


Electrophysiological Evidence for the Understanding of Maternal Speech by 9-Month-Old Infants

Eugenio Parise and Gergely Csibra

Central European University

Psychological Science
 XX(X) 1–6
 © The Author(s) 2012
 Reprints and permission:
sagepub.com/journalsPermissions.nav
 DOI: 10.1177/0956797612438734
<http://pss.sagepub.com>


Abstract

Early word learning in infants relies on statistical, prosodic, and social cues that support speech segmentation and the attachment of meaning to words. It is debated whether such early word knowledge represents mere associations between sound patterns and visual object features, or reflects referential understanding of words. By measuring an event-related brain potential component known as the N400, we demonstrated that 9-month-old infants can detect the mismatch between an object appearing from behind an occluder and a preceding label with which their mother introduces it. Differential N400 amplitudes have been shown to reflect semantic priming in adults, and its absence in infants has been interpreted as a sign of associative word learning. By setting up a live communicative situation for referring to objects, we demonstrated that a similar priming effect also occurs in young infants. This finding may indicate that word meaning is referential from the outset of word learning and that referential expectation drives, rather than results from, vocabulary acquisition in humans.

Keywords

infants, language acquisition, semantic priming, ERP, N400, infant development, language development, priming

Received 7/19/11; Revision accepted 1/10/12

Word learning in infancy is supported by a range of cognitive skills. By at least 8 months of age, infants are able to use statistical information to segment words from continuous speech (Saffran, Aslin, & Newport, 1996), and they rely on prosody both in word segmentation and in associating novel words with visual referents (Shukla, White, & Aslin, 2011). Mapping words onto semantic representations by matching the word form to a referent is a nontrivial problem, and there is no agreement on whether young infants are capable of referential word understanding that is genuinely semantic (Waxman & Gelman, 2009). Evidence suggests that infants conceive deictic signals, such as gazing and pointing, as having referential intent (e.g., Senju & Csibra, 2008), and they exploit such signals in word learning (e.g., Hollich et al., 2000). Furthermore, 1-year-olds expect that concurrent verbal and nonverbal expressions from the same source refer to the same object, which suggests that they appreciate the referential nature of both pointing and words (Gliga & Csibra, 2009). However, some researchers have argued that early word understanding reflects simple associations between auditory and visual stimuli and not an appreciation of the referential and symbolic nature of words (Robinson, Howard, & Sloutsky, 2005).

Whether infants form stable word-object associations that reflect referential understanding of speech is not yet known.

One way to test such understanding in infants is by measuring event-related potentials (ERPs), such as the N400 component. The N400 ERP component—a negative-going waveform that peaks approximately 400 ms after the onset of a stimulus—has been shown to reflect semantic priming in adults (Kutas & Hillyard, 1980). Friedrich and Friederici (2004, 2005a, 2005b) found that adults and 19- and 14-month-olds, but not 12-month-olds, produced an enhanced N400 amplitude to words that were incongruous with picture primes. In a follow-up study, only ERPs from 12-month-olds who were assessed as high word producers showed an N400 effect (Friedrich & Friederici, 2010). The absence of an N400 effect in infants younger than 12 months was interpreted as a sign of merely associative word understanding, which does not entail semantic processing. Word learning requires the learner to form associations between word forms and visual referents. However, temporal associations alone do not imply semantic representation of word meaning. In the present study, we attempted to clarify exactly what infants learn when they form

Corresponding Author:

Eugenio Parise, Cognitive Development Center, Central European University,
 Hattyú utca 14, 1015 Budapest, Hungary
 E-mail: eugenio.parise@tiscali.it

word-object associations: meaningless links or semantically meaningful sign-referent relations.

To shed light on whether young infants interpret words referentially and semantically, we developed a new paradigm that overcomes some shortcomings of earlier ERP studies with infants. In particular, we wanted to ensure that infants realized that the word they heard referred to the object they saw. Because both theoretical arguments (Csibra, 2010) and empirical findings (e.g., Senju, Csibra, & Johnson, 2008) suggest that referential expectation is primarily elicited by ostensive communicative contexts in infants, we modified the picture-word priming paradigm (Friedrich & Friederici, 2004) to provide an optimal stimulus environment for 9-month-olds. First, we presented infants with live rather than prerecorded speech, which was occasionally accompanied by nonverbal referential gestures, such as pointing and gazing toward upcoming objects. Second, we used the words as primes and the objects as probes because a known word is more likely to activate the associated semantic representation in preverbal infants than a picture of an object is (cf. Xu, 2007). Third, instead of flashing a picture on a monitor, the object appeared in a dynamic fashion from behind an occluder on a computer screen. Thus, on the one hand, live speech and interaction ensured optimal conditions for speech processing of the prime (Kuhl, 2007), and on the other hand, video presentation of the target objects made it possible to accurately control stimulus variables for measuring ERPs. Finally, because young infants display a preference for maternal speech (e.g., Cooper & Aslin, 1989), we used the mother as the speaker in one of our two conditions. As auditorily presented words have rarely been used as primes for visual probes in ERP studies (but see Nigam, Hoffman, & Simons, 1992, for a different paradigm), we also tested whether, using this paradigm, we could find a reliable N400 effect in adults.

Method

Participants

Twelve adults (6 females, 6 males; average age = 38 years 8 months, range = 25 to 56 years) who were native speakers of

Hungarian and 28 infants raised in Hungarian families participated in the study. Infants were assigned to two conditions: 14 to the mother-speech condition (5 females, 9 males; average age = 277 days, range = 269 to 286 days) and 14 to the experimenter-speech condition (6 females, 8 males; average age = 278 days, range = 266 to 285 days). One additional adult was excluded because of poor electroencephalogram (EEG) impedance, and 21 additional infants were excluded because of fussiness ($n = 11$), insufficient number of trials ($n = 6$), extensive body movements ($n = 2$), poor impedance ($n = 1$), and maternal interference ($n = 1$). (See the Supplemental Material available online for further participant details.)

Stimuli

Using data collected from parental reports, we selected 15 object labels that two thirds of Hungarian 1-year-old infants recognize. We used corresponding pictures of objects to create 15 animated video clips showing an occluder dropping to reveal one of the objects (Fig. 1; see Table S1 and Stimuli in the Supplemental Material for further details).

Procedure

Adults sat 70 cm in front of a CRT monitor. In each trial, they heard a prerecorded word (the name of 1 of the 15 objects) from a loudspeaker behind the monitor while a dynamic fixation stimulus was presented on top of an occluder. The duration of the auditory stimulus was between 419 and 784 ms (average = 559 ms). After the auditory stimulus ended, the fixation stimulus stopped moving, and the display remained frozen for 600 to 800 ms. Then the fixation stimulus disappeared, and the occluder started to fall forward, revealing an object behind it. This phase lasted for 480 ms. The object was fully visible for 1,000 ms before the occluder began to rise, hiding the object again. This was followed by an intertrial interval lasting 1,100 to 1,300 ms. Participants were presented with 240 trials in 4 blocks. In half of the trials, the object corresponded to the preceding auditory word (congruous trials); in the other half, it did not (incongruous trials). Congruous and incongruous trials were presented equiprobably in pseudorandom order.

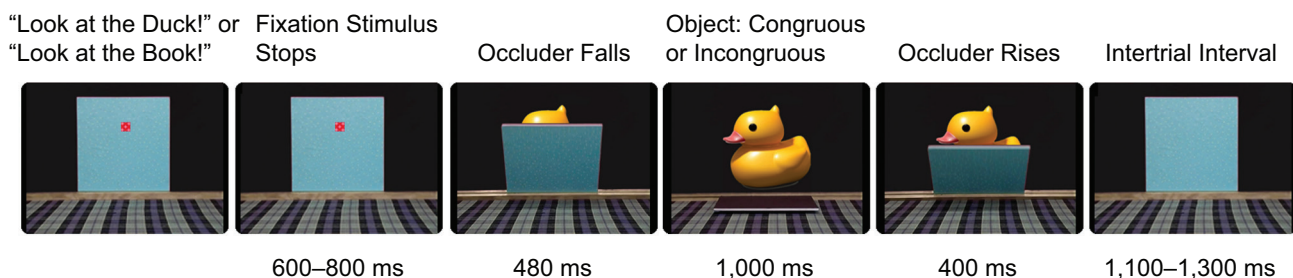


Fig. 1. Illustration of the trial sequence for infants. While a moving fixation stimulus was presented in front of an occluder, the infant's mother or an experimenter spoke a word or phrase that either named the object behind the occluder (congruous trial) or named some other object (incongruous trial). The fixation stimulus then stopped moving, and the display froze for 600 to 800 ms. After this, the fixation stimulus disappeared, and the occluder started to fall forward, revealing an object behind it. After 480 ms, the object was in full view, where it remained for 1,000 ms. Then the occluder began to rise, completely covering the object after 400 ms.

Infants sat on a high chair 70 cm in front of a CRT monitor. The infant's mother and an experimenter sat on chairs at either side of the infant. The trial sequence (Fig. 1) was the same as for adults, except that at the beginning of each trial, a word was presented over headphones either to the mother (mother-speech condition) or to the experimenter (experimenter-speech condition). She then repeated it (or uttered a phrase containing it) for the infant. Mothers were instructed to speak to the child as they would in everyday life, and they were allowed to gesture toward the monitor on which the occluder was seen. We asked them to utter the word at the very end of the phrase if they wanted to say more than just that word to their infant. In the experimenter-speech condition, the experimenter talked to the infant, attempting to reproduce the words, intonation, and style of a yoked mother from the mother-speech condition. In this design, the experimenter matched each individual mother from the mother-speech condition for an infant in the experimenter-speech condition.

Once the word was uttered and the infant attended the monitor, a second experimenter started a video clip revealing an object behind the occluder. Objects could be congruous or incongruous with the word prime, just as in the procedure for adult participants. Because the auditory stimulus was spoken live, the interstimulus interval between the word prime and the start of the visual stimulus varied across trials, averaging about 2,155 ms. Trials were presented as long as the infants were attentive. The position of the mother was counterbalanced across subjects in both conditions. The behavior of the infants was video-recorded throughout the session for off-line trial-by-trial editing and for additional behavioral scoring. (Further details of the procedure are reported in Experimental Setup in the Supplemental Material.)

EEG recording and analysis

Continuous EEG was recorded by 128-channel Geodesic Sensor Nets at a sampling rate of 500 Hz. The EEG was segmented into 1,700-ms epochs starting 200 ms before the occluder began to fall. We considered Time 0 the frame when the object started to become visible from behind the occluder (160 ms after motion onset). EEG segments were averaged to separate ERPs for word-congruous and word-incongruous objects, baseline-corrected to the first 200 ms of the segments, and rereferenced to the average reference. On the basis of previous reports that found the N400 component over parietal sites (Kutas & Hillyard, 1980), we identified as regions of interest the electrodes between C3 and P3 and between C4 and P4, over the left and right hemisphere, respectively (Fig. 2). On the basis of visual inspection of the grand averages and the existing literature on N400 latencies, we analyzed adults' mean ERP amplitude between 350 and 500 ms after the object appeared and infants' mean ERP amplitude between 500 and 650 ms after the object appeared (cf. Friedrich & Friederici, 2004, 2005b, for N400 time windows in adults and infants). (Further details of the EEG procedure are reported in EEG

Acquisition and EEG Data Reduction in the Supplemental Material.)

Results

Figure 2 shows grand-average ERP results for adults and infants. An analysis of variance (ANOVA) on the data from adults, with congruency (congruous object vs. incongruous object) and hemisphere (left vs. right) as within-subjects factors, revealed that incongruous objects elicited a more negative N400 amplitude than did congruous objects, $F(1, 11) = 7.50, p = .02, \eta_p^2 = .41$. Although the effect was bigger on the right side, the interaction between congruency and hemisphere was not significant. This result demonstrates that a semantic congruency effect on the N400 component can be elicited when an auditory word prime precedes an object image.

On average, infants contributed 20.3 congruous trials and 20.3 incongruous trials in the mother-speech condition, and 19.4 congruous trials and 19.9 incongruous trials in the experimenter-speech condition. An ANOVA on the infants' N400 amplitudes, with congruency and hemisphere as within-subjects factors and condition (mother speech vs. experimenter speech) as a between-subjects factor, revealed an interaction between congruency and condition, $F(1, 26) = 5.02, p = .03, \eta_p^2 = .16$. Two-way follow-up ANOVAs with congruency and hemisphere as factors revealed that this interaction was due to the fact that the N400 amplitude was more negative in response to the incongruous than to the congruous object in the mother-speech condition, $F(1, 13) = 5.45, p = .036, \eta_p^2 = .30$, but not in the experimenter-speech condition, $F(1, 13) = 0.29, p = .598, \eta_p^2 = .02$. A congruency-by-hemisphere interaction in the three-way ANOVA also indicated that the effect was more pronounced over the right hemisphere than over the left hemisphere, $F(1, 26) = 4.39, p = .05, \eta_p^2 = .15$.

To account for the difference in N400 amplitude between the conditions, we took five behavioral measures from infants: the number of times the speaker repeated the object name within a trial, whether the infant was looking to the mother or to the experimenter during the speech, and whether the infant looked to the mother or to the experimenter after seeing the object. Although we tried to match the mothers' behavior in the experimenter-speech condition, mothers uttered the prime word slightly, but significantly, more than the experimenter did (1.07 times vs. 1.01 times), $F(1, 26) = 13.67, p = .001, \eta_p^2 = .345$. As a control, we reran the statistics on ERPs that included only trials in which the word was uttered once in the mother-speech condition, and we obtained the same pattern of results as before. We did not find a significant difference between conditions in the number of times infants looked to the mother or to the experimenter. However, infants in the mother-speech condition tended to look more toward the mother than those in the other condition did, $F(1, 26) = 3.74, p = .064, \eta_p^2 = .126$. There was no effect of any factor on how many times infants looked at the mother or at the experimenter after having seen the object.

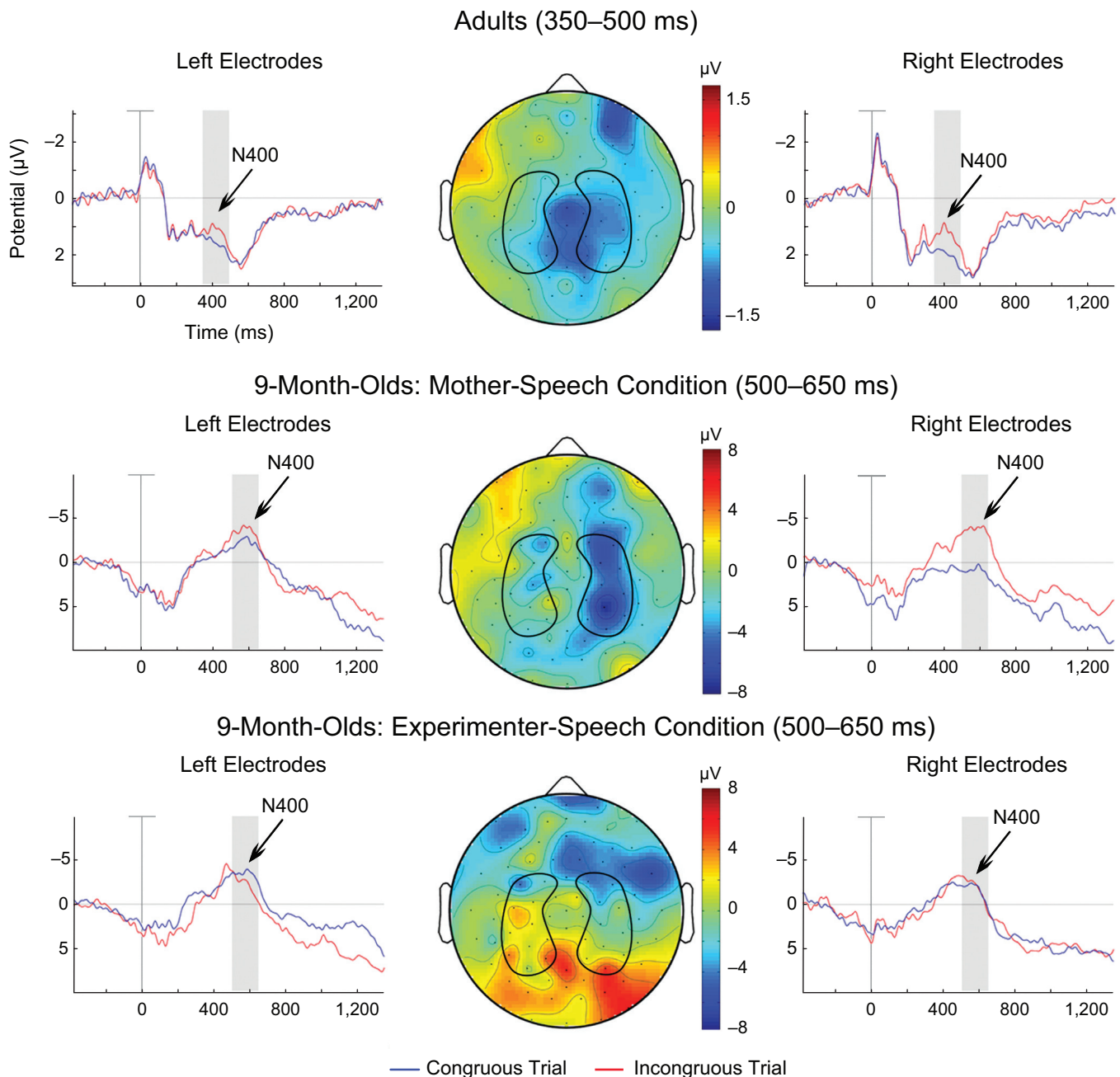


Fig. 2. Event-related potential (ERP) results. The figure shows grand-average waveforms on congruous and incongruous trials in left and right regions of interest (marked by black contours on the scalp maps). Results are shown separately for adults (top row), infants in the mother-speech condition (middle row), and infants in the experimenter-speech condition (bottom row). The gray shading indicates the time window of the N400 (350–500 ms in adults and 500–650 ms in infants), and the vertical line marks the time at which the object in each trial appeared from behind an occluder. The scalp maps depict the spatial distribution of the difference in ERP amplitude between incongruous and congruous trials in the given time windows.

(See Additional Analyses in the Supplemental Material for more information.)

Discussion

Our results suggest that infants as young as 9 months have a rudimentary receptive vocabulary. This has been previously suspected (Swingley, 2008) but difficult to prove. In agreement with our findings, the results of a recent eye-tracker

study by Bergelson and Swingley (2012) showed that 6- to 9-month-olds can follow their mother's verbal instructions and direct their gaze to objects. Six-month-old infants are also able to match novel words with arbitrary visual referents in a few trials by taking advantage of prosodic cues (Friedrich & Friederici, 2011; Shukla et al., 2011). However, the electrophysiological signs of semantic priming disappeared after 24 hr, which suggests strong limitations in memory processes (Friedrich & Friederici, 2011). Our study shows that

word-to-object priming occurs in 9-month-olds with familiar words when there is no requirement to learn new ones. Our method does not allow us to tell which infant understood which word, but infants in the mother-speech condition, as a group, seem to have activated the object features associated with the highly familiar words their mother uttered and matched them with the image that appeared on the screen in front of them. In this sense, infants in our study understood their mother's speech.

What is the nature of this understanding? In particular, is it possible that the word knowledge that infants displayed in this experiment goes beyond the formation of merely associative links between auditory and visual information, and reflects truly referential and semantic understanding of nouns? The kind of neuronal activation that we demonstrated here is correlational in nature and does not support unambiguous conclusions about the underlying processes. Nevertheless, three aspects of our results suggest that 9-month-olds appreciate the referential nature of words.

First, the N400 component is commonly interpreted to reflect semantic processing by exhibiting lower amplitude to stimuli semantically primed (rather than grammatically or associatively primed) by the context (Kutas & Federmeier, 2011). The differential N400 responses to congruent and incongruent object images thus indicate that the words activated neural processes in infants that are correlated with the extraction of meaning from stimuli. Second, we found these effects despite the relatively long delay (> 2 s on average) between the uttered word and the appearance of the object. Earlier studies, which used synchronous presentation (with the object being visible during the presentation of the word), should have had a better chance at demonstrating associative audiovisual links, but they failed to do so (e.g., Friedrich & Friederici, 2005b). Although pure associations could bridge temporal delays, they should work best with contiguous stimuli. In contrast, words can refer to absent referents, as they did in our study. Thus, finding differential brain activations for matches and mismatches between temporally separated stimuli supports the interpretation that a semantic, rather than an associative, link was formed between them.

Third, the success of this study was partly due to our effort to set up a situation in which infants had every reason to expect semantically interpretable referential expressions. Infants pay special attention to ostensive communicative signals, such as eye contact (Parise, Reid, Stets, & Striano, 2008) and their own name (Parise, Friederici, & Striano, 2010), as well as deictic referential signals, such as gaze (Hoehl, Wiese, & Striano, 2008) and pointing (Gredeback & Melinder, 2010), from early on. Ostensive communication has been suggested to generate referential expectation (Csibra, 2010) and expectation of coreference for concurrent referential signals (Gliga & Csibra, 2009). Although it is not clear why ostensive referential signals would facilitate associative processes, they could support the inference that the words infants hear are semantically

related to the object that would appear at the location referred to by the speaker. Thus, our paradigm, which closely resembles the natural joint-attention interaction in which infants at this age are regularly engaged with their parents (Bakeman & Adamson, 1984), provided an optimal environment for measuring the effect of semantic priming in young infants.

Our results suggest that adopting the mother as the speaker also contributed to this optimal environment. Newborns prefer their mother's voice to a stranger's voice (DeCasper & Fifer, 1980), and recent ERP research has confirmed that at the age of 4 months, infants respond faster and allocate more attention to their mother's voice than to an unfamiliar voice (Purhonen, Kilpeläinen-Lees, Valkonen-Korhonen, Karhu, & Lehtonen, 2005). In addition, the mother's voice elicits more activation in language-relevant brain areas in newborns than a stranger's voice does (Beauchemin et al., 2011). Our paradigm did not allow us to pinpoint the exact factors that made infants more responsive to the mother's than to the experimenter's communication. It could be that the familiar voice, intonation, or verbal expression (e.g., "Look, a duck!" vs. "Here comes the duck!") helped them to recognize the situation as a naming game. It is also possible that 9-month-old infants had difficulties in recognizing the target word in the slightly different phonetic production by the experimenter. Because the speaker sat next to the infants and did not always make eye contact with them during her speech, they might not have recognized from the experimenter's intonation alone that they were being addressed; however, the mother's voice alone could have achieved this effect.

Further studies will be needed to test at what age or under what conditions infants detect semantic violations in a stranger's speech. Nevertheless, the functional nature of the N400 component, and the similarity of the effects we found in adults and infants, suggests semantic priming by, and referential understanding of, familiar words at 9 months of age. This finding supports the view that the referential nature of speech may not have to be learned by human infants, but it may be expected and exploited by them during language acquisition.

Acknowledgments

We thank B. Kollod, M. Toth, and A. Volein for assistance, and H. Bortfeld, M. Chen, M. Friederich, A. D. Friederici, T. Gliga, A. M. Kovacs, and O. Mascaró for discussions.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by a European Research Council Advanced Investigator Grant (OSTREFCOM) to Gergely Csibra.

Supplemental Material

Additional supporting information may be found at <http://pss.sagepub.com/content/by/supplemental-data>

References

- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55, 1278–1289.
- Beauchemin, M., González-Frankenberg, B., Tremblay, J., Van-nasing, P., Martínez-Montes, E., Belin, P., . . . Lassonde, M. (2011). Mother and stranger: An electrophysiological study of voice processing in newborns. *Cerebral Cortex*, 21, 1705–1711.
- Bergelson, E., & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, USA*, 109, 3253–3258.
- Cooper, R. P., & Aslin, R. N. (1989). The language environment of the young infant: Implications for early perceptual development. *Canadian Journal of Psychology*, 43, 247–265.
- Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, 25, 141–168.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1174–1176.
- Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, 16, 1465–1477.
- Friedrich, M., & Friederici, A. D. (2005a). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *NeuroReport*, 16, 653–656.
- Friedrich, M., & Friederici, A. D. (2005b). Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, 17, 1785–1802.
- Friedrich, M., & Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and Language*, 114, 66–71.
- Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: Fast encoding–weak retention. *Journal of Cognitive Neuroscience*, 23, 3228–3240.
- Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, 20, 347–353.
- Gredeback, G., & Melinder, A. (2010). The development and neural basis of pointing comprehension. *Social Neuroscience*, 5, 441–450.
- Hoehl, S., Wiese, L., & Striano, T. (2008). Young infants' neural processing of objects is affected by eye gaze direction and emotional expression. *PLoS ONE*, 3(6), e2389. Retrieved from <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0002389>
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., . . . Roco, C. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning [Target article and commentary]. *Monographs of the Society for Research in Child Development*, 65, i–vi, 1–123.
- Kuhl, P. K. (2007). Is speech learning “gated” by the social brain? *Developmental Science*, 10, 110–120.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Nigam, A., Hoffman, J., & Simons, R. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, 4, 15–22.
- Parise, E., Friederici, A. D., & Striano, T. (2010). “Did you call me?” 5-month-old infants own name guides their attention. *PLoS One*, 5(12), e14208. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0014208>
- Parise, E., Reid, V. M., Stets, M., & Striano, T. (2008). Direct eye contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, 3, 141–150.
- Purhonen, M., Kilpeläinen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2005). Four-month-old infants process own mother's voice faster than unfamiliar voices: Electrical signs of sensitization in infant brain. *Cognitive Brain Research*, 24, 627–633.
- Robinson, C. W., Howard, E. M., & Sloutsky, V. M. (2005). The nature of early word comprehension: Symbols or associations? In B. G. Bara, L. Barsalou, & M. Bucciarrelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1883–1888). Mahwah, NJ: Erlbaum.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, 18, 668–671.
- Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, 108, 303–319.
- Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-month-old infants. *Proceedings of the National Academy of Sciences, USA*, 108, 6038–6043.
- Swingle, D. (2008). The roots of the early vocabulary in infants' learning from speech. *Current Directions in Psychological Science*, 17, 308–311.
- Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, 13, 258–263.
- Xu, F. (2007). Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences*, 11, 400–406.