

The decline of cross-species intersensory perception in human infants

David J. Lewkowicz*[†] and Asif A. Ghazanfar[‡]

*Department of Psychology, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431; and [‡]Program in Neuroscience, Department of Psychology, Princeton University, Green Hall, Princeton, NJ 08544

Communicated by Charles G. Gross, Princeton University, Princeton, NJ, March 13, 2006 (received for review January 30, 2006)

Between 6 and 10 months of age, infants become better at discriminating among native voices and human faces and worse at discriminating among nonnative voices and other species' faces. We tested whether these unisensory perceptual narrowing effects reflect a general ontogenetic feature of perceptual systems by testing across sensory modalities. We showed pairs of monkey faces producing two different vocalizations to 4-, 6-, 8-, and 10-month-old infants and asked whether they would prefer to look at the corresponding face when they heard one of the two vocalizations. Only the two youngest groups exhibited intersensory matching, indicating that perceptual narrowing is pan-sensory and a fundamental feature of perceptual development.

crossmodal | face processing | multisensory

From the moment of birth, infants find themselves in a socially rich environment where they see and hear other people. In order for them to have veridical and meaningful social experiences with such people, infants must be able to integrate particular faces and voices by detecting their correspondences. Indeed, a number of studies have shown that, beginning as early as 2 months of age, infants begin to exhibit the ability to perceive face–voice correspondences (1–8). Despite this fact, however, the developmental process underlying intersensory integration of faces and voices, as well as more general intersensory processes, remain poorly understood. Current theoretical views assume either that basic intersensory perceptual abilities are present at birth and become increasingly differentiated and refined over age (9) or that such abilities are not present at birth and only emerge gradually during the first years of life as a result of the child's active exploration of the world (10, 11).

Most empirical evidence supports the former, differentiation, view in showing that basic intersensory perceptual abilities are already present in infancy and that as infants grow these abilities change and improve in significant ways (12, 13). For example, young human infants can perceive lower-order intersensory relations based on such attributes as intensity (14), temporal synchrony (15, 16), and duration (17), but do not integrate auditory and visual spatial cues (18). In contrast, older infants can perceive higher-order intersensory relations based on such attributes as affect (6) and gender (3), become capable of learning arbitrary intersensory associations (19) and can integrate auditory and visual spatial cues (18). Findings from studies of underlying neural mechanisms of intersensory integration in cats and rhesus monkeys show a similar pattern. Whereas multisensory cells in the superior colliculus of adult cats and rhesus monkeys integrate auditory and visual cues in spatial localization tasks, these cells do not integrate them in neonatal cats and monkeys (20, 21). Together, extant findings suggest a general developmental pattern consisting of the initial emergence of low-level intersensory abilities, a subsequent age-dependent refinement and improvement of existing abilities, and the ultimate emergence of new higher-level intersensory abilities. This pattern is consistent with conventional notions of development as consisting of a process of increasing differenti-

ation and specificity for the detection of new stimulus properties, patterns, and distinctive features (9, 22).

Recent studies of the development of unisensory perceptual abilities have raised the possibility that the conventional theoretical view of intersensory perceptual development and the empirical findings on which it is based may, at a minimum, be incomplete. Studies of responsiveness to unisensory stimulation have shown that perceptual development also includes processes that lead to perceptual narrowing and that this results in reduced processing capacity. For example, in the speech domain, it has been shown that infants younger than 6–8 months of age can discriminate categorically native as well as nonnative phonetic contrasts, but that 10- to 12-month-old infants no longer can discriminate nonnative phonetic contrasts (23). For example, whereas 6- to 8-month-old English learning infants can detect the English labial/alveolar contrast such as /ba/ vs. /da/ and the Hindi unvoiced unaspirated retroflex/dental contrast such as /ta/ vs. /Ta/, 10- to 12-month-old infants no longer can discriminate the Hindi contrast. Similarly, in the visual domain, it has been shown that 6-month-old infants can learn and discriminate between different human faces as well as between different monkey faces but that 9-month-old infants only can learn and discriminate between different human faces (24). Together, these findings indicate that the perception of sounds and faces becomes reorganized during development, and that this results in a general narrowing of perceptual ability.

It is possible that the intersensory perceptual narrowing effects and the fact that they appear at similar time points in infancy reflect a general and, therefore, pan-sensory characteristic of perceptual development. Despite this possibility, however, none of the extant theories on the development of intersensory perception mention the possibility that perceptual narrowing may occur in the intersensory domain as well. As a result, we devised a cross-species intersensory matching task to investigate whether perceptual narrowing also may occur in the development of intersensory perception during human infancy. The task allowed infants to choose between one of two vocalizing monkey faces while listening to a vocalization that corresponded to one of the faces. We assumed that infants had not been previously exposed to vocalizing monkey faces and, thus, that their responsiveness reflected processing in the absence of prior experience. Furthermore, we assumed that intersensory perceptual tuning for face–voice relations is initially broad and that it then narrows over the first year of life as infants acquire increasingly greater experience with human faces and voices. As a result, we predicted that young infants would match monkey faces and voices, but that older infants would not. This finding would differ markedly from previous findings showing that infants can match human faces and voices as early as 2 months of age and as late as 12 months of age (1–8).

Conflict of interest statement: No conflicts declared.

[†]To whom correspondence should be addressed. E-mail: lewkowicz@fau.edu.

© 2006 by The National Academy of Sciences of the USA

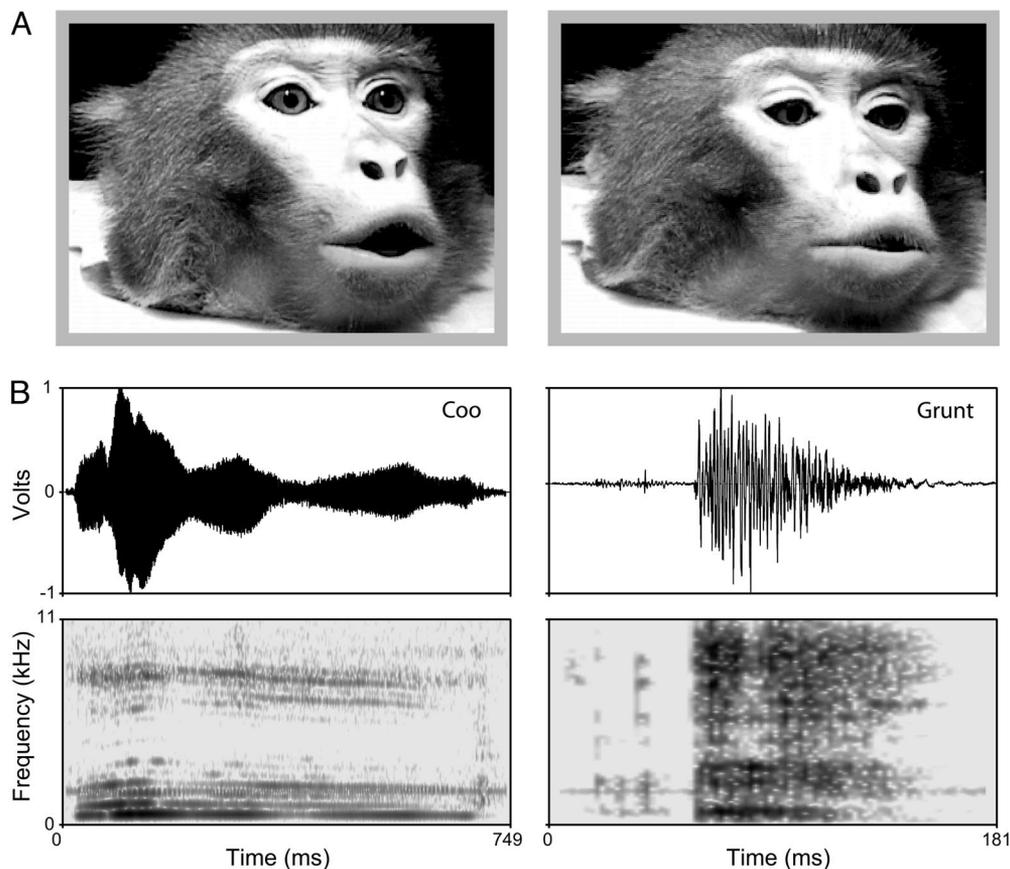


Fig. 1. Representative example of the stimuli. (A) Single video frames of facial gestures made during a coo (Left) and a threat (Right) call, shown at the point of maximal mouth opening. (B) Oscillograms (Upper) and spectrograms (Lower) of the coo and threat vocalizations: coo calls are long and tonal; threat calls are short, pulsatile, and noisy.

We tested separate groups of infants consisting of 34 4-month-old, 57 6-month-old, 54 8-month-old, and 32 10-month-old infants. The stimuli consisted of pairs of digitized videos of rhesus monkeys (*Macaca mulatta*) producing two different calls: a “coo” and a “grunt” (Fig. 1A). These calls are frequently produced by monkeys during affiliative encounters (25). The coo call is a long tonal signal and is accompanied by a small mouth opening and protruding lips, whereas the grunt call is a noisy, pulsatile signal of short duration and is characterized by mouth opening with little or no lip protrusion (Fig. 1). We used a visual paired-preference testing procedure to measure looking time to each of two side-by-side dynamic faces, first in silence and then during acoustic playback of a vocalization that corresponded to one of the two facial expressions (see *Materials and Methods* for details).

Each infant was first given two silent test trials followed by two in-sound test trials. During the silent test trials infants viewed two side-by-side videos showing the same monkey performing the coo call on one side and the grunt call on the other side in silence. The lateral position of the two calls was reversed on the second of these two silent test trials. During the two in-sound test trials, infants viewed the same videos again, but this time they either heard the audible coo or the audible grunt. The onset of the audible call was synchronized with the initiation of lip movement in both videos but its offset was only synchronized with the matching video. The lateral position of the two faces was reversed on the second in-sound test trial.

Ideally, we might have tested responsiveness to human faces and voices in the same infants and then compared their performance in an intraspecies versus a cross-species intersensory

matching task. However, given the nature of the methodology, this was impractical: infants typically do not tolerate more than the four test trials in a testing session. Also, a large body of evidence shows clearly that human infants as young as 2 months of age and as old as 12 months of age can make intraspecies (human face–voice) intersensory matches in the same kind of task used here (1–8). Thus, it is unlikely that the subjects used in the current study would be exceptional in this capacity.

Results

Our findings were consistent with the prediction that perceptual development is not modality-specific but is a general, pan-sensory ontogenetic process. When we gave infants a visual choice test between a video of a monkey producing a facial expression that matched a concurrently presented vocalization and another face that did not match, the 4- and the 6-month-old infants looked at the matching facial expression, whereas the 8- and the 10-month-old infants did not. For infants in each audible call group, we first computed the proportion of time that each infant looked at the matching face, separately for the silent and the in-sound trials. This consisted of dividing the total amount of time of looking at the matching face by the total amount of looking at both faces. An initial examination of the proportion of looking in the silent condition revealed that infants had an inherent preference for the face that produced the coo call. As a result, to demonstrate that infants were making intersensory matches, it was essential that we take this initial preference into account. Thus, we compared the proportion of time that infants looked at the matching face in the presence of the vocalization versus its absence. If infants were perceiving the relation be-

tween the audible and visible components of the calls, then they were expected to exhibit greater looking at the face in the presence of the vocalization than in its absence. To determine whether the particular call (i.e., coo or grunt), the particular animal producing the call, and/or the infant's age differentially affected the results, we first conducted a repeated ANOVA, with call (2), animal (2), and age (4) as the between-subjects factors and trial (silent, in-sound) as the within-subjects factor. Results of this analysis did not yield any meaningful interactions. As a result, the data at each age were collapsed across call and animal and, based on our *a priori* prediction of perceptual narrowing during development, we conducted separate two-tailed *t* tests comparing looking at the matching face in silence versus looking at it in the presence of the vocalization. These tests indicated that the 4-month-old infants looked significantly longer at the matching face in the in-sound condition than in the silent condition ($t = 2.14$, $df = 33$, $P < 0.05$) and that the 6-month-old infants did so as well ($t = 2.03$, $df = 56$, $P < 0.05$). In contrast, the results from the two older groups indicated that neither the 8-month-old infants ($t = 1.53$, $df = 53$, $P = 0.13$) nor the 10-month-old infants ($t = 1.47$, $df = 31$, $P = 0.15$) exhibited differential looking in the two conditions.

Discussion

The current findings provide clear and direct new evidence that perceptual narrowing also occurs during intersensory development. Our findings show that young human infants perform cross-species intersensory matching of facial and vocal expressions but that older infants no longer do. What might have made it possible for the younger infants to make intersensory matches? The onset of the audible call was in synchrony with the matching and nonmatching faces but its offset was only in synchrony with the matching face. In addition, the duration of the matching face corresponded to the duration of the matching vocalization. Thus, the younger infants most likely based their matches on the onset and offset synchrony of the facial expression and its corresponding vocalization and/or on the common duration of the two. If that is the case, the fact that older infants did not make intersensory matches suggests that these low-level features are no longer of interest to older infants and that they search for higher-level, more meaningful features when attempting to integrate vocal and facial information. This shift most likely reflects the effects of increasing experience with faces and voices and increasing sophistication in perceptual and cognitive processing mechanisms.

Similar to unisensory narrowing effects found in the auditory and visual modalities (23, 24, 26, 27), our findings show that intersensory perceptual narrowing effects also occur early in infancy. Pascalis *et al.* (24) have hypothesized that the parallel developmental pattern of narrowing found in the speech and face domain probably reflects a common perceptual-cognitive tuning apparatus that is not specific to a particular sensory modality and one that has experience-expectant properties (28). The current findings provide empirical evidence that a common tuning apparatus does, indeed, mediate perceptual narrowing effects and suggest that such effects are a general, pan-sensory property of perceptual development. There is certainly great adaptive value to be gained from such a process. Most of our real-world experiences are multisensory in nature and, thus, we must constantly integrate information across different sensory modalities (29). A developmental mechanism that is initially broad enough to capture intersensory relations, regardless of whether such relations originate in same-species or cross-species events, is highly adaptive because it helps initially inexperienced infants to discover the fact that sensory stimulation in different modalities can be, and often is, related in meaningful ways. With specific regard to faces and voices, this kind of mechanism enables infants to first discover that specific faces and voices

belong together and then, as they become more perceptually experienced, to learn to integrate the relevant, species-specific features of faces and voices and to ignore those that are not.

Although the current study examined cross-species intersensory perception, the broader significance of our findings lies in the fact that they are consistent with current ideas regarding the effects of early experience on ultimate structural/functional outcome. In essence, early experience with human faces and voices provides infants with experience-expectant inputs (28) that, through the process of canalization (30), result in the development of species-specific perceptual expertise. Our findings add to the evidence of intersensory integration in early human development and show that, initially, this ability is so broad as to accommodate nearly all forms of multisensory stimulation, including signals from a different species of primate. This ability, and the broad perceptual window that it provides at the start of development, is adaptive because it serves to bootstrap the infant's perceptual system to begin learning about the relevant features of its world and to ignore irrelevant ones. Such bootstrapping mechanisms have already been identified in other domains of perceptual learning. For example, mechanisms exist in early infancy that permit infants to discover speech segmentation procedures that are appropriate to their native language and, thus, enable them to begin the process of acquiring language-specific phonological knowledge and a lexicon. These mechanisms consist of an initial sensitivity to the general rhythmic properties of their native speech at birth which helps infants to segment incoming speech into meaningful categories (31). In a like manner, our results suggest that an early ability to integrate information across the different modalities bootstraps early multisensory learning and that, with experience, the specific kinds of intersensory connections become pruned down to those that are most relevant to the infant's species-specific needs and ecology.

Materials and Methods

Participants. All participants were full-term, healthy infants recruited by telephone from birth records. The 4-month group had a mean age of 19 weeks ($SD = 1.1$ weeks) and consisted of 16 boys and 17 girls, the 6-month group had a mean age of 28.2 weeks ($SD = 1.2$ weeks) and consisted of 29 boys and 28 girls, the 8-month group had a mean age of 36.9 weeks ($SD = 1.8$ weeks) and consisted of 26 boys and 28 girls, and the 10-month group had a mean age of 45.6 weeks ($SD = 1.5$ weeks) and consisted of 15 boys and 17 girls. Fifty-nine additional infants were tested but did not contribute usable data; 48 of these because of fussing (14 at 4 months, nine at 6 months, 14 at 8 months, and 11 at 10 months), six because of equipment failure, and five because of distractions.

Apparatus. Infants were tested individually in a quiet room with an ambient sound pressure level (SPL) of 58 dB (A level). They were seated in an infant seat 50 cm in front of two 17-inch liquid-crystal-display monitors. The monitors were placed side-by-side with a space in between them to accommodate the lens of a video camera and a circuit-board atop the camera containing several light-emitting-diodes (LEDs). The LEDs were used to attract the infant's attention to the center before the start of each trial. A speaker was located behind the camera.

Each video was shown inside a window measuring 19.05 cm high and 25.40 cm wide, and each monkey's head filled most of the window. The video was a 2-s digital recording of a facial/vocal expression looped continuously for 1 min. To permit generalization, and to rule out the possibility that responsiveness was based on some idiosyncratic property of the animal, we presented two different stimulus sets with two different stimulus animals. The SPL of the coo call for one animal was 77 dB and

72 dB for the other animal, and the corresponding grunt call was 81 dB and 77 dB, respectively.

Procedure. The experiment consisted of four 1-min test trials. The LEDs flashed on and off at the start of each test trial and as soon as the infant looked at them, they were turned off and the videos began to play. Infants were videotaped during the testing session, and, subsequently, the infants' visual behaviors were coded by observers who were blind to the testing conditions and to the stimuli being presented on a given trial; this consisted of recording the direction (i.e., left monitor, right monitor, or away from the monitors) and duration of each of the infant's looks

throughout each test trial. Interobserver reliability was calculated on a sample of randomly chosen infants representing the various ages tested. The average level of agreement on the total duration of looking to each side on each trial was 97%.

We thank Jennifer Hughes for her assistance in infant recruitment and data collection; Silvia Place and Ryan Sowinski for their help in coding; and Kari Hoffman, Olivier Pascalis, Barry Stein, and Janet Werker for their comments on an earlier draft. All digital videos were of socially housed monkeys at the Max Planck Institute for Biological Cybernetics, Tuebingen, Germany. This work was supported by National Institute of Child Health and Human Development Grant HD35849 (to D.J.L.).

1. Kuhl, P. K. & Meltzoff, A. N. (1982) *Science* **218**, 1138–1141.
2. Patterson, M. L. & Werker, J. F. (2003) *Dev. Sci.* **6**, 191–196.
3. Patterson, M. L. & Werker, J. F. (2002) *J. Exp. Child Psychol.* **81**, 93–115.
4. Walton, G. E. & Bower, T. G. (1993) *Infant Behav. Dev.* **16**, 233–243.
5. Kahana-Kalman, R. & Walker-Andrews, A. S. (2001) *Child Dev.* **72**, 352–369.
6. Walker-Andrews, A. S. (1986) *Dev. Psychol.* **22**, 373–377.
7. Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S. & Diaz, I. (1991) *Ecol. Psychol.* **3**, 55–75.
8. Poulin-Dubois, D., Serbin, L. A., Kenyon, B. & Derbyshire, A. (1994) *Dev. Psychol.* **30**, 436–442.
9. Gibson, E. J. (1969) *Principles of Perceptual Learning and Development* (Appleton, New York).
10. Piaget, J. (1952) *The Origins of Intelligence in Children* (International Universities Press., New York).
11. Birch, H. G. & Lefford, A. (1967) *Monogr. Soc. Res. Child Dev.* **32**, 1–87.
12. Lewkowicz, D. J. (2000) *Psychol. Bull.* **126**, 281–308.
13. Lewkowicz, D. J. (2002) *Cognit. Brain Res.* **14**, 41–63.
14. Lewkowicz, D. J. & Turkewitz, G. (1980) *Dev. Psychol.* **16**, 597–607.
15. Lewkowicz, D. J. (1996) *J. Exp. Psychol.* **22**, 1094–1106.
16. Lewkowicz, D. J. (1992) *Infant Behav. Dev.* **15**, 297–324.
17. Lewkowicz, D. J. (1986) *Infant Behav. Dev.* **9**, 335–353.
18. Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J. & Shimojo, S. (2006) *Dev. Sci.*, in press.
19. Reardon, P. & Bushnell, E. W. (1988) *Infant Behav. Dev.* **11**, 245–250.
20. Wallace, M. T. & Stein, B. E. (1997) *J. Neurosci.* **17**, 2429–2444.
21. Wallace, M. T. & Stein, B. E. (2001) *J. Neurosci.* **21**, 8886–8894.
22. Werner, H. (1973) *Comparative Psychology of Mental Development* (International Universities Press, New York).
23. Werker, J. F. & Tees, R. C. (1984) *Infant Behav. Dev.* **7**, 49–63.
24. Pascalis, O., Haan, M. D. & Nelson, C. A. (2002) *Science* **296**, 1321–1323.
25. Rowell, T. E. & Hinde, R. A. (1962) *Proc. Zool. Soc. London* **138**, 279–294.
26. Hannon, E. E. & Trehub, S. E. (2005) *Psychol. Sci.* **16**, 48–55.
27. Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992) *Science* **255**, 606–608.
28. Greenough, W. T., Black, J. E. & Wallace, C. S. (1987) *Child Dev.* **58**, 539–559.
29. Stein, B. E. & Meredith, M. A. (1993) *The Merging of the Senses* (MIT Press, Cambridge, MA).
30. Gottlieb, G. (1991) *Dev. Psychol.* **27**, 35–39.
31. Nazzi, T. & Ramus, F. (2003) *Speech Communication* **41**, 233–243.