Nonhuman primate vocalizations support categorization in very young human infants

Alissa L. Ferry[a,b,1], Susan J. Hespos[b], and Sandra R. Waxman[b]

[a]Cognitive Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, 34136 Trieste, Italy; and [b]Department of Psychology, Northwestern University, Evanston, IL 60208

Language is a signature of our species and our primary conduit for conveying the contents of our minds. The power of language derives not only from the exquisite detail of the signal itself but also from its intricate link to human cognition. To acquire a language, infants must identify which signals are part of their language and discover how these signals are linked to meaning. At birth, infants prefer listening to vocalizations of human and nonhuman primates; within 3 mo, this initially broad listening preference is tuned specifically to human vocalizations. Moreover, even at this early developmental point, human vocalizations evoke more than listening preferences alone: they engender in infants a heightened focus on the objects in their visual environment and promote the formation of object categories, a fundamental cognitive capacity. Here, we illuminate the developmental origin of this early link between human vocalizations and cognition. We document that this link emerges from a broad biological template that initially encompasses vocalizations of human and nonhuman primates (but not backward speech) and that within 6 mo this link to cognition is tuned specifically to human vocalizations. At 3 and 4 mo, nonhuman primate vocalizations promote object categorization, mirroring precisely the advantages conferred by human vocalizations, but by 6 mo, nonhuman primate vocalizations no longer exert this advantageous effect. This striking developmental shift illuminates a path of specialization that supports infants as they forge the foundational links between human language and the core cognitive processes that will serve as the foundations of meaning.

infancy | developmental tuning | language acquisition | conceptual development | language and thought

Human infants rapidly tune in to the signals of human language. At birth, they prefer listening to the vocalizations of human and nonhuman primates over artificial sounds; within months, as their listening preferences narrow, they become attuned to human speech, and to their own native language(s) in particular (1–7).

Interestingly, the developmental path underlying infants' rapid tuning to the faces of communicative partners follows a similar trajectory: Infants' prowess in human face-processing derives from a broader template that initially encompasses faces of both humans and nonhuman primates before tuning specifically to human faces (8–10). Moreover, these tuning effects are also evident in infants' cross-modal (auditory-visual) matching. Initially, human infants reveal a broad capacity to match species-specific vocalizations and faces (e.g., matching human vocalizations to human faces and nonhuman primate vocalizations to nonhuman primate faces), but within the first year of life, this broad pattern of cross-modal matching narrows, as infants tune specifically to the correspondences between human vocalizations and faces (11–14). Perceptual tuning effects like these, ubiquitous throughout the animal kingdom (15, 16), are adaptive: They ensure that infants direct their attention increasingly toward the communicative signals of our species and toward the individuals who will serve as their communicative and pedagogical partners.

However, to learn language, infants must do more than tune in to the sounds of their native language. They must also discover how these sounds are linked to meaning. Remarkably, infants make rapid headway toward establishing just such a link. For infants as young as 3 mo of age, human language evokes more than simple listening preferences: it also engages infants' core cognitive capacities. Listening to human vocalizations engenders in young infants more in-depth processing of the objects in their environment and supports their formation of object categories (17). Categorization, a building block of cognition, is inherent in a wide range of processes including conceptual organization and reasoning (18–20). In infants, a precocious link between language and object categorization was documented using a simple object categorization task. Infants first viewed a series of distinct images from one object category (e.g., dinosaurs). Each image was presented in conjunction with either a segment of human speech or a sequence of sine-wave tones, matched precisely to the human speech segment in mean frequency, amplitude, duration, and pause lengths. At test, all infants viewed two test images, presented in silence: a new member of the now-familiar category (a dinosaur) and a member of a novel category (a fish). Following the logic established for infant looking-time paradigms, if infants formed the object category (dinosaurs), they were expected to distinguish the test images, expressing this discrimination behaviorally with a reliable preference for either the familiar or the novel image (21).

The results were straightforward. Infants hearing the tone sequences failed to discriminate between the test images at any age. In sharp contrast, infants hearing human speech successfully formed the object category, expressing their distinction first as a familiarity preference (in 3-mo-old infants) and later as a novelty preference (in 4- to 12-mo-old infants) (17, 22, 23). This age-related shift is consistent with a systematic progression documented in infancy research: familiarity preferences tend to be expressed early (in very young infants and in older infants before they have gained sufficient familiarization), and novelty preferences tend to be expressed later in development, or after sufficient familiarization (21, 24, 25). Although the precise mechanism underlying this progression is not yet fully specified, advances in brain maturation, processing speed, and encoding efficiency are implicated. These results documented that human speech is not only a preferred signal, but also confers an adaptive advantage: By 3 mo, infants are not only tuned to the communicative signals of their partners, but also to a principled and surprisingly early link between these signals and the fundamental cognitive process of categorization.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

This evidence is intriguing, but it leaves open a key developmental question: What is the developmental origin of this link? Is the link specific to human vocalizations from the start, or does it derive from a broader template, perhaps one that initially also encompasses the vocalizations of nonhuman primates? Here, we address this question directly. To do so, we traced the effect of introducing two new auditory signals—a nonhuman primate vocalization and backward human speech—on infants' object categorization over the first 6 mo of life.

Experiment 1: Nonhuman Primate Vocalization

If the link between human language and cognition emerges from a broader template, then the vocalizations of nonhuman primates may initially confer the same adaptive advantage as human vocalizations.

Thirty-six infants participated, 12 at each of three ages (3, 4, and 6 mo old). We adapted the categorization task described earlier (17, 23), this time accompanying each of the familiarization images with a vocalization of a blue-eyed Madagascar lemur (*Eulemur macaco flavifrons*) (Fig. 1). By 3 mo, infants discriminate vocalizations of humans from those of nonhuman primates (1, 14). We selected a lemur vocalization because although it differs distinctly from human vocalizations, it nonetheless shares certain acoustic properties with infant-directed speech. More specifically, the lemur vocalization that we selected falls naturally within the mean frequency range of infant-directed human speech and is comparable in duration to the infant-directed speech segments that were presented to infants in previous object categorization tasks (17, 22, 23; Audio File S1). At test, infants viewed two images simultaneously: a new member of the familiar category (e.g., a new dinosaur) and a member of a novel category (e.g., a fish), presented in silence. We recorded infants' looking time to each image.

At 3 and 4 mo, infants' responses to the lemur vocalization mirrored precisely their responses to human vocalizations (17), reliably distinguishing the novel from familiar test images. Infants exhibited familiarity preferences at 3 mo [$t(11) = -3.427$, $P = 0.006$] and novelty preferences at 4 mo [$t(11) = 2.693$, $P = 0.021$] (Fig. 2). However, by 6 mo lemur vocalizations no longer exerted this facilitative effect [mean = 0.52, SD = 0.07, $t(11) = 1.039$, $P = 0.321$] (Fig. 2). An ANOVA documented this significant effect of age [$F(3,26) = 13.702$, $P < 0.001$].

A series of supplementary analyses compared infants' responses to lemur vocalizations (reported here) and human speech [reported previously (17), using the same inclusion criteria]. First, we compared 3- and 4-mo-olds' responses to human speech and lemur vocalizations directly. Infants responded identically to
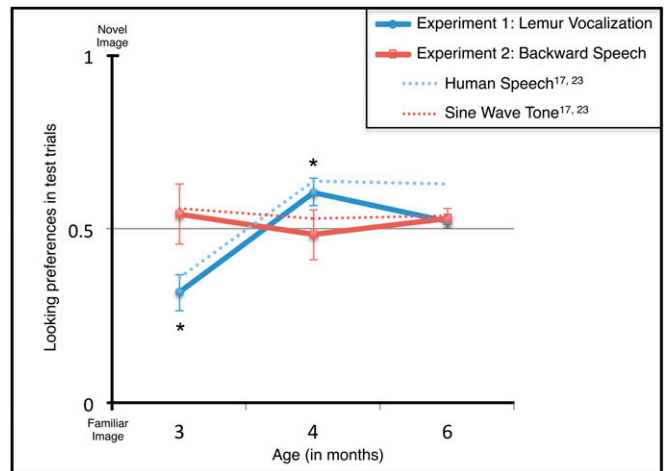

Fig. 2. Infants' responses to lemur vocalizations (experiment 1) and backward speech (experiment 2). At 3 and 4 mo, infants hearing lemur vocalizations were not only more likely than chance, but also more likely than those hearing backward speech, to discriminate between the novel and familiar test objects; at 6 mo, neither lemur vocalizations nor backward speech supported object categorization. For the purpose of comparison, we also show infants' responses to human vocalizations (e.g., "Look at the modi!") and to pure sine-wave tone sequences [e.g., 400- or 800-Hz tones; matched to the human vocalization stimuli for mean frequency, duration, and pause length (17, 23)]. Error bars represent ±1 SEM. Significant differences between preference score and chance performance (0.50) are marked by an asterisk ($P < 0.05$).

these two conditions [$F(1,42) = 0.004$, $P > 0.05$]. Second, we examined individual infants' responses to these vocalizations. The results were striking: 92% of the 3-mo-olds hearing lemur vocalizations (and 90% of those hearing human speech) exhibited familiarity preferences ($P = 0.006$, binomial comparison); 83% of the 4-mo-olds hearing lemur vocalizations (and 83% hearing human speech) exhibited novelty preferences ($P = 0.039$). Third, strong correlations between infants' age and preference scores, documented previously in infants' responses to human speech [$r(20) = 0.61$, $P < 0.05$], were echoed here in infants' responses to lemur vocalizations [$r(22) = 0.60$, $P < 0.05$] (Fig. 3A). Finally, to gain insight into the shift from familiarity to novelty preferences at 3 and 4 mo, we examined the performance of an additional group of infants: 4-mo-olds who heard lemur vocalizations but were excluded from the main analyses because they attended insufficiently to the familiarization images. These infants, like their 4-mo-old counterparts who heard human speech but who also accrued minimal exposure to the familiarization images, exhibited a significant familiarity preference [$t(10) = -1.935$, $P = 0.041$, one-tailed]. The finding that 4-mo-olds with minimal exposure to the familiarization performed identically to their more attentive 3-mo-old counterparts suggests that the systematic shift from familiarity to novelty preference in this task may be mediated by infants' processing and encoding of the materials, rather than by their maturational state. Together, these analyses highlight the compelling parallels between 3- and 4-mo-olds' responses to human speech and lemur vocalizations.

These results reveal that the link between human language and object categorization, evident at 3 mo of age, derives from a broader template that initially encompasses vocalizations of human and nonhuman primates and is rapidly tuned specifically to human vocalizations. However, what is the best account of this phenomenon? Is the facilitative effect conferred by lemur (and human) vocalizations specific to primate vocalizations? Or is it a consequence of the acoustic complexity of these vocalizations, especially compared with sine-wave tones? After all, it is well
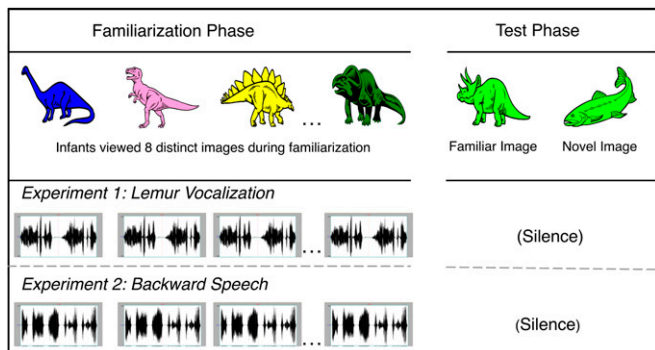
Fig. 1. Experimental design. During the familiarization phase, each infant viewed eight different exemplars, presented sequentially, in conjunction with either the lemur vocalization (experiment 1) or backward speech (experiment 2). During the test phase, each infant viewed images from the familiar and novel categories, presented simultaneously in silence.
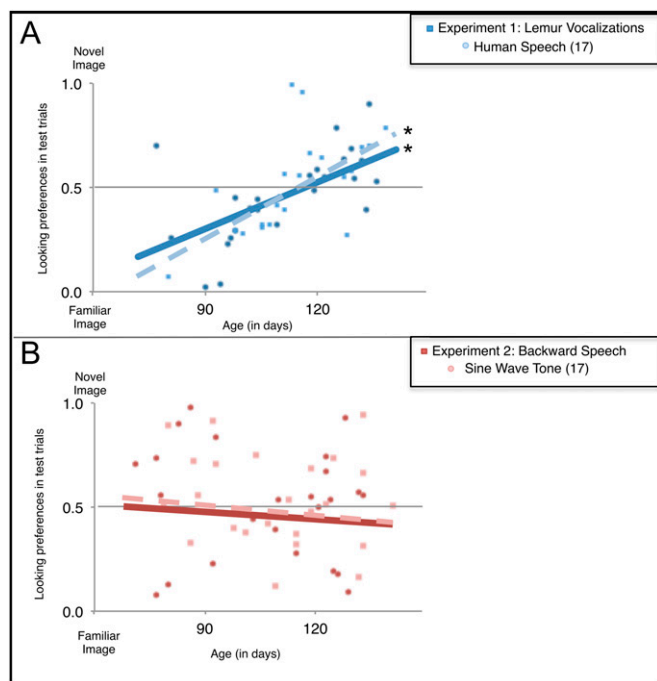
Fig. 3. Correlations between age (in days) and preference score for infants hearing (*A*) lemur vocalizations and (*B*) backward speech. For the purpose of comparison, we also show the correlations for the human vocalization and sine-wave tone conditions (17). Significant correlations are marked by an asterisk ($P < 0.05$).

documented that infants' attention and arousal are enhanced in the context of complex acoustic stimuli (26, 27).

Experiment 2: Backward Human Speech

To adjudicate between these alternatives, we presented another group of 36 infants (12 per age group) with backward speech in the context of the same object categorization task described earlier. In fact, we presented the same segment of human speech from ref. 17, but played it backward (Audio File S2). This segment provides an ideal point of comparison, for although backward speech is identical to forward speech in perceptual complexity, it violates fundamental acoustic properties of forward speech, is processed in different brain regions than forward speech from birth, and is impossible to produce with a mammalian vocal tract (3, 28). We reasoned as follows: If the cognitive advantage conferred by lemur (and human) vocalizations is a general consequence of young infants' heightened attention in the context of complex acoustic stimuli, then 3- and 4-mo-old infants hearing backward speech should also successfully form object categories.

However, this was not the case (Fig. 2). Instead, infants hearing backward speech failed to form object categories, performing at chance levels at each age: 3 mo [$t(11) = 0.504, P = 0.624$], 4 mo [$t(11) = -0.212, P = 0.836$], and 6 mo [$t(11) = 0.280, P = 0.280$]. An ANOVA comparing infants' responses at each age to lemur vocalizations (experiment 1) and backward speech (experiment 2) revealed a significant interaction [$F(5,66) = 5.021, P = 0.009$]: although infants' responses to these two signals differed at 3 and 4 mo, by 6 mo, infants' responses to these signals were indistinguishable.

Finally, a supplementary series of analyses revealed that infants' responses to backward speech (reported here) mirrored their responses to sine-wave tone sequences [reported previously (17), using the same inclusion criteria]. An ANOVA revealed no main effects or interactions between infants' responses to these signals (all $P > 0.05$). Moreover, as was the case for tone

sequences, there was no correlation between infants' age and preference scores for backward speech [$r(22) = -0.18, P > 0.05$] (Fig. 3*B*).

General Discussion

The current work reveals that tuning processes not only guide human infants to the communicative signals of our species, but also support the very first links between these signals and the core cognitive capacities that will ultimately constitute the foundations of meaning. We document that a link between human language and object categorization, evident at 3 mo of age, derives from a broader template that initially encompasses vocalizations of human and nonhuman primates and is rapidly tuned specifically to human vocalizations. This work sheds light on a path of developmental specialization that supports infants' first steps toward establishing the links between human language and cognition that are the hallmark of the human mind and the primary conduit for the cultural transmission of knowledge.

This work provides unique insight into the processes that undergird infant language development, concept development, and the links between them. First, we document that in human infants, the fundamental process of developmental narrowing extends well beyond purely perceptual discriminations and preferences (1, 5–7). Perceptual narrowing is well underway within the first months of life. For example, neonates' listening preferences, which include both human and nonhuman primate vocalizations, give way to a strong preference for human over nonhuman vocalizations by 3 mo (1, 5). However, despite their clearly-tuned preference for human vocalizations, nonhuman primate vocalizations conferred precisely the same conceptual advantages at 3 and 4 mo as human vocalizations, an advantage that by 6 mo was tuned specifically to human vocalizations. This constitutes evidence that the power of developmental tuning reaches beyond perceptual capacities alone to support the establishment of foundational cross-modal links between human language and the core cognitive processes that will serve as the foundations of meaning.

Second, we identify very detailed parallels in young infants' responses to human and nonhuman vocalizations. Despite their preference for human vocalizations (1) and the vast differences in their prior exposure to these stimuli, infants hearing human and nonhuman vocalizations revealed the same, precisely timed shift from familiarity preferences (at 3 mo) to novelty preferences (at 4 mo) (Figs. 2 and 3). This parallel, striking in itself, also sets limits on the role of prior exposure in infants' processing speed, encoding efficiency and their expression of familiarity or novelty preferences (21, 24, 25).

Third, these findings offer insight into the mechanisms underlying infants' initially broad link between language and object categorization (17). Certainly, experience with human language is instrumental in infants' first months of life (1, 5, 7, 29). Infants may quickly learn that human language heralds interesting visible events or brings with it rewarding affective experiences. Associations like these would likely engender in infants heightened attention in the context of human vocalizations. However, we document here that lemur vocalizations—an entirely unfamiliar signal—engender the same facilitative effect as human vocalizations at 3 and 4 mo. We therefore conclude that the starting point for human infants' link between language and core cognitive processes cannot be derived from learning alone.

These findings also open several avenues for additional research. First, it will be important to discover the boundary conditions on the signals that initially engage infants' cognitive processes, including object categorization. Is this link sufficiently broad to include naturalistic vocalizations beyond those of our closest genealogical cousins, or is it restricted to primates, whose vocalizations may be perceptually just close enough to our own to serve as early candidates for the platform on which human

language is launched? Does the initial link encompass communicative signals from other modalities (e.g., signed languages), or does it privilege vocalizations from the start (30, 31)? Additional research will also be required to ascertain whether the tuning process that we have documented here has begun even earlier, in infants younger than 3 mo of age. It will also be important to investigate whether and how infants' exposure to multiple languages influences the course of this developmental progression (32, 33).

Another key question concerns the boundary conditions on the kinds of cognitive capacities that are initially engaged by human and nonhuman vocalizations. The evidence reported here documents an advantageous effect of human and nonhuman primate vocalizations on one fundamental cognitive process: object categorization. Do vocalizations initially support other cognitive processes (e.g., object individuation), or does the link to cognition initially privilege object categorization? Finally, additional work will help to reveal whether and how the striking developmental effects documented here are gated by experience, maturation, or an interaction between these twin engines of development.

Methods

Participants. Seventy-two healthy, full-term infants from predominantly college-educated, white families living in the Greater Chicago Area participated. All procedures were approved by the Northwestern University Institutional Review Board and informed consent was obtained from the parents of all infants. There were 36 infants per experiment divided into 12 infants per age: 3-mo-olds (2 mo, 11 d to 3 mo, 20 d; mean = 3 mo, 2 d), 4-mo-olds (3 mo, 25 d to 4 mo, 16 d; mean = 4 mo, 6 d), and 6-mo-olds (5 mo, 18 d to 6 mo, 24 d; mean = 6 mo, 3 d). Infants who devoted insufficient attention (less than 50% overall) to familiarization images (43, distributed evenly across conditions and sex) were analyzed separately. Additional infants were excluded for fussiness (10) and parental interference (4). Excluded infants did not differ in either age or sex from included infants; this exclusion rate is consistent with prior work (17, 23).

Parents were asked whether their infant was exposed to any language(s) other than English and the percentage of time another language was spoken to the infant. Approximately 70% of the infants were raised in a monolingual environment. The remaining infants, who were exposed to more than one language (at least 20% of the time), were evenly distributed across ages and experiments, and performed comparably to those exposed only to English. Infants with bilingual exposure had a variety of linguistic backgrounds including exposure to Spanish (~50%) and a variety of other languages. This variation in language exposure prevented us from identifying effects of monolingual versus bilingual exposure in this experiment.

Stimuli. Visual. Line-drawn images of dinosaurs and fish formed two eight-item familiarization sets and two test pairs. Within each familiarization set,

images varied in color; within each test pair, images were matched in color. Images (~15 cm²) were projected onto a white screen ~75 cm from the infant's eyes.
Auditory. A vocalization of a blue-eyed Madagascar lemur (*E. macaco flavifrons*) and a segment of backward speech [a labeling phrase from ref. 17, played in reverse], each ~2.2 s, were played from a hidden speaker, located 56 cm below the center of the screen. (Audio Files S1 and S2).

Procedure. Infants were seated on a parent's lap, facing a wooden puppet stage (located 93 cm above the floor; the front opening was 61 cm high × 106 cm wide). The visual images were projected onto a screen resting on the stage floor. The left and right positions of the projected images were separated by 11 cm. Parents were instructed not to influence their infant's attention. Infant behavior was recorded by a video camera (hidden 12 cm below the screen). During the task, two trained coders, blind to condition assignment, recorded infants' visual attention; reliability between these coders was 90%.
Familiarization trials. Visual stimuli (either dinosaurs or fish) were presented on alternating sides of the screen (20 s each). Auditory stimuli (either lemur vocalization or backward speech) were presented as each image appeared and repeated 8 s later.
Test trials. Each test pair appeared side-by-side, in silence, and remained visible until the infant accumulated 10 s of looking at the test images. [Note: Three infants in each experiment, who looked exclusively at one test image for 10 s, were presented with a second test trial with new images. In accordance with previous implementations of this paradigm (17), the analyses reported here include these infants' performance on the second test trial; the same pattern of results emerges, whether or not these infants are included.] Across infants, left/right positions of the first familiarization image and the test images were counterbalanced.

Coding. Infants' looking time at test served as our dependent measure. For each infant, we calculated a preference score (looking time at novel image/looking time at both images). Test trials were coded offline using frame-by-frame software (34). Reliability between two trained observers, blind to condition, was 91%.

Infants' looking time during familiarization (coded online) revealed no differences across age or experimental condition ($P > 0.05$). This provides assurances that any differences between experiments cannot be attributed to differences in attention to the stimuli during familiarization.

1. Vouloumanos A, Hauser MD, Werker JF, Martin A (2010) The tuning of human neonates' preference for speech. *Child Dev* 81(2):517–527.
2. Kuhl PK (2004) Early language acquisition: Cracking the speech code. *Nat Rev Neurosci* 5(11):831–843.
3. Mehler J, et al. (1988) A precursor of language acquisition in young infants. *Cognition* 29(2):143–178.
4. Shultz S, Vouloumanos A (2010) Three-month-olds prefer speech to other naturally occurring signals. *Lang Learn Dev* 6(4):241–257.
5. Vouloumanos A, Werker JF (2004) Tuned to the signal: The privileged status of speech for young infants. *Dev Sci* 7(3):270–276.
6. Werker JF, Tees RC (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 7(1):49–63.
7. Werker JF, Yeung HH, Yoshida KA (2012) How do infants become experts at native-speech perception? *Curr Dir Psychol Sci* 21(4):221–226.
8. Di Giorgio E, Leo I, Pascalis O, Simion F (2012) Is the face-perception system human-specific at birth? *Dev Psychol* 48(4):1083–1090.
9. Pascalis O, de Haan M, Nelson CA (2002) Is face processing species-specific during the first year of life? *Science* 296(5571):1321–1323.
10. Scott LS, Monesson A (2009) The origin of biases in face perception. *Psychol Sci* 20(6):676–680.
11. Lewkowicz DJ, Ghazanfar AA (2009) The emergence of multisensory systems through perceptual narrowing. *Trends Cogn Sci* 13(11):470–478.
12. Lewkowicz DJ, Leo I, Simion F (2010) Intersensory perception at birth: Newborns match nonhuman primate faces and voices. *Infancy* 15(1):46–60.

13. Pons F, Lewkowicz DJ, Soto-Faraco S, Sebastián-Gallés N (2009) Narrowing of intersensory speech perception in infancy. *Proc Natl Acad Sci USA* 106(26):10598–10602.
14. Vouloumanos A, Druhen MJ, Hauser MD, Huizink AT (2009) Five-month-old infants' identification of the sources of vocalizations. *Proc Natl Acad Sci USA* 106(44):18867–18872.
15. Lorenz K (1971) *Studies in Animal and Human Behaviour* (Harvard Univ Press, Cambridge, MA), Vol I.
16. Sugita Y (2008) Face perception in monkeys reared with no exposure to faces. *Proc Natl Acad Sci USA* 105(1):394–398.
17. Ferry AL, Hespos SJ, Waxman SR (2010) Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Dev* 81(2):472–479.
18. Gelman SA, Meyer M (2011) Child categorization. *Wiley Interdiscip Rev Cogn Sci* 2(1):95–105.
19. Medin DL, Rips LJ (2005) Concepts and categories: Memory, meaning, and metaphysics. *The Cambridge Handbook of Thinking and Reasoning* (Cambridge University Press, Cambridge, UK), pp 37–72.
20. Pinker S, Jackendoff R (2005) The faculty of language: What's special about it? *Cognition* 95(2):201–236.
21. Aslin RN (2007) What's in a look? *Dev Sci* 10(1):48–53.
22. Balaban MT, Waxman SR (1997) Do words facilitate object categorization in 9-month-old infants? *J Exp Child Psychol* 64(1):3–26.
23. Fulkerson AL, Waxman SR (2007) Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition* 105(1):218–228.

24. Colombo J (2002) Infant attention grows up: The emergence of a developmental cognitive neuroscience perspective. *Curr Dir Psychol Sci* 11(6):196–200.

25. Roder BJ, Bushnell EW, Sasseville AM (2000) Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy* 1(4):491–507.

26. Trehub SE (2001) Musical predispositions in infancy. *Ann N Y Acad Sci* 930(1):1–16.

27. Zentner MR, Kagan J (1998) Infants' perception of consonance and dissonance in music. *Infant Behav Dev* 21(3):483–492.

28. Peña M, et al. (2003) Sounds and silence: an optical topography study of language recognition at birth. *Proc Natl Acad Sci USA* 100(20):11702–11705.

29. Fernald A (1985) Four-month-old infants prefer to listen to motherese. *Infant Behav Dev* 8(2):181–195.

30. Palmer SB, Fais L, Golinkoff RM, Werker JF (2012) Perceptual narrowing of linguistic sign occurs in the 1st year of life. *Child Dev* 83(2):543–553.

31. Weikum WM, et al. (2007) Visual language discrimination in infancy. *Science* 316(5828):1159.

32. Kuhl PK, Tsao FM, Liu HM (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc Natl Acad Sci USA* 100(15):9096–9101.

33. Sebastián-Gallés N, Albareda-Castellot B, Weikum WM, Werker JF (2012) A bilingual advantage in visual language discrimination in infancy. *Psychol Sci* 23(9):994–999.

34. The Language Archive (2007) *ELAN* (Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands).

PSYCHOLOGICAL AND COGNITIVE SCIENCES