Speaker gaze increases information coupling between infant and adult brains

Victoria Leongab,1, Elizabeth Byrne, Kaili Clacksona, Stanimira Georgieva, Sarah Lam, and Sam Wassd

aDepartment of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom; bDivision of Psychology, Nanyang Technological University, Singapore 637332, Republic of Singapore; cMRC Cognition & Brain Sciences Unit, University of Cambridge, Cambridge CB2 7EF, United Kingdom; and dDivision of Psychology, University of East London, London E16 2RD, United Kingdom

When infants and adults communicate, they exchange social signals of availability and communicative intention such as eye gaze. Previous research indicates that when communication is successful, close temporal dependencies arise between adult speakers’ and listeners’ neural activity. However, it is not known whether similar neural contingencies exist within adult–infant dyads. Here, we used dual-electroencephalography to assess whether direct gaze increases neural coupling between adults and infants during screen-based and live interactions. In experiment 1 (n = 17), infants viewed videos of an adult who was singing nursery rhymes with (i) direct gaze (looking forward), (ii) indirect gaze (head and eyes averted by 20°), or (iii) direct-oblique gaze (head averted but eyes orientated forward). In experiment 2 (n = 19), infants viewed the same adult in a live context, singing with direct or indirect gaze. Gaze-related changes in adult–infant neural network connectivity were measured using partial directed coherence. Across both experiments, the adult had a significant (Granger) causal influence on infants’ neural activity, which was stronger during direct and direct-oblique gaze relative to indirect gaze. During live interactions, infants also influenced the adult more during direct than indirect gaze. Further, infants vocalized more frequently during live direct gaze, and individual infants who vocalized longer also elicited stronger synchronization from the adult. These results demonstrate that direct gaze strengthens bidirectional adult–infant neural connectivity during communication. Thus, ostensive social signals could act to bring brains into mutual temporal alignment, creating a joint-networked state that is structured to facilitate information transfer during early communication and learning.

neural synchronization | dyadic interaction | mutual gaze | ostensive signals | intention

Temporally contingent social interactions between adults and infants play a vital role in supporting early learning across multiple domains of language, cognition, and socioemotional development (1, 2). Infants rely heavily on the temporal dynamics of facial cues such as eye contact and gaze direction to infer intention, meaning, and causality (3–5), which is unsurprising given that infants’ early visual experience is heavily composed of faces (6). Of all cues, direct gaze is thought to be one of the most salient ostensive signals in human communication for conveying communicative intent (4). Gaze also acts to release and reinforce infants’ own social responses such as smiling and vocalization (7, 8). From birth, infants prefer to look at pictures of faces with direct gaze over averted gaze (9). By 4 mo, direct gaze elicits a larger amplitude in the face-sensitive N170 event-related potential (ERP) relative to averted gaze (10), which suggests that gaze also enhances infants’ neural processing of face-related information.

Social Synchronization Through Gaze in Communication

According to the social brain hypothesis, human brains have fundamentally evolved for group living (11). Social connectedness is created when group members act jointly (e.g., synchronously) or contingently (e.g., turn-taking) with each other (12). Even infants show synchronization with their adult caregivers, and adult–infant temporal contingencies have long been observed in behavioral and physiological domains. For example, patterns of temporally synchronous activity between parent and child during social interaction have been noted for gaze (13), vocalizations (14), affect (15), autonomic arousal (16, 17), and hormones (18). The synchronization of gaze (through mutual gaze and gaze-following) is thought to foster social connectedness between infants and adults (19). Previous research has also suggested that infants, like adults (20), show neural synchronization (or phase-locking) of cortical oscillatory activity to temporal structures in auditory signals (21). However, adult–infant behavioral and physiological synchronization is typically observed over much slower timescales (e.g., minutes or seconds) than neural synchronization (tens or hundreds of milliseconds). Thus, it remains to be seen whether neural synchronization also develops between infants and adults during social interaction and if/how such neural coupling is related to social synchronizing signals like gaze.

Recently, researchers have begun to examine the neural mechanisms that support the contingency (temporal dependency) of one partner’s neural activity with respect to the other during social interactions (see refs. 22 and 23 for reviews). This work has revealed that during verbal communication (especially face-to-face communication, which permits mutual gaze), adult speaker–listener pairs develop synchronous patterns of activity between brain regions such as the inferior frontal gyrus, prefrontal, and parietal cortices.

Significance

During communication, social ostensive signals (like gaze) are exchanged in a temporally contingent manner. Synchronized behavior creates social connectedness within human dyads, and even infants synchronize behaviorally with adults. However, the neural mechanisms that support infant–adult synchronization are unknown. Here, we provide evidence that infants up-regulate neural synchronization with adult partners when offered direct ostensive gaze, as compared with gaze aversion. Gaze therefore brings infant–adult neural activity into mutual alignment, creating a joint-networked state that may facilitate communicative success. Further, infants’ own communicative attempts were positively associated with adults’ neural synchronization to them, indicating mutual regulation of synchronization within infant–adult dyads. Thus, interpersonal neural synchronization may provide a mechanism by which infants construct their own earliest social networks.

Author contributions: V.L. and S.W. designed research; V.L., E.B., K.C., and S.G. performed research; V.L., S.G., S.L., and S.W. analyzed data; and V.L. and S.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. U.H. is a guest editor invited by the Editorial Board.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence should be addressed. Email: vvec2@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1702493114/-/DCSupplemental.
Further, the strength of speaker-listener neural synchronization predicts communication success (26). Thus, in adults, effective communication involves the mutual alignment of brain activity, as well as the temporal alignment of behavior (e.g., conversational turn-taking and mutual gaze). However, to our knowledge, no previous research has yet investigated whether infants’ neural activity also shows contingency on an adult partner’s neural activity and whether gaze acts as a neural synchronization cue during adult–infant communication.

Gaze-Cuing of Interpersonal Neural Synchronization

Here, we assessed whether the temporal dependency (synchronization) between adult and infant neural signals differed between direct and indirect gaze. Two experiments were performed to assess gaze-cuing of interpersonal synchronization in video and live modalities, respectively. In experiment 1, infants watched a prerecorded video of an experimenter singing nursery rhymes. Patterns of temporal dependency were assessed between infants’ neural activity recorded “live” and adult’s prerecorded neural activity (Fig. 1). We manipulated the adult speaker’s gaze to be either direct to the infant, indirect (head averted by a 20° angle), or direct-oblique (head averted but eyes toward the infant). The direct-oblique condition was included to control for the side view of the face that was presented during indirect gaze and to preclude the possibility that infants were responding to superficial visual differences between stimuli. In experiment 2, which used an entirely separate cohort, infants listened live to an adult reciting nursery rhymes while she presented direct or indirect gaze to the infant. Partial directed coherence (27), a statistical measure of Granger causality (28), was used to measure gaze-related changes in interpersonal neural synchronization within the adult–infant dyadic social network.

Predictions

In terms of affect and physiological changes, research has shown that the influence of infants and parents on one another is bidirectional (29, 30). Accordingly, we predicted that (i) significant neural coupling would exist between adults and infants during social interaction, (ii) direct (and direct-oblique) gaze would both be associated with higher interpersonal neural connectivity than indirect gaze, and (iii) in experiment 1 (video), only unidirectional [adult-to-infant (A → I)] coupling would be observed, but in experiment 2 (live), bidirectional [adult-to-infant (A → I) and infant-to-adult (I → A)] coupling would be observed. Further, as temporally contingent social interactions with adults are known to facilitate infants’ own vocalizations (8, 31), we predicted that infants’ vocalization efforts would be greater during direct than indirect gaze.

Results

Gaze Modulation of Interpersonal Neural Connectivity. General Partial Directed Coherence (GPDC) measures the degree of influence that each electrode channel directly has on every other electrode channel in the network (27). Here, GPDC values were computed for real and surrogate (shuffled) data, for all nonself channel pairs (connections), for each participant dyad, for each gaze condition, and in Theta and Alpha EEG bands (Fig. 1 C and D). In the subsequent network diagrams (Figs. 2 and 3), only connections whose GPDC values significantly exceeded their surrogate threshold are plotted. A breakdown of GPDC values for each neural connection is provided in SI Appendix, section 1 (SI Appendix, Tables S1 and S2). Here we focus our analysis on mean A → I and I → A connectivity.

Experiment 1: Video. Only unidirectional A → I connectivity was observed in experiment 1; no significant I → A connectivity was detected (Fig. 2). This confirmed the validity of the GPDC measure as infants could not have affected the adult’s prerecorded neural activity. Dunnett’s tests revealed that, as predicted, A → I connectivity was (i) significantly stronger for direct > indirect gaze in both Theta and Alpha bands (P < 0.01 and P < 0.05, respectively, one-tailed) and (ii) significantly stronger for direct-oblique > indirect gaze in both Theta and Alpha bands (P < 0.0001 for both, one-tailed). However, while connectivity in the direct and direct-oblique conditions was not significantly different in the Theta band (P = 0.30) as predicted, for the Alpha band a significant difference between these conditions was observed (direct-oblique > direct, P < 0.01).

Experiment 2: Live. During the live experiment, bidirectional connectivity was observed with significant A → I as well as I → A influences (Fig. 3). Regarding A → I connectivity, consistent with experiment 1, Dunnett’s tests revealed that the adult’s influence on infants was significantly stronger for direct > indirect gaze in both Theta and Alpha bands (P < 0.05 and P < 0.0001, respectively, one-tailed).

For I → A connectivity, Dunnett’s tests indicated that infants’ influence on the adult was likewise significantly stronger for direct > indirect gaze in both Theta and Alpha bands (P < 0.01 and P < 0.05, respectively, one-tailed).

Infant Vocalization Analysis. For experiment 1 (video), there was no difference in the number of infant vocalizations (summed over all categories) between gaze conditions (means: direct = 8.2 per infant, indirect = 7.4, direct-oblique = 7.1), F(2, 32) = 0.29, P = 0.75, ηp² = 0.02. There was also no difference in the duration of vocalizations across gaze conditions (means: direct = 0.69 s per
utterance, indirect = 0.82 s, direct-oblique = 0.70 s), $F(2, 24) = 0.37$, $P = 0.70$, $\eta^2_p = 0.03$. However, for experiment 2 (live), we observed a significantly higher number of vocalizations during direct gaze (mean 6.3 per infant) than indirect gaze (mean 5.0 per infant), $t(18) = 2.41$, $P < 0.05$, but no difference in the duration of vocalizations (mean: direct = 0.80 s per utterance, indirect = 0.85 s), $t(15) = -0.79$, $P = 0.44$.

Further, during experiment 2 (live), individual differences in infants’ vocalization durations were significantly associated with their $I \rightarrow A$ GPDC values [$r = 0.67$, $P < 0.05$, Benjamini–Hochberg false discovery rate (FDR) corrected] (32) (see Fig. 4). However, this correlation only emerged during direct gaze and was absent for indirect gaze ($r = 0.07$, $P = 0.78$). Therefore, infants who produced longer vocalizations also influenced the adult more strongly—but only when she offered direct gaze. *SI Appendix, section 2* provides further analyses of infants’ vocalizations.

**Discussion**

Temporally contingent social interactions between adults and infants scaffold early learning and development. Here, we tested the hypothesis that gaze acts as an interpersonal neural synchronization cue between dyadic (adult–infant) partners. Two experiments were performed to assess the effect of direct speaker gaze on interpersonal synchronization using video (experiment 1) and live (experiment 2) modalities. Across both experiments, significant neural coupling between infants and adults was observed during social interaction, relative to rigorous control analyses that accounted for nonspecific neural coupling. Adult–infant neural coupling was observed consistently across video and live presentation formats, using two separate cohorts of infants. Further, during unidirectional interactions in experiment 1 (i.e., infants watching a prerecorded adult speaker), the adult had a significant influence on infants’ neural activity, but (as expected) infants had no influence on the adult’s neural activity. Conversely, during live (bidirectional) social interactions (experiment 2), there were significant and bidirectional patterns of influence between adult and infant.

Across both experiments, we consistently observed that direct gaze produced higher interpersonal neural synchronization than indirect gaze in both Theta and Alpha frequency bands. Further, in experiment 2 (live), the synchronizing effect of gaze was observed bidirectionally: During direct gaze, the adult had a stronger

---

**Fig. 2.** (Left) Network depiction of experiment 1 Theta (3–6 Hz, Top) and Alpha (6–9 Hz, Bottom) connectivity, plotting GPDC values for direct (Left), indirect (Middle), and direct-oblique gaze (Right) conditions. Nodes represent C3 (L) and C4 (R) electrodes for adult (A) and infant (I). Arrows indicate the direction and strength of connectivity (higher GPDC value, thicker arrow). Connections that do not significantly exceed the surrogate threshold are excluded. (Right) Grand mean GPDC values averaged across all adult-to-infant ($A \rightarrow I$) connections for Theta (Top) and Alpha (Bottom) in direct (D), indirect (I), and direct-oblique (D-O) gaze conditions. Error bars show the SEM. *$P < 0.05$.

**Fig. 3.** (Left) Network depiction of experiment 2 Theta (3–6 Hz, Top) and Alpha (6–9 Hz, Bottom) connectivity, plotting GPDC values for direct (Left) and indirect (Right) gaze conditions. Nodes represent C3 (L) and C4 (R) electrodes for adult (A) and infant (I). Arrows indicate the direction and strength of connectivity (higher GPDC value, thicker arrow). Connections that do not significantly exceed the surrogate threshold are excluded. (Right) Grand mean GPDC values averaged across all adult-to-infant ($A \rightarrow I$, Left) and infant-to-adult ($I \rightarrow A$, Right) connections for Theta (Top) and Alpha (Bottom) in direct (D) and indirect (I) gaze conditions. Error bars show the SEM. *$P < 0.05$.
influence on the infant, and the infant also had a stronger influence on the adult. This gaze-related increase in synchronization was not due to power differences in the EEG spectra, nor was it a metaphenomenon of changes in basic sensory processing of the speech signal (which remained unchanged across gaze conditions). In experiment 1, we further showed that the gaze effect was not driven by superficial visual differences in the stimuli, since direct-oblique stimuli were visually similar to indirect stimuli but produced greater synchronization. It was also not the case that infants were more inattentive during indirect gaze, as infants looked just as long at indirect and direct-oblique stimuli in experiment 1 and at indirect and direct stimuli in experiment 2. Therefore, the increased interpersonal neural synchronization produced by direct gaze appears to reflect stronger mutual oscillatory phase alignment between adult and infant.

A Mechanism for Interpersonal Neural Synchronization. One mechanism that might mediate this effect is mutual phase resetting in response to salient social signals. The phase of cortical oscillations (the neural feature used in GPDC computations) reflects the excitability of underlying neuronal populations to incoming sensory stimulation (33). Sensory information arriving during high-receptivity periods is more likely to be encoded than information arriving during low-receptivity periods. Consequently, neuronal oscillations have been proposed to be a mechanism for temporal sampling of the environment (20). Specifically, salient events are thought to reset the phase of ongoing neuronal oscillations to match the temporal structure of these events and optimize their encoding (33). Consequently, interpersonal neural synchronization could increase within a dyad during the course of social interaction because each partner is continuously producing salient social signals (such as gaze, gestures, or vocalizations) that act as synchronization triggers to reset the phase of his or her partner’s ongoing oscillations. As a result, infants’ most receptive periods become well-aligned to adults’ speech temporal patterns (e.g., prosodic stress and syllable patterns) (34), optimizing communicative efficiency. This mechanism could also allow slow-varying behavioral synchronization signals (like gaze) to hierarchically control fast-varying neural synchronization between partners (33).

Direct Gaze Supports Communication Through Synchronization. Our findings suggest that direct gaze from the adult may reset the phase of infants’ oscillations to align with that of the adults’, thereby increasing mutual synchronization (i.e., stronger A → I connectivity). One aspect of our results was, however, unpredicted. In experiment 1, we had predicted an equal effect for direct and direct-oblique gaze, yet we found that Alpha neural synchrony was higher for direct-oblique than direct gaze. One possible explanation for this is that infants are less frequently exposed to direct eye contact when the speaker’s head is averted, which could therefore present greater novelty. However, infants did not look for longer at the speaker during the direct-oblique condition relative to the direct gaze condition, which is inconsistent with this explanation. A second potential explanation is that the direct-oblique condition provided a stronger intentional ostensive cue because the speaker’s gaze was intentionally forward while her face and body were averted. This predicts that social cues that are perceived as the most intentional will produce the strongest increases in interpersonal connectivity. Further, since phase resetting optimizes information transfer between dyadic partners (33), stronger intentional signals could produce more effective phase resetting, which would increase the potential for mutual communication and learning within the dyad. Future work should investigate this hypothesis in more detail.

As observed in previous studies (8), we also found that infants vocalized more frequently toward the adult during live direct gaze (when interpersonal synchronization was higher) than indirect gaze. Further, individual infants who vocalized for longer under live direct gaze also had stronger neural connectivity with their adult partner (i.e., stronger I → A connectivity), even during segments when no vocalizations were occurring. One possible reason for this could be that infants’ vocalizations (which were communicative signals to the adult and could potentially trigger phase resetting) acted as a social feedback mechanism to positively reinforce and sustain dyadic synchronicity (8, 31, 35).

Our present findings may offer the potential for integrating three separate strands of research into early learning: first, research that has pointed to the importance of eye gaze as an ostensive cue during learning (3); second, research into the importance of contingent social feedback, which is thought to energize early learning (31); and third, research into the role of bidirectional parent-child synchrony in structuring and scaffolding learning experiences (36). Phase resetting due to synchronization triggers that are more prevalent during mutual than indirect gaze may, potentially, offer the means for providing contingent feedback (in which the child responds to the parent, and vice versa) within the framework of the periodic oscillatory activity that structures and scaffolds early learning (36). Over longer time frames, infants’ neural synchrony with adults may also offer an implicit mechanism for learning adult-like response patterns via entrainment.

Limitations and Conclusion. Our results converge with previous dual functional near-infrared spectroscopy (fNIRS) studies (24, 37) where greater frontal neural synchronization between adults was observed during eye contact. However, one limitation of the current work is that due to the adult’s speech production artifacts, only two EEG channels, C3 and C4, could be analyzed from any individual. Thus, unlike the fNIRS studies, we were unable to make inferences about the potential neural sources of these effects. A second limitation of the current work is that, by excluding a large proportion of infants’ “active” data by technical necessity, the results could present a selective view of the neural dynamics underlying adult–infant engagement. Nonetheless, the current data are still valuable in providing insight into adult–infant neural coupling during social communication. The current study demonstrates that adults and infants show significant mutual neural coupling during social interactions and that direct gaze strengthens adult–infant neural connectivity in both directions during communication. Further, live gaze appeared to stimulate infants’ own communicative efforts, which could help to reinforce dyadic synchronization. Thus, gaze and speech act as cues for interpersonal synchronization. The contingent exchange of these social signals acts to bring adults’ and infants’ brains into temporal alignment, creating a joint-networked state that is structured to optimize information transfer during communication and learning.
Methods

Participants. Experiments 1 and 2 involved separate infant cohorts—experiment 1: 19 infants (13 male, 6 female), median age 8.2 m (SE, 0.26 m), and experiment 2: 29 infants (15 male, 14 female), median age 8.3 m (SE, 0.44 m). Infants’ mothers were native English speakers, and all infants had no neurological problems as assessed by maternal report. The same female adult experimenter participated in both experiments with all infants. The study received ethical approval from the Cambridge Psychology Research Ethics Committee. Parents provided written informed consent on behalf of their infants.

Materials. For both experiments, seven familiar nursery rhymes were used as sung stimuli (SI Appendix, section 3). Sung nursery rhymes were used because these are integral to play and caretaking routines with infants, such as during feeding and putting to sleep (38). Infants are equally or more behaviorally responsive to sung compared with spoken language (39); thus, sung speech is likely to evoke a robust neural response from infants. In experiment 1, pre-recorded video stimuli were used with mean pitch, pitch variability, duration, and loudness matched across gaze conditions (SI Appendix, Table S5). For experiment 2 (live), the experimenter was recorded during each session to ensure acoustic consistency across gaze conditions (SI Appendix, Table S6). Paired t tests indicated no significant differences between conditions for all acoustic parameters. The experimenter was instructed to maintain a neutral facial expression across all gaze conditions, varying only her gaze direction.

Protocol.

Experiment 1. Infants sat upright in a high chair 70 cm from a display monitor (90 cm width × 60 cm height), showing a life-sized image of a female experimenter’s head against a black background. Each nursery rhyme was presented in three gaze conditions (Fig. 1): direct, indirect (head averted by 20°), and direct-oblique (head averted by 20°, but direct gaze). The direct-oblique condition was included to control for the side view of the face that was presented during indirect gaze. During stimulus recording, the experimenter gazefixed on a life-sized picture of an infant to standardize her visual input across conditions. Each nursery rhyme was presented six times (twice per gaze condition, order counterbalanced).

Experiment 2. Infants sat upright in a high chair facing the female experimenter at a distance of 70 cm. Each nursery rhyme was presented in two gaze conditions. In the direct condition, the experimenter looked directly at the infant while singing; in the indirect condition, she fixated at a target 20° to the left or right side of the infant (see Fig. 1 and SI Appendix, section 4 for the experimenter's view). Each nursery rhyme was presented four times (twice direct, twice indirect, order counterbalanced).

EEG Acquisition. In experiment 1, EEG was recorded separately from infants (during testing) and from the female adult experimenter (during stimulus recording) from 32 electrodes according to the international 10–20 placement system. In experiment 2, EEG was recorded simultaneously from the infant and the adult experimenter from two central electrodes (C3 and C4), referenced to the vertex (Cz). Further details of EEG acquisition are given in SI Appendix, section 5.

EEG Artifact Rejection and Preprocessing. To ensure that the analyzed EEG data reflected only attentive and movement-free neural activity, a two-stage artifact rejection procedure was performed. First, manual artifact rejection was performed to further exclude segments where the EEG amplitude exceeded +100 μV. Full descriptions of the artifact rejection procedures and inclusion rules following artifact rejection are given in SI Appendix, section 6. Data were then downsampled to 200 Hz, low-pass filtered <45 Hz to suppress electrical line noise, and segmented into 1-s epochs for connectivity analysis.

EEG Analyses: Speech Artifacts, Power Spectrum, and GPDC Network Connectivity. Speech production artifacts were present in the EEG signal of the adult speaker. To assess the topography and spectral profile of these artifacts, we compared the adult’s EEG during speech production relative to resting state (SI Appendix, section 7). Despite rigorous analyses, we were able to identify no evidence of EEG signal distortion by speech artifacts in the central region (e.g., C3/C4) in Theta and Alpha bands, although evidence of artifacts at other frequency bands and for more peripheral electrode positions was clearly present. Therefore, to avoid spurious results arising from speech artifacts, the connectivity analysis used only Theta and Alpha bands for C3 and C4 electrodes for both adult and infant. To confirm the representativeness of this region of analysis for the infant, we assessed infants’ whole-head (32-channel) connectivity to adults’ C3 and C4 electrodes (Fig. 5 and SI Appendix, section 12). Across gaze conditions, the strongest connectivity between infant and adult was topographically observed over infants’ central and posterior regions (including C3 and C4) for both Theta and Alpha bands. Therefore, C3 and C4 were indeed representative regions of analysis for the infant.

A detailed description of EEG analysis methods is given in SI Appendix, sections 8 and 9. Briefly, first the EEG power spectra of infant and adult signals were assessed for each experimental condition to confirm that the gaze manipulation did not generate any detectable power changes that might systematically bias the connectivity analysis. Second, to assess network connectivity in each gaze condition, GPDC—a directional causal measure of direct information flow between channels in a network—was computed (27). GPDC measures the degree of influence that channel i directly has on channel j with respect to the total influence of i on all channels in the network. Here, each electrode [infant left (IL), infant right (IR), adult left (AL), adult right (AR)] was one channel (Fig. 1C).

Control Analyses. The first control analysis established a threshold for non-specific connectivity between brains that was unrelated to the experimental task (SI Appendix, section 10). A surrogate dataset was generated for each participant pair where the fine-grained temporal correspondence between adult and infant neural signals was disrupted by randomly pairing adult and infant epochs from different timepoints within the same experimental session (i.e., shuffling). An identical connectivity analysis was then performed on this surrogate dataset. For each participant pair, neural connection, and frequency band, a threshold value was computed by taking the average surrogate value across all gaze conditions. Paired t tests [Benjamini-Hochberg (FDR)-corrected at P < 0.05 (32), one-tailed] were then used to assess whether the real data significantly exceeded their respective threshold values.

The second control analysis examined basic sensory processing of the speech stimulus, which could indirectly affect adult–infant neural coupling. Entrainment (oscillatory phase-locking) between the EEG signal and the speech amplitude envelope was measured in each gaze condition. As described in SI Appendix, section 11, no significant differences in neural entrainment to the speech signal were found between gaze conditions in either experiment.

Statistical Analysis of Gaze Effects on Interpersonal GPDC Connectivity. We hypothesized that interpersonal neural connectivity would be higher during direct (and direct-oblique) gaze than indirect gaze (i.e., direct > indirect–indirect). We also wished to assess whether the adult’s influence on the infant (i.e., A → I GPDC) and the infant’s influence on the adult (i.e., I → A GPDC) would show the same pattern of gaze modulation. As previous work with infants has not found hemisphere differences for gaze effects (30), interhemispheric connectivity patterns were not explored further. Accordingly, the four interhemispheric connections (L→R, R→L) were collapsed into one average each for A → I and I → A directional influences. These two directional indices were computed for each gaze condition, for Theta and Alpha bands. For experiment 1, only A → I connections were analyzed, as all I → A connections were not significantly above threshold (this was expected, as adults’ EEG was prerecorded).

The effects of gaze on A → I and I → A connectivity were assessed using two statistical approaches. First, to assess overall patterns and interactions, repeated-measures ANOVAs were performed, taking frequency and gaze condition as within-subjects factors. Second, to assess specific contrasts
between pairs of gaze conditions at each frequency, Dunnett’s multiple range t tests (40) were conducted, which independently control for the familywise error rate. For Theta and Alpha bands, the following pairwise procedures were performed for experiment 1: (i) direct → indirect, (ii) direct-oblique → indirect, and (iii) direct → direct-oblique. For experiment 2, only the direct → indirect test was performed. Dunnett’s test results are reported in the main text, and ANOVA results are provided in SI Appendix, section 13. Separate analyses were also performed to examine infants’ looking times (SI Appendix, section 14) and the effects of infant age on neural connectivity (SI Appendix, section 15). Finally, a permutation analysis was performed (SI Appendix, section 16) to assess the internal reliability of the gaze findings, both within and across experiments. All statistical tests were two-tailed unless there was a priori directional hypotheses (i.e., Dunnett’s test for direct/direct-oblique → indirect; data > surrogate threshold), for which one-tailed tests were used.

Infant Vocalizations. Infants’ vocalizations were coded from session videos according to Oller’s (41) infraphonological acoustic classification system (SI Appendix, section 2). Each infant’s (i) number and (ii) duration of vocalizations were computed during each gaze condition. To explore the relationship between neural coupling and infants’ communicative attempts, vocalization indices were correlated with A → I and I → A GPDC values for both experiments. Of note, the connectivity analyses only included segments of EEG data when no vocalizations were occurring.

ACKNOWLEDGMENTS. This research was funded by a UK Economic and Social Research Council (ESRC) Transforming Social Sciences Grant ES/N006461/1 (to V.L. and S.W.), a Lucy Cavendish College Junior Research Fellowship (to V.L.), Nanyang Technological University start-up Grant M4081585.S50 (to V.L.), and a British Academy Post-Doctoral Fellowship and ESRC Future Leaders Fellowship ES/N017560/1 (to S.W.).

**SUPPORTING INFORMATION (SI) APPENDIX**

1  GPDC values by EEG frequency band

1.1  Experiment 1: Video

<table>
<thead>
<tr>
<th></th>
<th>Across Individuals</th>
<th>Within Individuals</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adult -&gt; Infant</td>
<td>Infant -&gt; Adult</td>
<td>Infant -&gt; Infant</td>
<td>Adult -&gt; Adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>R</td>
<td>L</td>
<td>R</td>
<td>L</td>
<td>R</td>
<td>L</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>R</td>
<td>L</td>
<td>R</td>
<td>L</td>
<td>R</td>
<td>L</td>
<td>R</td>
</tr>
<tr>
<td>Direct</td>
<td>0.076</td>
<td>0.082</td>
<td>0.087</td>
<td>0.090</td>
<td>0.068</td>
<td>0.067</td>
<td>0.071</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>0.016</td>
<td>0.014</td>
<td>0.008</td>
<td>0.009</td>
<td>0.008</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>Indirect</td>
<td>0.076</td>
<td>0.082</td>
<td>0.081</td>
<td>0.084</td>
<td>0.069</td>
<td>0.075</td>
<td>0.067</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>0.015</td>
<td>0.013</td>
<td>0.007</td>
<td>0.007</td>
<td>0.011</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Direct-Oblique</td>
<td>0.079</td>
<td>0.084</td>
<td>0.090</td>
<td>0.086</td>
<td>0.067</td>
<td>0.068</td>
<td>0.064</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>0.013</td>
<td>0.012</td>
<td>0.012</td>
<td>0.006</td>
<td>0.007</td>
<td>0.007</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>Surrogate</td>
<td>0.065</td>
<td>0.068</td>
<td>0.074</td>
<td>0.077</td>
<td>0.068</td>
<td>0.069</td>
<td>0.068</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>0.004</td>
<td>0.005</td>
<td>0.006</td>
<td>0.008</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.011</td>
<td>0.010</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>0.078</td>
<td>0.088</td>
<td>0.091</td>
<td>0.094</td>
<td>0.043</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>0.012</td>
<td>0.015</td>
<td>0.016</td>
<td>0.012</td>
<td>0.004</td>
<td>0.005</td>
<td>0.007</td>
<td>0.005</td>
</tr>
<tr>
<td>Indirect</td>
<td>0.080</td>
<td>0.087</td>
<td>0.085</td>
<td>0.091</td>
<td>0.043</td>
<td>0.048</td>
<td>0.042</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>0.016</td>
<td>0.017</td>
<td>0.010</td>
<td>0.005</td>
<td>0.006</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>Direct-Oblique</td>
<td>0.084</td>
<td>0.088</td>
<td>0.096</td>
<td>0.095</td>
<td>0.042</td>
<td>0.045</td>
<td>0.041</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>0.013</td>
<td>0.011</td>
<td>0.017</td>
<td>0.004</td>
<td>0.004</td>
<td>0.005</td>
<td>0.006</td>
</tr>
<tr>
<td>Surrogate</td>
<td>0.071</td>
<td>0.075</td>
<td>0.081</td>
<td>0.085</td>
<td>0.042</td>
<td>0.044</td>
<td>0.042</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>0.006</td>
<td>0.007</td>
<td>0.008</td>
<td>0.010</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Shaded connections were not significantly above surrogate threshold (BH FDR-corrected at p<.05). See Section 10 for surrogate analysis.

Table S1 – Experiment 1 GPDC values by EEG frequency band (mean in bold, SD in italics)
1.2 Experiment 2: Live

<table>
<thead>
<tr>
<th></th>
<th>Across Individuals</th>
<th>Within Individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adult -&gt; Infant</td>
<td>Infant -&gt; Adult</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>R</td>
</tr>
<tr>
<td>Theta (3-6 Hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>0.072</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>0.025</td>
<td>0.026</td>
</tr>
<tr>
<td>Indirect</td>
<td>0.072</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>0.022</td>
<td>0.017</td>
</tr>
<tr>
<td>Surrogate</td>
<td>0.058</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>0.012</td>
<td>0.010</td>
</tr>
<tr>
<td>Alpha (6-9 Hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>0.075</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>0.022</td>
<td>0.029</td>
</tr>
<tr>
<td>Indirect</td>
<td>0.072</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>0.024</td>
<td>0.027</td>
</tr>
<tr>
<td>Surrogate</td>
<td>0.061</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>0.019</td>
<td>0.018</td>
</tr>
</tbody>
</table>

*Shaded connections* were not significantly above surrogate threshold (BH FDR-corrected at p<.05). See Section 10 for surrogate analysis.

*Table S2 – Experiment 2 GPDC values by EEG frequency band (mean in bold, SD in italics)*
Infant vocalisation analysis and correlations with neural connectivity

Infants’ vocalisations were manually coded from videos recorded during the experimental session according to Oller’s [1] infraphonological acoustic classification system. This coding scheme incorporates acoustic features (such as fundamental frequency and formant transitions) with qualitative descriptors (e.g. phonetic categories) to distinguish between four categories of vocalisations: quasi-resonant vowel nuclei, fully-resonant vowel nuclei, marginal syllables and canonical syllables. The infants in both studies (median age of 8/8.5 months) were expected to produce all four categories of vocalisations. The total mean number and (utterance) duration of infants’ vocalisations in each experiment and gaze condition are shown in Table S3.

Table S3. Mean number and duration of infants’ vocalisations in each experiment and gaze condition. Means are shown in bold, standard errors are shown in italics.

<table>
<thead>
<tr>
<th>Gaze Condition</th>
<th>Mean number per infant</th>
<th>Mean duration per infant (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expt 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>8.22 (2.43)</td>
<td>0.69 (.10)</td>
</tr>
<tr>
<td>Indirect</td>
<td>7.44 (1.80)</td>
<td>0.82 (.15)</td>
</tr>
<tr>
<td>Direct-Oblique</td>
<td>7.11 (1.69)</td>
<td>0.70 (.07)</td>
</tr>
<tr>
<td>Expt 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>6.32 (1.11)</td>
<td>0.80 (.10)</td>
</tr>
<tr>
<td>Indirect</td>
<td>5.00 (1.20)</td>
<td>0.85 (.08)</td>
</tr>
</tbody>
</table>

As discussed in the main manuscript, the total number of utterances increased significantly for Direct relative to Indirect gaze in Expt 2 (live), but not for Expt 1 (video) where interactions were uni-directional (the infant could not influence the adult). To assess whether the social interaction context significantly moderated the effect of gaze on infants’ vocalisations, we computed the mean difference between infants’ number of vocalisations under Direct versus Indirect gaze, predicting that this difference would be larger for Expt 2 (live) than for Expt 1 (video). A normalized (i.e. Direct minus Indirect) index was used rather than raw values to account for differences in the baseline number of vocalisations between experiments. As there were two Direct gaze conditions in Experiment 1 (Direct and Direct-Oblique), we used the average number of vocalisations across these two conditions. An ANOVA was then conducted on the normalized vocalisation index, taking Experiment as the between-subjects factor, and controlling for infants’ looking time and age. This analysis revealed a trend toward a significant difference between experiments (F(1,23)=2.17, p=.077, one-tailed), with a larger benefit of Direct gaze for vocalisations in the live interaction context (Expt 1) than for video (Expt 2), as predicted.

This live benefit is reminiscent of Kuhl et al.’s study [2] where infants showed phoneme learning from live speakers but not from video DVDs of the same speakers. Similarly, Goldstein & Schwade [3] found that only infants who received live contingent feedback from their mothers showed re-structuring of their babbling patterns. Consistent with these studies, here, infants produced more vocalisations only when Direct gaze was offered in a live contingent context.

2.1 Vocalisations by category

A breakdown of the mean number of infant vocalisations by category of complexity is provided in Figure S1. For each experiment, a Repeated Measures ANOVA was conducted to assess whether the complexity of vocalisations differed across gaze conditions, taking Complexity (4 levels) and Gaze (3 or 2 levels) as within-subjects factors. For Experiment 1, there was no main effect of Gaze (F(2,32) = .29, p=.75, η²p = .02) and no interaction between Gaze and Complexity (4 levels) and Gaze (3 or 2 levels) as within-subjects factors.
vocalisation Complexity (F(6,96) = .39, p = .88, η^2_p = .02). However, there was a significant main effect of Complexity (F(3,48) = 8.94, p<.001, η^2_p = .36. Significantly more quasi-resonant and fully-resonant nuclei were produced than marginal and canonical syllables, but there was no difference within these sub-categories.

For Experiment 2, there was a significant main effect of Gaze (F(1,18) = 5.80, p<.05, η^2_p = .24) but no interaction between Gaze and vocalisation Complexity (F(3,54) = 1.67, p = .18, η^2_p = .09). However, there was again a significant main effect of Complexity (F(3,54) = 8.20, p<.001, η^2_p = .31. As for Expt 1, significantly more quasi-resonant and fully-resonant nuclei were produced than marginal and canonical syllables, but there was again no difference within these sub-categories. Therefore, these results indicate that the adult speaker’s gaze did not change the complexity of infants’ utterances.

Figure S1. Mean number of vocalisations in each category for Expt 1 (left) and Expt 2 (right). QR Nuc = Quasi-Resonant nucleus, FR Nuc = Fully-Resonant nucleus, M Syll = Marginal syllable; C syll = Canonical syllable. Error bars indicate the standard error of the mean.

2.2 Correlations with neural coupling

Table S4 shows the correlation between adult-to-infant and infant-to-adult GPDC values (averaged across Theta and Alpha bands) and vocalisation duration, for each Experiment. As there was no significant infant-to-adult sending in Experiment 1, these correlations were not computed. Infants’ vocalisation duration was only correlated to their own neural sending patterns (i.e. infant-to-adult), and not the adults’ sending patterns. Thus, infants were not vocalising for longer in response to the adult, rather, their longer vocalisations were having a stronger synchronizing effect on the adult. Since the analysed EEG segments excluded periods of infant vocalisations (motion), speech artifacts could not account for this effect. Further, the neural-vocalisation relationship emerged only under Direct gaze from the adult, and was absent during Indirect gaze, consistent with the availability of the adult providing a stimulus for infants to vocalise with stronger communicative intent toward her. There were no significant correlations between neural connectivity and number of vocalisations for any gaze condition. This suggests that not every vocalization was equally effective in increasing neural connectivity with the adult. Rather, sustained vocalisations of a longer duration were more effective in influencing the adult.
<table>
<thead>
<tr>
<th>Gaze Condition</th>
<th>Adult-to-Infant ( r ) (raw ( p )-val)</th>
<th>Infant-to-Adult ( r ) (raw ( p )-val)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expt 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>-0.46 (.11)</td>
<td>N.A.</td>
</tr>
<tr>
<td>Indirect</td>
<td>0.03 (.93)</td>
<td>N.A.</td>
</tr>
<tr>
<td>Direct-Oblique</td>
<td>-0.25 (.41)</td>
<td>N.A.</td>
</tr>
<tr>
<td><strong>Expt 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>-0.09 (.47)</td>
<td>*0.07 (.07)</td>
</tr>
<tr>
<td>Indirect</td>
<td>-0.12 (.28)</td>
<td>0.07 (.78)</td>
</tr>
</tbody>
</table>

Table S4. Pearson correlation \( r \)-values and raw (uncorrected) \( p \)-values for adult-to-infant and infant-to adult connectivity (GPDC averaged over Theta and Alpha) and infant vocalisation duration in each gaze condition. \(*p<.05\) (Benjamini-Hochberg FDR corrected)
### 3 Nursery rhyme stimuli

#### 3.1 Experiment 1: Video

<table>
<thead>
<tr>
<th></th>
<th>Direct gaze</th>
<th>Indirect gaze</th>
<th>Direct-Oblique gaze</th>
<th>Direct gaze</th>
<th>Indirect gaze</th>
<th>Direct-Oblique gaze</th>
<th>Direct gaze</th>
<th>Indirect gaze</th>
<th>Direct-Oblique gaze</th>
<th>Direct gaze</th>
<th>Indirect gaze</th>
<th>Direct-Oblique gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>If You’re Happy</td>
<td>14.05</td>
<td>13.92</td>
<td>13.78</td>
<td>261.6</td>
<td>261.9</td>
<td>261.1</td>
<td>34.9</td>
<td>37.6</td>
<td>36.3</td>
<td>70.0</td>
<td>70.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Hickory Dickory Dock</td>
<td>6.84</td>
<td>6.76</td>
<td>6.93</td>
<td>224.2</td>
<td>224.4</td>
<td>224.4</td>
<td>39.3</td>
<td>33.0</td>
<td>33.7</td>
<td>70.0</td>
<td>70.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Humpty Dumpty</td>
<td>7.58</td>
<td>7.75</td>
<td>7.61</td>
<td>211.4</td>
<td>211.3</td>
<td>211.8</td>
<td>25.0</td>
<td>24.2</td>
<td>23.3</td>
<td>70.0</td>
<td>70.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Old MacDonald</td>
<td>19.29</td>
<td>19.11</td>
<td>19.33</td>
<td>246.5</td>
<td>246.2</td>
<td>246.8</td>
<td>36.7</td>
<td>35.4</td>
<td>36.4</td>
<td>70.0</td>
<td>70.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Where is Thumbkin</td>
<td>13.32</td>
<td>13.48</td>
<td>13.14</td>
<td>257.5</td>
<td>258.4</td>
<td>257.4</td>
<td>53.7</td>
<td>53.3</td>
<td>49.9</td>
<td>70.0</td>
<td>70.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Twinkle Twinkle</td>
<td>21.04</td>
<td>20.87</td>
<td>21.11</td>
<td>245.5</td>
<td>245.9</td>
<td>245.6</td>
<td>37.2</td>
<td>35.5</td>
<td>37.5</td>
<td>70.0</td>
<td>70.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Wheels on the Bus</td>
<td>10.54</td>
<td>10.6</td>
<td>10.8</td>
<td>243.6</td>
<td>243.2</td>
<td>243.3</td>
<td>42.0</td>
<td>43.8</td>
<td>41.2</td>
<td>70.0</td>
<td>70.0</td>
<td>70.0</td>
</tr>
<tr>
<td><strong>Average (SD)</strong></td>
<td><strong>13.24</strong></td>
<td><strong>13.21</strong></td>
<td><strong>13.24</strong></td>
<td><strong>241.47</strong></td>
<td><strong>241.61</strong></td>
<td><strong>241.49</strong></td>
<td><strong>38.41</strong></td>
<td><strong>37.55</strong></td>
<td><strong>36.90</strong></td>
<td><strong>70.0</strong></td>
<td><strong>70.0</strong></td>
<td><strong>70.0</strong></td>
</tr>
</tbody>
</table>

Table S5. Acoustic parameters of pre-recorded video nursery rhyme stimuli used in Experiment 1. Note that the loudness of all stimuli was equalized to 70dB. Pitch variability was computed as the standard deviation of pitch values across all timepoints in each stimulus.

Repeated Measures ANOVA for **Duration**: F(2,12) = .10, p = .91 (n.s.)
Repeated Measures ANOVA for **Mean Pitch**: F(2,12) = .38, p = .69 (n.s.)
Repeated Measures ANOVA for **Pitch Variability**: F(2,12) = 1.31, p = .31 (n.s.)
### Table S6. Acoustic parameters of live nursery rhyme stimuli used in Experiment 2.

Nursery rhymes were videoed live and the timings analysed post hoc. For each nursery rhyme, the average duration, mean pitch, pitch variability and loudness during Direct and Indirect conditions is given, and the SD is shown in brackets. Paired sample t-tests were calculated to assess whether the average duration, mean pitch, pitch variability or loudness of any of the nursery rhymes was significantly different across gaze conditions. No significant differences for any acoustic parameter or nursery rhyme were identified at the Benjamini-Hochberg (BH) FDR-corrected threshold of \( p<.05 \).
4 Experiment 2: Experimenter’s gaze perspective

During Direct gaze, the experimenter fixated on the infant and during Indirect gaze, she fixated on a red visual target placed 20° to the right or left side of the infant (in Figure S2, the target is placed on the right). Note that even during Indirect gaze, the infant was still clearly visible in her visual field. The infants’ image was more peripheral and also very slightly larger during Indirect gaze because by rotating her head, she also brought her contralateral eye slightly closer to the infant.

Of note, in Experiment 2, we observed some, but much-reduced, infant-to-adult coupling in the Indirect condition. This was not unexpected, since the infant was facing the adult directly in both conditions, and, for the adult, the infant was positioned at 20° eccentricity from the fixation point, and so still clearly visible when her gaze was averted.

5 EEG acquisition

In both experiments, EEG signals were acquired using wireless amplifiers to reduce distraction for the infant during testing. In Experiment 1, EEG signals were obtained using a 32-channel wireless Biopac Mobita Acquisition System and 32-channel Easycap caps with electrodes placed at Fp1, Fp2, AFz, Fz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, T7, T8, FT9, FT10, Cz, C3, C4, CP1, CP2, CP5, CP6, Pz, P3, P4, P7, P8, TP9, TP10, POz and Oz according to the International 10–20 placement system. EEG was recorded at 500 Hz with no online filtering using AcqKnowledge software (Biopac Systems Inc). In Experiment 2, EEG signals were recorded from C3 and C4 locations at 1000 Hz using a 2-channel Biopac MP150 Acquisition System with filters set at 0.1 Hz highpass and 100 Hz lowpass using AcqKnowledge software (Biopac Systems Inc). Both adult and infants’ data was recorded concurrently in a single acquisition session on the same computer, ensuring accurate time synchronisation of the two data streams.

Prior to electrode or cap attachment, electrode sites were marked and wiped with alcohol. Conductive electrode gel was used to affix the electrodes/cap to the scalp. In Experiment 2, EEG was recorded from central sites to reduce potential confounding influences of muscle artefacts and blinking while still capturing a robust neural response (see analysis of speech production artifacts in Section 7). Across both experiments, a vertex reference location was used because it produces comparable results to other reference sites [4], and is the least invasive for young infants.
6 EEG artifact rejection

To ensure that the EEG data used for analysis reflected only attentive and movement-free behavior we performed a two-stage artifact rejection procedure. First, each experimenter-infant dyad was video-taped and the videos were reviewed frame-by-frame (30 fps) to identify the onset and offset times of movement artifacts, including blinks, head and limb motion, and chewing. Only periods when infants were still and looking directly at the experimenter were accepted. Next, manual artifact rejection was performed on this still, attentive data to further exclude segments where the amplitude of infants' or adults' EEG exceeded +100 μV.

6.1 Experiment 1 : Video

Following the two-stage artifact detection and rejection process, 17/19 infants (12M/5F), gave sufficient data for inclusion in the final analysis. The median (st. err.) age of the retained infants was 8.0 months (0.28 months). On average, the retained infants contributed 94.82 seconds (range = 25s to 171s, SD = 43.02s) of attentive and artifact-free data in the Direct gaze condition, 86.29 seconds (range = 31 to 169s, SD = 41.67s) in the Indirect gaze condition and 89.00 seconds (range = 35 to 170s, SD = 43.08s) in the Direct-Oblique gaze condition. Adult data was only analysed for those segments in which the infant data was retained. A larger quantity of clean and attentive data was obtained in the Direct gaze condition than in the Indirect gaze condition (t(16) = 2.83, p< .05), but data quantity did not differ between Direct-Oblique and Indirect gaze conditions (t(16) = 1.16, p=.26) or between Direct and Direct-Oblique gaze conditions (t(16) = 1.54, p=.14). However, additional analyses performed to assess the effect of these data quantity differences (e.g. sub-sampling an equal number of epochs across gaze conditions) confirmed that the main effects of gaze were not affected by data quantity.

6.2 Experiment 2 : Live

Following artifact rejection, 19/29 infants (10M/9F), gave sufficient data for inclusion in the final analyses. The median (st.err.) age of retained infants was 8.52 (0.57) months. On average, the retained infants contributed 45.52 seconds (range = 8s to 107s, SD = 28.18s) of attentive and artifact-free data in the Direct gaze condition, and 43.92 seconds (range = 11 to 123s, SD = 30.07s) in the Indirect gaze condition. A paired t-test confirmed that there was no significant difference in the amount of clean data obtained between Direct and Indirect gaze conditions (t(18) = 0.44, p = .66) therefore all the clean data was used for analysis. Adult data was only analysed for those segments in which the infant data were retained.

7 Adult speech artifact analysis (speaking versus rest)

Speech production artifacts were present in the EEG signal of the adult speaker, and these articulatory motions are known to reduce the signal-to-noise ratio of neural signals that relate to cognition [5]. For instance, the temporalis muscle is used for closing the lower jaw and this muscle spreads widely over the scalp locations that correspond to the frontal/temporal/parietal junction of the brain, generating large artifacts in the EEG signals measured over these regions [5]. Muscle artifact contamination is greatest over frontal and temporal scalp regions [6] and generally less severe over central regions, where our recording electrodes were placed. Several methods have been proposed for removing speech artifacts
from the EEG signal. These include the use of low-pass filtering to remove muscle artifacts that most prominently occur at frequencies over 12 or 20 Hz [7,8], and blind source separation based on Canonical Correlation Analysis [6] or Independent Component Analysis [9] to separate cortical sources from electromyographic (EMG) responses. However, none of these methods are able to completely remove motion artifacts from the EEG signal, and may even remove some genuine neural activity of interest.

Therefore, in order to understand whether these speech production artifacts could have introduced a pattern of bias into our results, it is first necessary to quantify the spatial (i.e. scalp topography) and spectral signature of the exact speech production artifacts that were generated by the adult speaker whilst singing nursery rhymes. According, we performed a control analysis to systematically document the topographical and spectral differences in the EEG signals of the speaker during speech production (in each gaze condition) as compared to rest.

7.1 Protocol

All recordings were performed by the same female speaker as in the main studies.

Nursery rhymes. Twenty repetitions of each of the 7 nursery rhymes were recorded by the speaker in each of three gaze positions (Direct, Indirect and Direct-Oblique), in which the speaker maintained the same head and body position as in the original experiments. During recording, her gaze was fixated on a life-sized head image of an infant.

Resting State. The adult was instructed to remain relaxed with her eyes open and to focus her gaze on the image of the infant. She was told to avoid eye, head or other movements. Resting state EEG was recorded for 12 minutes.

7.2 EEG acquisition

32-channels of EEG data were acquired from the adult at 500 Hz using a Biopac Mobita amplifier and Acqknowledge v5.0 software. No online referencing or filtering was used. Impedance for all channels was under 10KΩ.

7.3 EEG pre-processing and analysis

Average re-referencing was performed offline. No filtering was applied to the raw signal. Eye-movement and blink artifacts, as well as segments with raw amplitude above 100 µV were manually identified and removed from the raw recordings. After cleaning, the 20 repetitions of each nursery rhyme were concatenated for each gaze condition. A Fast Fourier Transform was applied to the nursery rhyme and resting state data in non-overlapping 1.0s windows for each EEG channel. As the frequency spectra of individual nursery rhymes did not differ, we collapsed the data across nursery rhymes and analysed the grand average frequency spectrum over all nursery rhymes.

7.4 Scalp topography during resting state and speech production

The scalp topography of EEG power in 5 frequency bands (Delta[1-3Hz]; Theta[3-6Hz]; Alpha[6-9Hz]; Beta[9-25Hz]; Gamma[25-42Hz]) is shown in Figure S3 for resting state condition, and during speech production for each gaze condition.
From visual inspection, it may be observed that during speech production (as compared to resting state), there were distinct increases in power, especially at Beta and Gamma frequencies, and particularly over left and right fronto-temporal regions. However, central regions (e.g. C3 and C4) appeared to be the least affected by speech production power artifacts. To assess these differences more closely, a detailed spectral analysis on the power spectrum at C3 and C4 was performed to test for frequency-specific changes in power during speech production as compared to resting state, as described next.

7.5 Spectral analysis at C3 and C4

To identify spectral differences between speech production conditions and resting state, one-way ANOVAs with 4 levels (RS, Direct gaze, Indirect gaze, Direct-Oblique Gaze) were conducted at each frequency between 0 to 40 Hz, for C3 and C4 channels. We performed all post-hoc comparisons (RS vs each gaze condition; each gaze condition against the other two gaze conditions) by running unpaired t-tests. For all tests, boot-strapping was performed by randomly selecting an equal subset (~1500) of 1.0s segments in each gaze condition, and permutating this selection 100 times. Only comparisons in which t-tests were significant at the alpha-level of p<0.05 for over 95% of all permutations are reported.

As shown in Figure S4, the results of the ANOVA revealed that there were significant spectral differences between speaking and rest conditions at C3 at 12 Hz, 13 Hz and between 15 – 39 Hz. At C4, significant differences were observed between 21 – 25 Hz, and between 29 - 39 Hz. In each case, speech production increased power in the EEG signal relative to
rest. Note that for both electrodes, no overall differences in power were observed across conditions at frequencies under 12 Hz.

To assess the specific pattern of differences between conditions, post-hoc t-tests were conducted at every frequency between 0 and 40 Hz (as described above). For C3, significant differences between Rest and Direct gaze were observed at 12, 13, and 15–39 Hz. For Indirect gaze, differences were additionally observed in the Delta band at 2 Hz. For Direct-Oblique gaze, differences were observed at 12, 13, 15–23 and 28–39 Hz. For C4, Direct gaze differed from Rest at 21–23 and 29–39 Hz. For Indirect gaze, differences were observed at 21–25 and 29–39 Hz. For Direct-Oblique gaze, differences were only observed between 29–39 Hz. Over all comparisons, no significant differences between speaking and rest were observed in Theta (3–6 Hz) and Alpha (6–9 Hz) bands.

In summary, our analysis of the adults’ speech production artifacts confirmed that speech gestures did indeed produce increases in EEG power that were most prominent over frontal and temporal scalp regions, consistent with previous studies [5,6]. However, our fine-grained spectral analyses revealed that, relative to resting state EEG (when no overt motor activity was present), electrodes in the central scalp region (C3 and C4) showed no significant change in power at Theta (3–6 Hz) and Alpha (6–9 Hz) band frequencies for any gaze condition. Accordingly, our main connectivity analyses focused on this scalp region and frequency range.

8 EEG power analysis

As our main aim was to assess changes in connectivity between gaze conditions, it was important to first establish whether there were any properties of the underlying EEG
signal in each condition that might artifactually generate increases (or decreases) in computed connectivity. One such potential confounding factor is the composition of the power spectrum of the EEG signal. The accuracy of the partial directed coherence (PDC) metric can be sensitive to even moderate changes in signal-to-noise ratio [10]. For example, Adhikari et al [10] reported that a 10% decrease in signal power from 67% to 57% was associated with ~15% lower accuracy in PDC directionality estimation, although a similar 11% power change from 57% to 46% only caused an accuracy drop of <5%. Therefore, if the EEG signal in one experimental condition has higher noise than in another condition (or if the spectral composition of the signal changes substantially), this can lead to greater error in estimation of connectivity patterns.

To assess the power spectra of the EEG signals, their power spectral density (PSD) was estimated using the Matlab 'periodogram.m' function, which performs a discrete Fourier transform on the signal. One PSD estimate was computed for each channel (left and right electrodes for adult and infant respectively), for each participant pair, and for each experimental condition. The resulting power spectra were then divided into EEG frequency bands, and averages were taken for each frequency band used for analysis.

To assess whether there were differences in EEG power between the gaze conditions, for each experiment, a repeated measures ANOVA was conducted taking Gaze ([3 (Expt 1) or 2 (Expt 2) levels]), Frequency band ([2 levels, Theta 3-6 Hz and Alpha 6-9 Hz]) and Channel ([4 levels, infant and adult x left and right]) as within-subjects factors. For Experiment 1, there was no overall difference in EEG power between the Direct, Indirect and Direct-Oblique conditions (F(2,32) = 0.25, p = .78). There was also no interaction between Gaze x Channel (F(6,96) = .23, p = .97), no interaction between Gaze x Frequency (F(2,32) = 1.94, p = .16), and no interaction between Gaze x Channel x Frequency (F(6,96) = 2.04, p = .07). For Experiment 2, there was again no overall difference in EEG power between the Direct and Indirect conditions (F(1,18) = 0.30, p = .59). There was no interaction between Gaze x Channel (F(3,54) = .14, p = .93), no interaction between Gaze x Frequency (F(1,18) = .00, p = .98), and no interaction between Gaze x Channel x Frequency (F(3,54) = .90, p = .45). Therefore, the gaze manipulation did not generate any detectable power changes that might systematically bias the PDC metric.

9 Neural connectivity analysis: Partial Directed Coherence (PDC)

Partial Directed Coherence (PDC) is a directional causal measure of direct flows between channels [11-13]. It is based on the principles of Granger Causality [14], and measures the degree of influence that channel j (the ‘Sender’) directly has on channel i (the ‘Receiver’) with respect to the total influence of j on all channels in the network. Here, each individual electrode (Infant L, Infant R, Adult L, Adult R) was taken as one channel and the entire network consisted of 4 electrodes in total. We computed directed coherence values for 12 possible pairwise connections, both within individual (e.g. Infant L -> Infant R) as well as across individuals (e.g. Infant L -> Adult L).

For the current analysis, we used Generalised Partial Directed Coherence (GPDC; [12]), which is an adapted version of PDC with better variance stabilization properties and the advantage of scale-invariance [15]. As a first step in the analysis, a multivariate autoregressive (MVAR) model is fitted to the EEG time series, which has the advantage of providing information about causal linear interaction effects in addition to estimating the coupling strength between channels. A frequency representation of the MVAR model parameters is then generated via a Fourier Transform, as follows:
\[ A(f) = I - \sum_{p=1}^{P} A_p e^{-2\pi i p f_s} \]  
(eq.1)

where \( A_p \) are the model coefficients, \( I \) refers to the M-dimensional identity matrix, \( f_s \) is the sampling frequency, and \( i^2 = -1 \). For each pair of channels \((i \text{ and } j)\), GPDC\(_{ij}\) is then computed as:

\[ GPDC_{ij}(f) = \frac{1}{\sigma_i} \frac{|A_{ij}(f)|}{\sqrt{\sum_{m=1}^{M} \frac{1}{\sigma_m^2} |A_{mj}(f)|^2}} \]  
(eq.2)

where \( \sigma_i^2 \) refers to the variance of the innovation process \( x_i(t) \). GPDC takes values between \([0,1]\) and is normalized across receivers (i.e. total outflow = 1 at each frequency), with larger values indicating strong connectivity.

The MVAR model was estimated using the Burg-type Nuttall-Strand method [16] which is thought to perform best for small sample sizes [17], and a model order (MO) of 5 was used. The model order (MO) indicates the number of preceding samples that are used to predict the data at sample time \( t \), and determines the number of observed frequency components for each pair of channels, which is typically half the model order. Following prior studies on autoregressive modeling [18,19] and multivariate autoregressive modeling of EEG time series [20-23], here a model order of 5 was used for this analysis. For example, Jansen et al [18] reported that a fifth order AR model was sufficient in 90% of cases to adequately capture variance in EEG time series data. Vaz et al [19] also noted that “a 5th order AR model represents adequately 1- or 2-s EEG segments with the exception of featureless background, where higher order models are necessary”. Model orders used in other MVAR EEG studies typically range between 3 and 6 [20-23].

One MVAR model and the resulting set of GPDC estimates (spanning the entire frequency spectrum) was computed for each non-overlapping 1.0s EEG epoch (200 data samples), and these estimate GPDC values were averaged across all epochs for each participant pair, for each experimental condition. The resulting epoch-averaged GPDC spectrum was then divided into discrete Theta (3-6 Hz) and Alpha (6-9 Hz) EEG frequency bands. Note that as infants’ Theta and Alpha EEG bands are lower in frequency as compared to adults [24], our frequency banding was adjusted lower accordingly. The mean GPDC value was taken within each frequency range, for each pairwise connection, condition and participant.

10 Control analysis 1: Surrogate connectivity data

As a control analysis, we generated a surrogate dataset comprising 1000 temporally-shuffled versions for each participant pair. The aim of this control analysis was to disrupt the fine-grained temporal correspondence between adult and infant neural signals by randomly pairing each adult 1.0s epoch with a non-matching infant 1.0s epoch from a different timepoint within the same experimental session. For example, the adult neural signal whilst singing “Twinkle Twinkle” may be paired to the infant signal whilst listening to “Wheels on the Bus”. The pairing of adult and infant time-shuffled epochs was determined by random permutation, and was non-identical for each of the 1000 shuffled versions generated for each participant pair, as well as for different participant pairs.
This shuffled control allowed us to establish a baseline level of non-specific connectivity between brains that could have arisen, for example, from commonalities in the physical environment during a particular experimental session, or due to general increases in infants’ and adults’ arousal. We could then be assured that any neural connectivity which could be detected over and above this baseline was specifically related to the time-contingent neural coupling between speaker and listener for the given experimental stimulus. Identical connectivity analyses were then performed on the real and surrogate datasets. All GPDC analyses were performed using the eMVAR (Extended Multivariate Autoregressive Modelling) Toolbox [25] in Matlab (The Mathworks Inc). The resulting GPDC values are shown in Tables S1 and S2.

11 Control analysis 2: Neural entrainment to the speech stimulus

In order to assess whether interpersonal connectivity gaze effects could be attributed to differences in basic speech processing across gaze conditions, we examined whether neural oscillatory entrainment to the amplitude envelope (temporal structure) of the adult’s speech signal differed between gaze conditions. The EEG data was first low-pass filtered under 45 Hz using an inverse fft filter to remove line noise (EEGLAB eegfiltfft.m function [26]). Next, the wholeband amplitude envelopes of the speech signal of the nursery rhyme stimuli were extracted using the Hilbert transform. To assess the degree of entrainment between the neural EEG signal and the speech amplitude envelope the phase-locking value (PLV, [27]) was computed. The PLV takes values between [0, 1], where a value of 0 reflects the absence of phase synchrony and a value of 1 reflects perfect synchronisation.

Figure S5. Speech-brain entrainment for Experiment 1 (top panel) and Experiment 2 (bottom panel) infants and adults by gaze condition for Theta (left) and Alpha (Right) frequency bands respectively. Error bars show the standard deviation.
Prior to calculating the PLV, a continuous wavelet transform was applied to the neural and speech data, which convolves each time series with scaled and translated versions of a wavelet function [28]. Here, the wavelet function chosen was the complex Morlet wavelet (bandwidth of mother wavelet = 1 Hz, time resolution = 0.1 Hz). The wavelet time-frequency decomposition was performed at 40 log-spaced frequencies. The phase series at each frequency was extracted from the complex wavelet coefficients, and divided into matching EEG and speech epochs of length 2.0s (with no overlap). The PLV for each epoch was then computed, and averaged over all epochs for each participant. Finally, frequency band-averaged PLV values were computed for Theta (3-6 Hz) and Alpha (6-9 Hz) frequency bands for each gaze condition, as shown in Figure S5.

For each experiment, a Repeated Measures ANOVA was conducted taking Gaze (3 or 2 levels) and Frequency (2 levels) as within-subjects factors, and Group (Infant or Adult) as the between-subjects factor. For both experiments, there was no significant difference in speech-brain entrainment between gaze conditions (Expt 1: $F(2,64)=1.89$, $p=.16$; Expt 2: $F(1, 36)=.06$, $p=.80$), and no significant interaction between Gaze and Frequency (Expt 1: $F(2, 64)=.72$, $p=.49$; Expt 2: $F(1, 36)=.42$, $p=.52$), suggesting that gaze did not change the pattern of speech-brain entrainment for Theta or Alpha bands. Therefore, any interpersonal connectivity gaze effects cannot be attributed to differences in basic speech processing.

12 Full infant scalp topography of receiving GPDC values (Expt 1)

To assess the scalp topography of infants’ neural receiving patterns with respect to adults’ C3 and C4 electrodes, 4-channel GPDC analyses were conducted for all hemispherically-dichotomous pairs of infants’ electrodes (e.g. infants’ left temporal [T7] and right temporal [T8] electrodes). The results indicated that across both EEG frequency bands, and across all gaze conditions, the strongest adult-to-infant connectivity was observed over infants’ central and posterior scalp locations (including C3 and C4). By contrast, lower connectivity was observed over infants’ frontal and temporal regions, particularly for the Alpha band. This topographical pattern confirms that the connectivity data from C3 and C4 (reported in the main manuscript) is indeed representative of infants’ overall neural response to the adult.

13 ANOVA results of gaze effects on interpersonal neural connectivity

To recap our analysis approach, the average (a) infant-to-adult GPDC (I→A) and (b) adult-to-infant (A→I) GPDC was computed for each gaze condition, and for Theta and Alpha bands separately. We then conducted Repeated Measures (RM) ANOVAs using these average indices, taking Frequency (2 levels) and Gaze (3/2 levels) as within-subjects factors. From Table S2, it may be noted that a few individual connections in Expt 2 were not significantly above threshold for one of the gaze conditions (but this never occurred across both gaze conditions). These values were included in the grand averages in order to maintain representativeness and a balanced ANOVA structure, since they were still statistically meaningful in terms of potentially revealing a difference between gaze conditions. For all analyses, infants’ looking times across each gaze condition were entered as co-variates to control for individual differences in attentiveness. For the I→A analysis, age was entered as an additional co-variate to control for individual differences in infants’ maturation. To assess the specific gaze effects at each frequency, we conducted planned pairwise comparisons using Dunnett’s multiple range t-test [29], which independently controls for familywise error rate without a prior F-test. Requiring a significant F-test before performing multiple
comparison tests (like Dunnett’s) is not recommended as this inflates the false negative rate [30,31]. At each frequency, we performed 3 planned pairwise comparisons using Dunnett’s test: (1) Direct > Indirect (Expt 1 & Expt 2); (2) Direct-Oblique > Indirect (Expt 1 only) and (3) Direct = Direct-Oblique (Expt 1 only). The results of these pairwise tests are reported in the main text. Here we provide a breakdown of the RM ANOVA results.

13.1 Experiment 1: Adult-to-infant GPDC (A → I)

<table>
<thead>
<tr>
<th>RM ANOVA Effect</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze</td>
<td>F(2,26) = .66, p = .53, η²p = .05</td>
</tr>
<tr>
<td>Frequency</td>
<td>F(1,13) = 6.92, p&lt;.05, η²p = .35</td>
</tr>
<tr>
<td>Gaze x Frequency</td>
<td>F(2,26) = 1.61, p = .22, η²p = .11</td>
</tr>
</tbody>
</table>

*Table S7 – Experiment 1 adult-to-infant RM ANOVA results*

It may be noted that although there were strong pairwise differences between individual gaze conditions (as revealed by Dunnett’s test and reported in the main text), the overall F-test for the Gaze effect was not significant. This apparent discrepancy could arise from the fact that the null hypothesis for the ANOVA F-test is that the means across all gaze conditions (and frequencies) are equal. However, this null hypothesis is inconsistent with our a-priori predictions that Direct/Direct-Oblique gaze would both differ from Indirect gaze, but Direct gaze would not differ from Direct-Oblique gaze. Therefore, we expected 2 out of our 3 condition means to be equal, and only 1 to differ. Conversely, if we had conducted the ANOVA analysis with only 2 gaze conditions that were predicted to differ (such as Direct-Oblique versus Indirect), then there would indeed be a significant main effect of Gaze (F(1,14) = 6.05, p<.05, η²p = .30). However, conducting 3 separate ANOVAs for each pairwise gaze contrast would be unparsimonious and lead to Type I error inflation. Accordingly, the ANOVA F-test was ill-suited to evaluate our predicted hypotheses in Experiment 1. To address this, we relied on the findings of the Dunnett’s tests (which independently control for Type 1 error) to assess our specific predictions in Experiment 1.

13.2 Experiment 2: Adult-to-infant GPDC (A → I)

<table>
<thead>
<tr>
<th>RM ANOVA Effect</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze</td>
<td>F(1,16) = 5.51, p&lt;.05, η²p = .26</td>
</tr>
<tr>
<td>Frequency</td>
<td>F(1,16) = .00, p=.96, η²p = .00</td>
</tr>
<tr>
<td>Gaze x Frequency</td>
<td>F(1,16) = 5.48, p&lt;.05, η²p = .26</td>
</tr>
</tbody>
</table>

*Table S8 – Experiment 2 adult-to-infant RM ANOVA results*

As expected, the results of the RM ANOVA showed a significant main effect of Gaze (Direct > Indirect), which corroborated with findings from pairwise Dunnett’s tests (reported in the main manuscript) showing that A → I connectivity was higher for Direct > Indirect gaze in both Theta and Alpha bands.
13.3 Experiment 2: Infant-to-adult GPDC (I → A)

<table>
<thead>
<tr>
<th>RM ANOVA Effect</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze</td>
<td>F(1,15) = 6.18, p&lt;.05, $\eta^2_p = .29$</td>
</tr>
<tr>
<td>Frequency</td>
<td>F(1,15) = 38.8, p&lt;.001, $\eta^2_p = .72$</td>
</tr>
<tr>
<td>Gaze x Frequency</td>
<td>F(1,15) = 10.5, p&lt;.01, $\eta^2_p = .41$</td>
</tr>
</tbody>
</table>

Table S9 – Experiment 2 infant-to-adult RM ANOVA results

For infant-to-adult connectivity in Experiment 2, the results of the RM ANOVA showed a significant main effect of Gaze (Direct > Indirect), which corroborated with findings from pairwise Dunnett’s tests (reported in the main manuscript) showing that I → A connectivity was higher for Direct > Indirect gaze in both Theta and Alpha bands.

14 Infant looking times

14.1 Experiment 1: Video

For Direct gaze stimuli, infants’ average looking time was 101.61s (SD = 43.04s). Their looking time was 92.73s (SD = 41.73s) for Indirect gaze stimuli, and 95.13s (SD = 43.02s) for Direct-Oblique gaze stimuli. A repeated measures ANOVA analysis with Gaze (3 levels) as the within-subjects factor revealed that there was a significant main effect of Gaze (F(2,32) = 4.46, p<.05) on infants’ looking times. Tukey HSD post hoc analysis indicated that infants looked for significantly longer at the Direct gaze nursery rhymes as compared to the Indirect gaze stimuli (p<.05), but there was no difference in looking time between Direct gaze and Direct-Oblique gaze (p=.10) or between Indirect gaze and Direct-Oblique gaze (p = .72).

As the acoustic parameters of the video stimuli were tightly controlled across conditions, these differences in infants’ looking patterns could not have arisen from inconsistencies in the speakers’ presentation of the stimuli.

14.2 Experiment 2: Live

For Direct gaze stimuli presented in a live format, infants’ mean looking time was 61.01s (SD = 31.61s) and for Indirect live stimuli, infants’ mean looking time was 61.11s (SD = 34.21s). A paired t-test confirmed that there was no significant difference in infants’ looking time for Direct and Indirect gaze conditions (t(18) = 0.03, p = .98). Therefore, infants were not more inattentive during Indirect gaze for live stimuli.

It is interesting that infants showed a different pattern of looking for Direct versus Indirect gaze stimuli across the two experiments. In Experiment 1 (video), consistent with previous screen-based studies (Farroni et al, 2002), infants looked longer at Direct gaze than Indirect gaze stimuli. However, in Experiment 2 (live), infants looked equally long at both types of gaze stimuli. This apparent attentional benefit for live speech was also observed in a phonetic learning experiment by Kuhl et al (2003), in which infants were more attentive to (and showed more phonetic learning from) live adult speakers than DVD movies of the same speakers. However, even though infants were equally attentionally-engaged for Direct and Indirect gaze stimuli in Experiment 2, their neural connectivity to the adult differed across gaze conditions, suggesting that attention did not underlie the neural gaze effect.
15 Effect of infant age

We examined the effect of age based on a median split analysis that divided our data into younger and older infants (Experiment 1 = 8.0 months, Experiment 2 = 8.52 months), entering this as an additional between-subjects factor in the RM ANOVA analyses. For both Expt 1 and 2, there was no main effect of Age on adult-to-infant GPDC (Expt 1 : F(1,12) = 1.38, p=.26, η²p = .10; Expt 2 : F(1,15) = .00, p=.96, η²p = .00). There was also no significant interaction between Age and other factors (Frequency Band, Gaze, p>.13 for all). For infant-to-adult GPDC in Expt 2 (Alpha band), there was similarly no effect of Age (F(1,15) = .15, p=.70, η²p = .01). Thus, the effects of Gaze did not differ as a function of infants’ age.

16 Internal replicability of gaze findings

We conducted a permutation analysis to assess the internal replicability of our two main gaze findings [1] Direct > Indirect (E1 and E2) and [2] Direct-Oblique > Indirect (E1). In the permutation analysis, 71%/75% of the E1/E2 cohort data (N=12 or 14 out of 17 or 19) was randomly selected in all possible ways (=6,188 or 11,628 permutations). For each cohort permutation, one main test statistic was computed for each Gaze contrast. To permit direct comparison across Experiments 1 and 2, we selected the same test statistic for the Direct v Indirect contrast in each experiment : Alpha band adult-to-infant GPDC. For completeness and to avoid bias, a different frequency band was selected for the Direct-Oblique v Indirect contrast : Theta band adult-to-infant GPDC. For each permutation, an RM ANOVA was performed on the test statistic, taking Gaze as the within-subjects factor and controlling for infant looking time. The effect size (r²) was recorded for each permutation to yield a distribution of possible effect sizes over all permutations.

15.1 Experiment 1

For the Direct vs Indirect gaze contrast (left subplot in Figure S6), the effect size (r²) obtained across all permutations was 0.219 (mean) / 0.212 (median), indicating the presence of a medium-large effect size in the data. For the Direct-Oblique vs Indirect gaze contrast (right subplot, Figure S6), the effect size (r²) obtained across all permutations was 0.192 (mean) / 0.183 (median), indicating the presence of a medium effect size in the data.

Figure S6. Experiment 1 : Distribution of effects sizes for Direct vs Indirect (left) and Direct-Oblique vs Indirect (right) contrasts obtained for 71% (n=12) sub-samples of the data.
15.2 Experiment 2

For the Direct vs Indirect gaze contrast (which used the same test statistic as Experiment 1), the mean effect size \( r^2 \) obtained across all permutations was 0.332 (mean) / 0.321 (median), indicating the presence of a large effect size in the data (see Figure S7). It is also interesting to note that the effect size distribution appeared to be bimodal, which could indicate that there is a subset of infants who show particularly strong sensitivity to adult gaze.

![Figure S7. Experiment 2: Distribution of effect sizes for the Direct vs Indirect contrast obtained for 75% (n=14) sub-samples of the data.](image)

Comparing across Experiments 1 and 2, the Direct vs Indirect gaze contrast consistently yielded at least a medium-sized effect across both experiments, indicating that this gaze finding is replicable across two different testing modalities (video versus live presentation) and two different infant cohorts.
SI REFERENCES