The multisensory representation of number in infancy

Kerry E. Jordan* and Elizabeth M. Brannon

Center for Cognitive Neuroscience and Department of Psychological and Brain Sciences, Duke University, Box 90999, Durham, NC 27708

Edited by Charles R. Gallistel, Rutgers, The State University of New Jersey, Piscataway, NJ, and approved January 7, 2006 (received for review September 16, 2005)

Human infants can discriminate visual and auditory stimuli solely on the basis of number, suggesting a developmental foundation for the nonverbal number representations of adult humans. Recent studies suggest that these language-independent number representations are multisensory in both adult humans and nonhuman animals. Surprisingly, however, previous studies have yielded mixed evidence concerning whether nonverbal numerical representations independent of sensory modality are present early in human development. In this article, we use a paradigm that avoids stimulus confounds present in previous studies of cross-modal numerical mapping in infants. We show that 7-month-old infants preferentially attend to visual displays of adult humans that numerically match the number of adult humans they hear speaking. These data provide evidence that by 7 months of age, infants connect numerical representations across different sensory modalities when presented with human faces and voices. Results support the possibility of a shared system between preverbal infants and nonverbal animals for representing number.

cognitive development | comparative cognition | multisensory processes | numerical cognition

Adult humans easily recognize the numerical equivalence between two objects they see and two sounds they hear. Even the nonverbal number representations of adults and children are not tied to the sensory modality in which they were originally perceived (1, 2). Human infants and nonhuman animals have been shown to discriminate number in both the auditory and visual sensory modalities (3–6). Field and laboratory studies demonstrate that nonhuman animals also connect these number representations in the auditory and visual modalities (7–11). For example, rats sum across sounds and sights, and monkeys spontaneously match the number of conspecific faces they see to the number of conspecific voices they hear (8, 9). If nonhuman primates, human infants, and human adults rely on a shared, developmentally and evolutionarily ancient system for representing number nonverbally, then human infants should also detect numerical equivalences across sensory modalities.

Studies investigating the ability of human infants to map numerical representations across sensory modalities, however, have yielded mixed results (12–16). Starkey *et al.* (12, 13) first used a preferential looking method to examine whether infants would spontaneously look at a visual stimulus that numerically matched an auditory stimulus. Research in fields other than the study of numerical cognition has already shown that infants look preferentially toward visual stimuli that correspond to a soundtrack (17–19); for example, when infants hear a specific speech sound, they look preferentially at a face that articulates that speech sound compared with a face that articulates a different speech sound (18). Starkey *et al.* (12, 13) presented infants with side-by-side slides of two or three household objects while an experimenter who was out of the infants' view hit a drum two or three times. Infants preferentially looked toward the visual display that numerically matched the number of drumbeats they heard. Unfortunately, when other researchers varied the parameters (such as the rate and duration of the tones) that often covary with number, or even the identity of the visual objects displayed, they had difficulty replicating these results (14, 15). In some cases, infants had no preference for the matching visual array, whereas in other cases, they preferred to look at the nonmatching array.

A more recent study by Kobayashi and colleagues (16) reported that 6-month-old infants represented the numerical equivalence between objects and sounds when the visual and auditory stimuli were given a natural, explicit relationship in a violation of expectation procedure. Infants were first familiarized with a digitized presentation of two and three fully visible Mickey Mouse-like objects sequentially impacting a surface, with each object emitting a tone at impact. Infants were then tested during trials in which an occluder blocked the infants' view, but the infants heard two or three of the tones from familiarization (varied in rate and total sequence duration). When the occluder was removed, two or three of the Mickey Mouse-like objects were revealed. Infants looked significantly longer at the numerically nonequivalent events, suggesting that they had formed an expectation of how many objects they should see based on how many impact sounds they had heard and were surprised that this expectation was violated. As in studies testing infants' cross-modal numerical knowledge, however, this study used a within-subject design in which each infant heard two and three sound stimuli and saw two and three element arrays in the familiarization and/or test trials of a single experimental session. In this way, infants could have compared stimuli on the basis of nonnumerical attributes (e.g., intensity of stimulation) across the multiple trials each infant experienced. It is possible that infants have a natural proclivity to match the more intense of two sounds with the more intense of two sights (20, 21). In addition, because individual auditory stimulus (tone) duration was never varied, the total duration of sound presented in a two-tone-sequence trial was always less than the total duration of sound presented in a three-tone-sequence trial. It is conceivable, therefore, that infants could have associated the visual stimuli having a greater surface area with the auditory stimuli that had longer total sound duration and looked longer when the outcome was in violation of this match.

In sum, although some evidence suggests that infants may numerically map across sensory modalities, critical stimulus controls have not yet been implemented. Furthermore, previous studies used within-subject designs, providing infants an opportunity to learn within the experimental session about nonnumerical relations across modalities or to compare relative intensity and express an untrained tendency to match the more intense of two sounds with the more intense of two sights.

In this article, we ask whether multisensory representation is a fundamental part of infants' numerical knowledge or whether this level of abstraction highlights a unique developmental discontinuity between human infant and adult number representation. We examine whether 7-month-old infants, like rhesus monkeys, spontaneously match the number of entities they see with the number of events they hear. We perform this study by testing whether infants preferentially attend to dynamic visual displays of two or three women that numerically

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

*To whom correspondence should be addressed. E-mail: kej8@duke.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0508107103

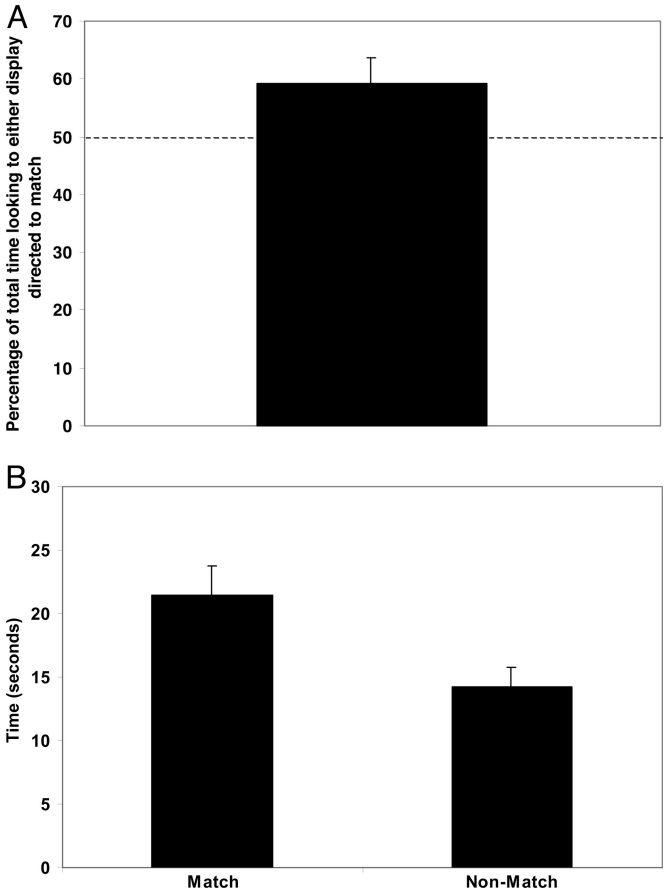

Fig. 1. Infants can match the number of faces they see with the number of voices they hear. (*A*) Mean percentage of total looking time to the matching video display; chance is 50%. On average, infants spent a significantly greater proportion of time looking at the display that matched the number of voices they heard. (*B*) Mean duration of looking time to the match versus nonmatch displays.

match the number of voices they hear simultaneously speaking the word "look." There are three crucial aspects of our experimental design that are unique for studies on cross-modal numerical perception in infants. First, the to-be-enumerated items were temporally and spatially synchronous sounds and sights, which eliminated cues such as duration or rate. Second, the sounds and sights used were both ecologically relevant and nonarbitrarily related (16, 22). Third, a between-subject design was used, which prevented learning or anchoring on the basis of intensity. By using a testing procedure modeled after a previous study conducted with rhesus monkeys (7), we were also able to directly compare results between two populations of nonlinguistic organisms.

Results

Infants spent a greater proportion of time looking at the display that numerically matched the number of women they heard compared with the numerically nonmatching display. They directed 59.2% of the total time they spent looking at either display to the matching display, which differed significantly from chance [t (19) = 2.09; $P < 0.05$] (Fig. 1*A*). On average, infants looked at the matching display for 21.5 ± 2.2 s and the nonmatching display for 14.2 ± 1.5 s (Fig. 1*B*). This difference was significant [t (19) = 2.64; $P < 0.02$]. A two (match vs. nonmatch) × two (two sounds vs. three sounds) × three (missing individual 1, 2, or 3 in the two-woman video) ANOVA revealed no other main effects or interactions. Thus, the effect held both for the infants who heard two women (average looking time to match, 20.9 s; average looking time to nonmatch, 15.8 s) and for the infants who heard three women (average looking time to match, 22.1 s; average looking time to nonmatch, 12.6 s). Finally, 14 of 20 infants tested looked longer at the matching display than at the nonmatching display ($P < 0.039$; sign test).

Table 1. Parallels between the behavior of rhesus monkeys and human infants when tested in a preferential looking paradigm on matching the number of faces with the number of voices

	Rhesus monkeys	Human infants
n	20	20
LT to either display directed to match, %	60	59.2
Average LT to match, s	14.2	21.5
Average LT to nonmatch, s	9.2	14.2
No. of subjects looking longer to match	15	14

Data from rhesus monkeys taken from ref. 7. LT, looking time.

Discussion

Our results demonstrate that by 7 months of age, infants can represent the equivalence between the number of voices they hear and the number of faces they see. The parallel between infants' and rhesus monkeys' performance on this task is particularly striking (Table 1). The present study tested a sample size of 20 infants and found that they directed 59.2% of the total time they spent looking at either display to the numerically matching display; Jordan and colleagues (7) tested 20 rhesus macaques and found that they directed 60.0% of the total time they spent looking at either display to the matching display. Both populations looked significantly longer at the matching display than the nonmatching display, and in both species, the effect held for individuals who heard two sounds or three sounds. Finally, in both populations, a significant number of individuals tested looked longer at the matching display than at the nonmatching display (15 of 20 monkeys and 14 of 20 infants). Although the neurobiological underpinnings of multisensory numerical perception remain unknown, these clear behavioral parallels argue for the possibility of a shared neurobiological substrate in two disparate populations of nonlinguistic organisms.

Importantly, multiple aspects of the present study's experimental design differ from previous studies that tested infants' abilities to match numerosities across sensory modalities. First, by presenting auditory stimuli simultaneously and equating the duration of the composite audio streams and videos, we avoided the possibility that infants could use rate or duration as a basis for judgment, a potential confounding factor in past studies. This procedure also eliminated other amodal cues such as synchrony. Second, infants could not have detected the multisensory numerical correspondence in the present study by learning to match the more intense or more complex auditory stimulus to the more intense or more complex visual stimulus. Because previous studies used within-subject designs in which infants heard two and three sound stimuli and saw two and three element arrays in a single experimental session, positive results could have been due to an infant's proclivity to match the more intense or complex stimuli in each modality. In the present study, we tested each participant on a single two or three-sound trial (between-subject design), and therefore, participants had no opportunity to compare the intensity or complexity of various auditory stimuli. Third, controlling for auditory cues that often covary with number made it unlikely that infants could use *a priori* expectations to map two (or three) sounds to a continuous property of the visual stimulus (i.e., when one hears speech at

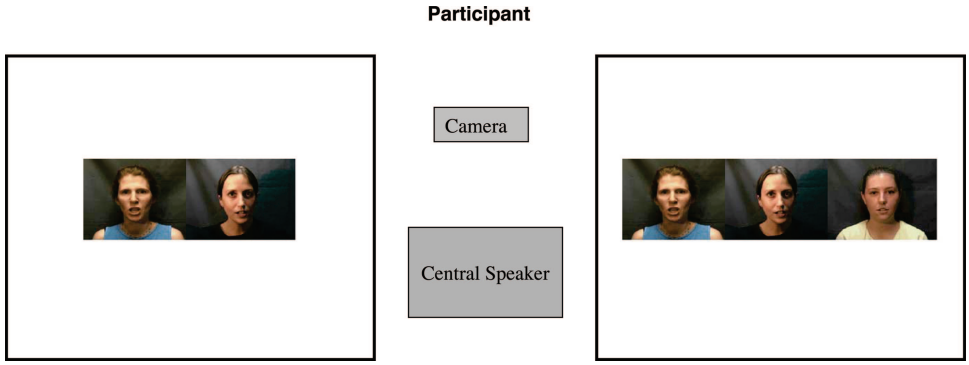

Fig. 2. Schematic illustrating the experimental setup and sample presentation of stimuli during a trial. Participants sat in front of two video screens that played concurrent movies synchronized with a single soundtrack from a central speaker. Visual and auditory stimuli were 0.5 s in duration and looped for 60 s/trial. The infant's looking time to each movie was recorded. Still frames for this diagram were extracted from the videos by using Adobe PREMIERE 6.0.

this amplitude, it is usually paired with a certain surface area of face). Finally, because the women in the stimulus videos were unfamiliar to the participants, infants could not have based their matches on the presence or absence of a face that was associated with a voice they heard and recognized. Thus, our data provide the solid evidence that human infants can spontaneously match numerosities across divergent sensory modalities by the age of 7 months.

It is also important to note that the nonarbitrary connection between the visual stimuli of human faces and auditory stimuli of human voices differs dramatically from the relationship between the types of stimuli used in previous experiments concerning cross-modal representations of number in human infants (12–15), even though previous studies testing a human infant's visual numerical knowledge have sometimes used ecologically relevant contexts (23). Except for a study by Kobayashi and colleagues (16), previous cross-modal studies with human infants have not allowed a natural connection between the auditory and visual stimuli presented to the infants. Matching three drumbeats to a visual display of three unrelated objects (such as household items or black dots) is in fact quite difficult because even 3-year-old children failed to perform a similar task that required matching the number of hand claps heard to the number of black dots seen [see ref. 24; however, preschool children succeed in adding and comparing numerosities across modalities (2)]. Our use of voices and faces may also have made the situation more ecologically relevant and meaningful to the infant, along with making the intermodal correspondence more salient (e.g., ref. 22). It remains to be determined whether infants can detect the numerical correspondence between other types of nonarbitrarily related sights and sounds (for example, match the number of sticks seen striking a drum to the number of drumbeats heard), which have less obvious ecological relevance.

Future studies should also attempt to pinpoint the format of the numerical representations underlying behavior in this task. Previous research suggests that infants' numerical representations are modulated by two different systems: (*i*) an analog magnitude system that represents number approximately and obeys Weber's Law in that discrimination between two quantities depends on their ratio and (*ii*) an object file system that tracks and represents small numbers of individual objects precisely (3, 25, 26). Because the numerical values tested are within the capacity of either system, future studies are needed to determine whether infants can cross-modally match larger numerical values and whether matching depends on the ratio between the two numerical values compared.

Data from the present study suggest two important conclusions. First, when stimulus attributes that often covary with number are properly controlled, human infants tested in a between-subject design recognize the numerical correspondence between two or three dynamic faces and two or three concurrent voices. Second, in contrast to adults, infants may need nonarbitrarily related auditory and visual stimuli to detect such an equivalence between modalities. Future research will be necessary to determine whether infants can succeed at numerically matching arbitrarily related stimuli across sensory modalities or whether this is a difference between the numerical abilities of infants and adults. In summary, our results provide clear evidence of a developmental basis for language-independent numerical representations that extend across different sensory

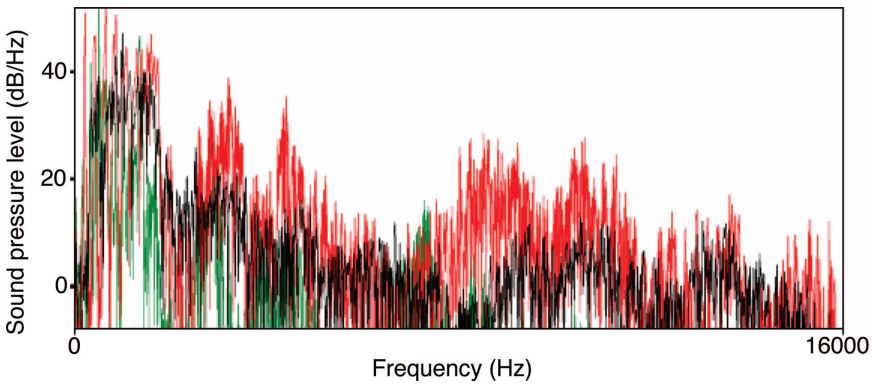

Fig. 3. Power spectra from sample auditory stimuli. Illustrated is a stimulus containing three women's voices playing concurrently, with each color representing a different individual's voice. The spectra were created by using PRAAT 4.2.

modalities. This spontaneous matching of numerosities across the visual and auditory modalities supports the contention that human infants, human adults, and nonhuman primates share at least one common nonverbal numerical representational system.

Methods

Twenty 7-month-old infants (age range, 6 months 19 days to 7 months 26 days), 12 of whom were male, participated in this study. No infants were excluded from the final sample. Informed consent from the parent of each participant was obtained before starting the experiment.

The stimuli were individual digital video recordings of three distinct adult women between the ages of 25–35 speaking the word "look". All three women were unfamiliar to the infants and were not involved in their testing. Individual videos were then acquired onto a computer and manipulated in Adobe PREMIERE 6.0 (San Jose, CA) to make composite videos of two and three individuals. Each video was edited for duration so that the onset and offset of the two or three women's mouth movements were synchronous in each composite video. We extracted the auditory stimuli from the digital video samples. Auditory stimuli were sampled at 32 kHz and normalized to the peak amplitude, and then two of the three auditory stimuli were temporally expanded to match the stimulus of the longest duration. We constructed two- or three-stimulus auditory tracks by mixing them down in Adobe AUDITION 1.0 and then equating their average rms power. Composite auditory stimuli were equated for amplitude. The fundamental frequencies of the stimuli did not overlap. Because all visual and auditory components were identical in duration and synchronized, the participants could not use cues such as rate to make a match (for a sample composite video, see Movie 1, which is published as supporting information on the PNAS web site).

In each session, participants were presented with one trial consisting of unfamiliar adult women mouthing the word "look"; one video showed two women and one video showed three women. Infants heard two or three women concurrently saying "look". The "two" versus "three" composite video stimuli were played simultaneously on side-by-side 40-cm liquid crystal display monitors. The two composite videos contained two common individuals such that the two-women display was a subset of the three-women display (Fig. 2). The individual identities of the two-women display were counterbalanced across participants, as was the left-right position of the two- and three-women composite video displays. Audio tracks, synchronized with both composite videos, were played through a hidden speaker placed directly between and slightly behind the monitors (Fig. 3). A REALBASIC program (REAL Software, Austin, TX) was used to play the video and audio stimuli in synchrony.

Infants were seated on a parent's lap in front of the two monitors at a distance of 72 cm. Parents were instructed to keep their eyes closed and to refrain from interacting with their infant, beside holding them. The monitors were 56 cm apart (center-to-center distance) and at eye-level with the infants. All trials were videotaped by using a microcamera placed above and between the monitors. All equipment was concealed by a thick black curtain, except for the monitor screens and the lens of the camera. The experimenter monitored infant activity remotely. A session began when the infant looked centrally. A trial consisted of the two 0.5-s videos played in a continuous loop for 60 s, with one of the two 0.5-s sounds also playing in a loop through the central speaker. Each participant was only tested once, and all trials were recorded on videotape. No familiarization or reward was provided.

We collected high-quality, close-up videos of the participants' behavior with a microcamera that fed directly into a video cassette recorder. Videos were digitized and acquired at 30 frames per s (frame size, 720 $\times$ 480 pixels) onto a computer. Clips for analysis were edited down to 60 s, starting with the onset of the auditory track. The total duration of a participant's time spent looking at each composite video (left or right) was assessed. The screens were far apart in the horizontal dimension, fairly close to the infant's face, and at eye level. Thus, the infant had to make large eye/head movements to look to one screen or the other, and this setup made scoring the direction of the look unambiguous. To assess reliability of the look direction measurements, 50% of the trials were scored by a second observer blind to the experimental condition; inter-observer reliability was 0.99 ($P < 0.0001$) as measured by a Pearson r test.

We thank Evan Maclean for help with writing experimental programs; Klaus Libertus, Melissa Libertus, Umay Suanda, and all members of the Duke Infant Cognition Center for help collecting data and feedback on the experimental design; Yukyung Jung and Sweta Saxena for help with coding the videos; Marc Hauser for extensive feedback on the experimental design and an earlier draft of the manuscript; and Asif Ghazanfar for discussions that led to the experimental design. This work was supported by a National Science Foundation (NSF) Graduate Fellowship (to K.E.J.) and National Institute of Mental Health Grant R01 MH066154, an NSF Research on Learning and Education/Developmental Learning Sciences Award 0132382, an NSF CAREER award, and a Merck Scholars award (to E.M.B.).

1. Barth, H., Kanwisher, N. & Spelke, E. (2003) *Cognition* **86,** 201–221.
2. Barth, H., La Mont, K., Lipton, J. & Spelke, E. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 14116–14121.
3. Brannon, E. & Roitman, J. (2003) in *Functional and Neural Mechanisms of Interval Timing*, ed. Meck, W. (CRC, New York), pp. 143–182.
4. Feigenson, L., Dehaene, S. & Spelke, E. (2004) *Trends Cognit. Sci.* **8,** 307–314.
5. Brannon, E. M. & Terrace, H. S. (1998) *Science* **282,** 746–749.
6. Hauser, M. D., Tsao, F., Garcia, P. & Spelke, E. S. (2003) *Proc. Biol. Sci.* **270,** 1441–1446.
7. Jordan, K., Brannon, E. M., Logothetis, N. K. & Ghazanfar, A. A. (2005) *Curr. Biol.* **15,** 1034–1038.
8. Church, R. & Meck, W. (1984) in *Animal Cognition*, eds. Roitblat, H. L., Bever, T. G. & Terrace, H. S. (Erlbaum, Hillsdale, NJ), pp. 445–464.
9. McComb, K., Packer, C. & Pusey, A. (1994) *Anim. Behav.* **47,** 379–387.
10. Wilson, M. L., Hauser, M. D. & Wrangham, R. W. (2001) *Anim. Behav.* **61,** 1203–1216.
11. Kitchen, D. M. (2004) *Anim. Behav.* **67,** 125–139.
12. Starkey, P., Spelke, E. & Gelman, R. (1983) *Science* **222,** 179–181.
13. Starkey, P., Spelke, E. & Gelman, R. (1990) *Cognition* **36,** 97–127.
14. Moore, D., Benenson, J., Reznick, J., Peterson, M. & Kagan, J. (1987) *Dev. Psychol.* **23,** 665–670.
15. Mix, K., Levine, S. & Huttenlocher, J. (1997) *Dev. Psychol.* **33,** 423–428.
16. Kobayashi, T., Hiraki, K. & Hasegawa, T. (2005) *Dev. Sci.* **8,** 409–419.
17. Spelke, E. (1976) *Cognit. Psychol.* **8,** 533–560.
18. Kuhl, P. & Meltzoff, A. (1982) *Science* **218,** 1138–1141.
19. Patterson, M. L. & Werker, J. F. (2002) *J. Exp. Child Psychol.* **81,** 93–115.
20. Rose, S. & Ruff, H. (1987) in *Handbook of Infant Development,* ed. Osofsky, J. (Wiley, New York), pp. 318–362.
21. Turkewitz, G., Gardner, J. & Lewkowicz, D. (1984) in *Conference on Levels of Integration and Evolution of Behavior*, eds. Greenberg G. & Tobach, E. (Erlbaum, Hillsdale, NJ), pp. 167–195.
22. Hauser, M. D. (2000) *Wild Minds: What Animals Really Think* (Henry Holt, New York).
23. Feigenson, L., Carey, S. & Hauser, M. (2002) *Psychol. Sci.* **13,** 150–156.
24. Mix, K., Huttenlocher, J. & Levine, S. (1996) *Child Dev.* **67,** 1592–1608.
25. Hauser, M. D. & Spelke, E. (2004) in *The Cognitive Neurosciences*, ed. Gazzaniga, M. S. (MIT Press, Cambridge, MA), pp. 853–864.
26. Hauser, M. & Carey, S. (1998) in *The Evolution of Mind*, eds. Cummins Dellarosa, D. & Allen, C. (Oxford Univ. Press, London), pp. 51–106.