanish-speaking adults made random choices whereas English-learning adults made correct intersensory matches on over 90% of the trials [47].

Is perceptual narrowing unique to humans?

The evidence for multisensory perceptual narrowing raises questions about the evolution of this developmental mechanism. Naturally, the only way to get at this issue is through comparative studies with non-human primates (hereafter, primates). Given that both humans and other extant primates use both facial and vocal expressions as communication signals, it is perhaps not surprising that many primates recognize the correspondence between the visual and auditory components of vocal signals. Macaque monkeys (*Macaca mulatta, Macaca fuscata*), capuchins (*Cebus apella*) and chimpanzees (*Pan troglodytes*) all perceive A-V coherence [48–53].

Given the apparent cross-species homology in the multisensory perception of communication signals, are the developmental processes leading to the emergence of these abilities similar or different across species? Because the timing of neural development in primates and humans differs (it is *heterochronous*), the developmental emergence of multisensory perception probably differs across species (see Box 2 for putative neural mechanisms). For example, relative to humans, monkeys are neurologically precocial, possessing ~65% of their adult brain size at birth, whereas human infants only possess ~25% of their adult brain size at birth [54,55]. In addition, the myelination of fiber tracts is more mature in monkeys than in humans at the same postnatal age [54,56]. Thus, given that primates exhibit multisensory perception of social signals, it is interesting to ask whether the mechanisms leading to the development of this ability are similar in primates and, thus, whether this includes the process of perceptual narrowing. Certainly, the neural precocity of non-human primates relative to humans suggests that they might not be so 'open' to the effects of early sensory experience and thus might not exhibit narrowing.

Does multisensory perception narrow in monkeys?

If a relatively immature state of neural development leaves a developing infant more open to the effects of early sensory experience, then a more advanced state of neural development might result in a different outcome. Thus, monkeys might be born with a perceptual system that is already tuned to a much narrower range of sensory input and might only be able to integrate the faces and

vocalization of their own species. In other words, they might be closed to the effects of early sensory experience. Alternatively, monkeys might be born with a perceptual system that is tuned to a broad range of sensory input, but because of their advanced state of neural development they might not be as susceptible as humans are to the effects of early experience. As a result, monkeys might either be permanently tuned to a broader range of sensory input or might require a greater amount of experience before perceptual narrowing exerts its full effects. In either case, monkeys would not be expected to exhibit perceptual narrowing effects.

These possibilities were investigated in developing infant vervet monkeys (an Old World monkey species; formerly *Cercopithecus aethiops*, now classified as *Chlorocebus pygerethrus*) by testing whether they can match the faces and vocalizations of another species with which they had no prior experience [57]. As in the human infant study described above [43], infant vervets ranging in age from 23 to 65 weeks (~6 to 15 months) were tested in a preference task in which they viewed pairs of the same rhesus monkey face producing a coo call on one side and a grunt call on the other side and heard one of the calls at the same time (Figure 2a). Even though the vervets had no prior exposure to rhesus monkey faces and vocalizations, they matched them. Interestingly, however, they did so by looking at the non-matching face for a greater proportion of overall looking time (but see below) (Figure 3a). Importantly, they *exhibited matching well beyond the age of perceptual narrowing in human infants*.

The seemingly 'opposite' pattern of the vervets' responsiveness (i.e. looking more at the non-matching face) was due to the increased affective salience of the matching face-vocalization combination (i.e. either a visible and audible coo or a grunt). This finding was interpreted to mean that the vervets were more fearful when presented with the matching than the mismatching combination. This conclusion was confirmed in a follow-up experiment in which the affective value of the audible call was eliminated by

replacing the voice with a complex tone [57]. Under these conditions, the vervets looked longest at the matching pairs. Moreover, an analysis of pupillary responses revealed that the vervets' pupils were more dilated (an affective response) when they looked at the matching natural face/vocalization combination than when they looked at the face/tone combination (Figure 3b). Because the infant vervets in this study exhibited cross-species multisensory matching far later in development than do human infants, these findings suggest either that multisensory perceptual narrowing does not occur in Old World monkeys or that it occurs later in their development.

Why do infant vervets continue to match hetero-specific faces and voices at a postnatal and neurological age that, relative to human infants, is beyond the time when multisensory perceptual narrowing should have occurred? One possibility is that monkeys are 'stuck' with a broader range of sensitivity because of the more precocial nature of their nervous system. The other is that monkeys' precocial brains are not stuck *per se* but, rather, are less plastic because of their more advanced developmental state [58]. Thus, vervets might still be sensitive to social experience, but it might take them longer to incorporate the effects of such experience and, consequently, to exhibit perceptual narrowing. The latter possibility is consistent with the development of vocal behavior in vervets in that their ability to produce vocalizations, use them in appropriate contexts, and respond appropriately to the vocalizations of conspecifics emerges gradually during the first four years of life [59]. For example, 3-month-old infant vervets produce different alarm calls in response to three general categories of threat: 'terrestrial predator', 'aerial predator' and 'snake-like object'. However, they do not distinguish between real predators and non-predators. Only as they grow older do they restrict their alarm calling to the small number of genuine predators within each category. It is also consistent with the fact that in Japanese macaques (another Old World monkey species), unisensory and multisensory representations are influenced by the amount of
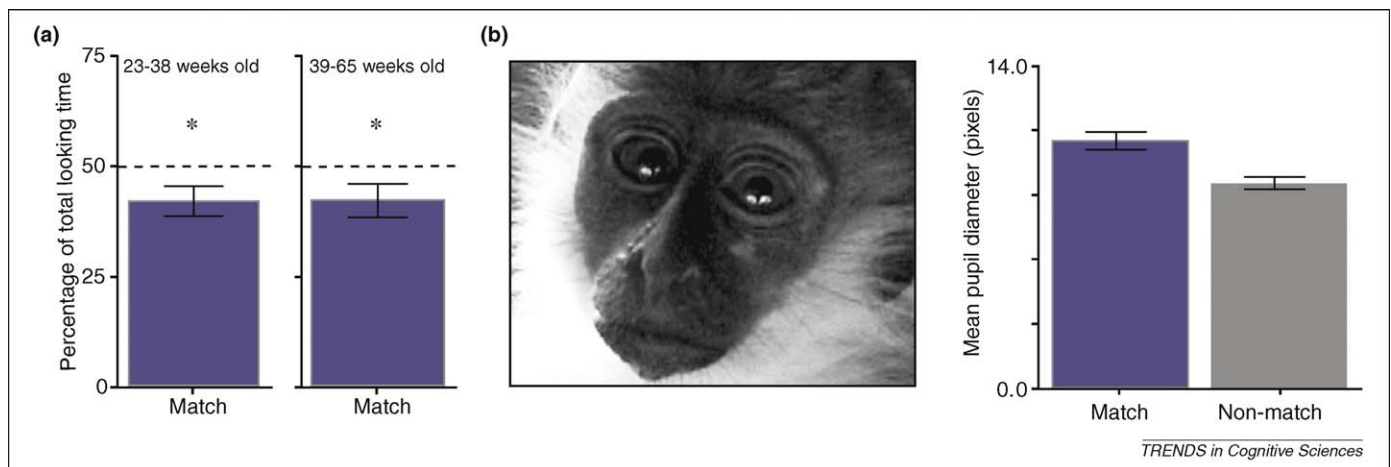

Figure 3. Cross-species multisensory integration in young vervet monkeys. **(a)** The percentage of total looking time the vervets looked at the matching video in the presence of the vocalization at two different ages. Although the amount of time looking at the matching video was significantly different from chance, please note that it was significantly *less* than was the amount of looking time directed at the mis-matching facial gesture. **(b)** Picture depicts the pupils of a vervet monkey showing that this monkey's pupils were dilated when it looked at the matching natural face/vocalization combination. The bar graph shows that vervets' pupils were more dilated when they were looking at the matching natural face/vocalization combination than when they were looking at the non-matching natural face/vocalization combination, suggesting that the former elicited a greater affective response than did the latter.

exposure they have to conspecifics and heterospecifics [60,61] and that adults in many primate species show a behavioral advantage for processing the faces of their own species [62].

Vervets and other Old World monkeys do seem to exhibit developmental narrowing in other domains. For example, in the vocal domain, there is an apparent decline in infant vervet monkeys' ability to produce certain vocalizations (e.g. 'wrr-like' vocalizations) but this ability returns later in life [63]. It should be noted, however, that this claim might be too strong given the evidence: it is very difficult to determine (particularly in field studies) whether a vocalization is truly absent from the repertoire or whether the appropriate context to elicit the vocalizations simply did not arise during the periods when the infant monkeys were observed. Another example of narrowing in vervets, more rapid than alarm-calling behavior, is the use of grunt vocalizations [59]. One- to 8-week-old vervets produce grunts in two broad social contexts, but between 17 weeks and 2 years of age produce a specific type of grunt only towards dominant individuals. Finally, some neonatal (1- to 3-day-old) rhesus monkeys imitate lip-smacking and tongue-protrusion gestures, but this ability declines after a few days [64]. This is in stark contrast to chimpanzees and humans: both species exhibit neonatal imitation but retain this capacity (i.e. there is no narrowing) [65,66], again suggesting the importance of neural developmental differences across primate species and individuals [67].

Conclusions

Experience can have complementary effects on the development of perceptual functions. On the one hand, it can induce and facilitate the emergence of a particular perceptual function and it can maintain that function through continued exposure to specific sensory inputs. On the other, it also can have the seemingly opposite but, in reality, complementary effect. That is, as experience with native sensory input accumulates, the scope of initially broad perceptual abilities is narrowed to best match the infant's native perceptual ecology. Such perceptual tuning is achieved through regressive rather than progressive developmental processes and, importantly, this does not lead to a permanent loss of function but rather to its re-organization [68] (Figure 1c). At the perceptual level, the regressive processes are reflected in a decline in responsiveness to non-native sensory attributes, whereas at the neural level the regressive processes are probably reflected in increasing synaptic connections rather than the pruning of an excess of neurons and exuberant connections (Box 2). Finally, the effects of early experience can differ for different functions. For example, narrowing of responsiveness to non-native vowels occurs several months before the narrowing of responsiveness to consonants. Similar developmental heterochronies have been found in the development of multisensory perceptual abilities [8]. Thus, the specific timing of multisensory narrowing also might depend on the nature of the information, the modality within which it is processed, the specific modalities involved, the timing of sensory system development, the rate of neural development, and the ecological context of the organism (i.e. the

> **Box 3. Outstanding questions**
>
> - How general is multisensory perceptual narrowing; in other words, does it occur in other species besides humans?
> - Does the presence and/or absence of multisensory perceptual narrowing depend on the ecological context of the particular species? That is, is it a mechanism that is important for only those species that are in close proximity to other similar looking species?
> - Does the presence and/or absence of multisensory perceptual narrowing depend on the rate of neural development and/or ecological context or does one factor dominate the other?
> - Does multisensory perceptual narrowing occur for modality pairings besides auditory-visual and might it occur even when more than two modalities are involved?

species involved)—see Box 3 for some additional outstanding questions.

References
1 Chandrasekaran, C. *et al.* (2009) The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5, e1000436
2 Munhall, K.G. and Vatikiotis-Bateson, E. (2004) Spatial and Temporal Constraints on Audiovisual Speech Perception. In *The handbook of multisensory processes* (Calvert, G.A. *et al.*, eds), pp. 177–188, MIT Press
3 Bahrick, L.E. *et al.* (2004) Intersensory Redundancy Guides the Development of Selective Attention, Perception, and Cognition in Infancy. *Curr. Directions Psychol. Sci.* 13, 99–102
4 Lewkowicz, D.J. and Kraebel, K. (2004) The value of multimodal redundancy in the development of intersensory perception. In *Handbook of multisensory processing* (Calvert, G. *et al.*, eds), MIT Press
5 Kingsbury, M.A. and Finlay, B.L. (2001) The cortex in multidimensional space: where do cortical areas come from? Commentary. *Dev. Sci.* 4, 125–142
6 Krubitzer, L. (2007) The magnificent compromise: Cortical field evolution in mammals. *Neuron* 56, 201–208
7 Oyama, S. (2000) *The ontogeny of information*, Duke University Press
8 Lewkowicz, D.J. (2002) Heterogeneity and heterochrony in the development of intersensory perception. *Cogn. Brain Res.* 14, 41–63
9 Lickliter, R. and Bahrick, L.E. (2000) The development of infant intersensory perception: advantages of a comparative convergent-operations approach. *Psychol. Bull.* 126, 260–280
10 Gibson, E.J. (1969) *Principles of perceptual learning and development*, Appleton
11 Turkewitz, G. and Kenny, P.A. (1982) Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Dev. Psychobiol.* 15, 357–368
12 Birch, H.G. and Lefford, A. (1967) Visual Differentiation, Intersensory Integration, and Voluntary Motor Control. *Monogr. Soc. Res. Child Dev.* 32, 1–87
13 Piaget, J. (1952) *The origins of intelligence in children*, International Universities Press
14 Gibson, E.J. (1984) Perceptual development from the ecological approach. In *Advances in developmental psychology* (Lamb, M.E. and et, al., eds), pp. 243–286, Lawrence Erlbaum Associates
15 Kuhl, P.K. and Meltzoff, A.N. (1982) The bimodal perception of speech in infancy. *Science* 218, 1138–1141
16 Patterson, M.L. and Werker, J.F. (2003) Two-month-old infants match phonetic information in lips and voice. *Dev. Sci.* 6 (2), 191–196
17 Walton, G.E. and Bower, T.G. (1993) Amodal representations of speech in infants. *Infant Behav. & Deve.* 16, 233–243

18 Patterson, M.L. and Werker, J.F. (2002) Infants' ability to match dynamic phonetic and gender information in the face and voice. *J. Exp. Child Psychol.* 81, 93–115

19 Kahana-Kalman, R. and Walker-Andrews, A.S. (2001) The role of person familiarity in young infants' perception of emotional expressions. *Child Dev.* 72, 352–369

20 Walker-Andrews, A.S. (1986) Intermodal perception of expressive behaviors: Relation of eye and voice? *Dev. Psy.* 22, 373–377

21 Maurer, D. *et al.* (2007) Sleeper effects. *Dev. Sci.* 10, 40–47

22 Horng, S.H. and Sur, M. (2006) Visual activity and cortical rewiring: activity-dependent plasticity of cortical networks. *Prog. Brain Res.* 157, 3–11

23 Wallace, M.T. and Stein, B.E. (2007) Early experience determines how the senses will interact. *J. Neurophysiol.* 97, 921

24 King, A.J. (2008) Visual influences on auditory spatial learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci* 364, 331–339

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

26 Best, C.T. *et al.* (1995) Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behav. & Dev.* 18, 339–350

27 Kuhl, P.K. *et al.* (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608

28 Cheour, M. *et al.* (1998) Development of language-specific phoneme representations in the infant brain. *Nature Neurosci.* 1, 351–353

29 Kuhl, P.K. *et al.* (2003) Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9096–9101

30 Pascalis, O. *et al.* (2002) Is face processing species-specific during the first year of life? *Science* 296, 1321–1323

31 Pascalis, O. *et al.* (2005) Plasticity of face processing in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5297–5300 Epub 2005 Mar 5224

32 Chiroro, P. and Valentine, T. (1995) An investigation of the contact hypothesis of the own-race bias in face recognition. *Q. J. Exp. Psychol. A: Human Experimental Psychology* 48A, 879–894

33 Kelly, D.J. *et al.* (2009) Development of the other-race effect during infancy: evidence toward universality? *J. Exp. Child Psychol.* 104, 105–114

34 Kelly, D.J. *et al.* (2007) The Other-Race Effect Develops During Infancy: Evidence of Perceptual Narrowing. *Psychol. Sci* 18, 1084–1089

35 Sangrigoli, S. and de Schonen, S. (2004) Recognition of own-race and other-race faces by three-month-old infants. *J. Child Psychol. Psychiatr.* 45, 1219–1227

36 Quinn, P.C. *et al.* (2008) Infant preference for female faces occurs for same - but not other -race faces. *J. Neuropsychol. (Special Issue on Face Processing)* 2, 15–26

37 Weikum, W.M. *et al.* (2007) Visual language discrimination in infancy. *Science* 316, 1159

38 Hannon, E.E. and Trehub, S.E. (2005) Metrical Categories in Infancy and Adulthood. *Psychol. Sci* 16, 48–55

39 Hannon, E.E. and Trehub, S.E. (2005) Tuning in to musical rhythms: infants learn more readily than adults. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12639–12643

40 Rosenblum, L.D. (2005) The primacy of multimodal speech perception. In *Handbook of speech perception* (Pisoni, D. and Remez, R., eds), pp. 51–78, Blackwell

41 Ghazanfar, A.A. (in press) The default mode of primate vocal communication and its neural correlates. In *Multisensory Object Perception in the Primate Brain* (Naumer, M.J., and Kaiser, J., eds), Springer

42 Lewkowicz, D.J. and Turkewitz, G. (1980) Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Dev. Psychol.* 16, 597–607

43 Lewkowicz, D.J. and Ghazanfar, A.A. (2006) The decline of cross-species intersensory perception in human infants. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6771–6774

44 Lewkowicz, D.J. *et al.* (2008) The decline of cross-species intersensory perception in human infants: Underlying mechanisms and its developmental persistence. *Brain Res.* 1242, 291–302

45 Poulin-Dubois, D. *et al.* (1994) Infants' intermodal knowledge about gender. *Dev. Psychol.* 30, 436–442

46 Lewkowicz, D.J., *et al.* Intersensory Perception at Birth: Newborns Match Non-Human Primate Faces & Voices. *Infancy* (in press)

47 Pons, F., *et al.* (2009) Narrowing of intersensory speech perception in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.0904134106

48 Jordan, K. *et al.* (2005) Monkeys match the number of voices they hear to the number of faces they see. *Curr. Biol.* 15, 1034–1038

49 Ghazanfar, A. *et al.* (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr. Biol.* 17, 425–430

50 Adachi, I. *et al.* (2006) Japanese macaques form a cross-modal representation of their own species in their first year of life. *Primates* 47, 350–354

51 Ghazanfar, A.A. and Logothetis, N.K. (2003) Facial expressions linked to monkey calls. *Nature* 423, 937–938

52 Izumi, A. and Kojima, S. (2004) Matching vocalizations to vocalizing faces in a chimpanzee (Pan troglodytes). *Anim. Cogni.* 7, 179–184

53 Parr, L.A. (2004) Perceptual biases for multimodal cues in chimpanzee (Pan troglodytes) affect recognition. *Anim. Cogn.* 7, 171–178

54 Malkova, L. *et al.* (2006) Longitudinal magnetic resonance imaging study of rhesus monkey brain development. *Eur. J. Neurosci.* 24, 3204–3212

55 Sacher, G.A. and Staffeldt, E.F. (1974) Relation of gestation time to brain weight for placental mammals: implications for the theory of vertebrate growth. *Am. Naturalist* 108, 593–615

56 Gibson, K.R. (1991) Myelination and behavioral development: A comparative perspective on questions of neoteny, altriciality and intelligence. In *Brain maturation and cognitive development: comparative and cross-cultural perspectives* (Gibson, K.R. and Petersen, A.C., eds), pp. 29–63, Aldine de Gruyter

57 Zangenehpour, S. *et al.* (2009) Heterochrony and cross-species intersensory matching by infant vervet monkeys. *PLoS ONE* 4, e4302

58 Kaas, J.H. (1991) Plasticity of sensory and motor maps in adult animals. *Annu. Rev. Neurosci.* 5, 137–167

59 Seyfarth, R.M. and Cheney, D.L. (1986) Vocal development in vervet monkeys. *Anim. Behav.* 34, 1640–1658

60 Adachi, I. *et al.* (2009) Plasticity of the ability to form cross-modal representations in infant Japanese macaques. *Dev. Sci.* 12, 446–452

61 Sugita, Y. (2008) Face perception in monkeys reared with no exposure to faces. *Proc. Nat.l Acad. Sci. U.S.A.* 105, 394–398

62 Dufour, V. *et al.* (2006) Face processing limitation to own species in primates: A comparative study in brown capuchins, Tonkean macaques and humans. *Behav. Processes* 73, 107–113

63 Hauser, M.D. (1989) Ontogenetic changes in the comprehension and production of vervet monkey (Cercopithecus aethiops) vocalizations. *J. Comp. Psychol.* 103, 149–158

64 Ferrari, P. *et al.* (2006) Neonatal imitation in rhesus macaques. *PLoS Biology* 4, 1501

65 Bard, K. (2007) Neonatal imitation in chimpanzees (Pan troglodytes) tested with two paradigms. *Anim. Cogn.* 10, 233–242

66 Meltzoff, A.N. and Moore, M.K. (1977) Imitation of facial and manual gestures by human neonates. *Science* 198, 75–78

67 Ferrari, P. *et al.* (2009) Interindividual Differences in Neonatal Imitation and the Development of Action Chains in Rhesus Macaques. *Child Dev* 80, 1057–1068

68 Werker, J.F. and Tees, R.C. (2005) Speech Perception as a Window for Understanding Plasticity and Commitment in Language Systems of the Brain. *Dev. Psychobiol. Special Issue: Critical Periods Re-examined: Evidence from Human Sensory Development* 46, 233–234

69 Holt, E.B. (1931) *Animal drive and the learning process*, Holt

70 Kuo, Z.Y. (1976) *The dynamics of behavior development: An epigenetic view*, Plenum

71 Gottlieb, G. (1991) Experiential canalization of behavioral development: Results. *Dev. Psychol.* 27, 35–39

72 Lewis, M.D. (2000) The promise of dynamic systems approaches for an integrated account of human development. *Child Dev.* 71, 36–43

73 Thelen, E. and Smith, L.B. (1994) *A dynamic systems approach to the development of cognition and action*, MIT Press

74 Cowan, W.M. *et al.* (1984) Regressive events in neurogenesis. *Science* 225, 1258–1265

75 Low, L.K. and Cheng, H.J. (2006) Axon pruning: an essential step underlying the developmental plasticity of neuronal connections. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1531–1544

76 Scott, L.S. *et al.* (2007) A domain general theory of the development of perceptual discrimination. *Curr. Directions Psychol. Sci.* 16, 197–201

77 Purves, D. *et al.* (1996) Is neural development darwinian? *Trends Neurosci.* 19, 460–464

78 Quartz, S.R. and Sejnowski, T.J. (1997) The neural basis of cognitive development: A constructivist manifesto. *Behav. Brain Sci.* 20, 537–556

79 Agmon, A. *et al.* (1993) Organized growth of thalamocortical axons from the deep tier of terminations into layer IV of developing mouse barrel cortex. *J. Neurosci.* 13, 5365–5382

80 Crowley, J.C. and Katz, L.C. (1999) Development of ocular dominance columns in the absence of retinal input. *Nat. Neurosci.* 2, 1125–1130

81 Bourgeois, J.P. and Rakic, P. (1993) Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. *J. Neurosci.* 13, 2801–2820