nature communications

Article

https://doi.org/10.1038/s41467-026-75831-x

Adult-to-infant unidirectional neural coupling mediates selective social learning in infants from the United Kingdom and Singapore

Received: 14 July 2025

Accepted: 9 July 2026

Published online: 22 July 2026

Check for updates

Wei Zhang (张维)[1,2], Kaili Clackson[1], Stanimira Georgieva[1], Lorena Santamaria[1], Vanessa Reindl[3,4,5], Valdas Noreika[6], Nicholas Darby[7], Vaka Valsdóttir[7], Priyadharshini Santhanakrishnan[1] & Victoria Leong[1,8]

Infants possess powerful learning abilities, but their neural resources must be deployed parsimoniously toward the most relevant information in a given social context. Adults curate infants' information selection through ostensive marking of valuable content using social cues such as eye-contact. However, the neural mechanisms that support social gating of early learning remain unclear. Here, we demonstrate that – in infants from UK and Singapore – unidirectional adult-speaker to infant-listener neural coupling mediates infants' selection of socially-relevant information for learning. Specifically, speaker-to-listener neural coupling is modulated by the adult speaker's eye contact and predicts selective learning of an artificial language that is paired with full (compared to partial or no) eye contact. Speaker-to-listener neural coupling is a better predictor of selective learning than infants' own neural activity, suggesting that this signal effectively captures important social influences on early cognition, and may provide insights into social valuation decision-processes that dictate early social learning.

In a busy world, infants are surrounded by myriad stimuli that compete for attention. Amidst different actors, sights and sounds, infants must parsimoniously select and attend to information that is relevant for learning and action. Social ostensive cues - such as gaze, infant-directed speech, facial expressions, and contingent reactivity - indicate the social partner's communicative intent and enhance infants' receptivity to learning and communication[1]. These multimodal cues are integrated probabilistically, with infants weighing their reliability and relevance to infer the speaker's intent and the significance of the conveyed information[2]. This integration is supported by emerging evidence that ostensive cues dynamically reinforce one another; for example, vocal prosody increases the likelihood of gaze following[3], and the combination of direct gaze and infant-directed speech synergistically enhances neural and behavioural responses in infants, supporting their ability to process and learn from social cues[4]. The stages of social perception (detecting the social context and integrating cues), social valuation (inferring intent and assigning subjective value, based on factors like relevance and reliability) and social action

[1]Early Mental Potential and Wellbeing Research (EMPOWER) Centre, Nanyang Technological University, Singapore, Singapore. [2]Cognitive Neuroimaging Centre, Nanyang Technological University, Singapore, Singapore. [3]Psychology, Nanyang Technological University, Singapore, Singapore. [4]Section Child Neuropsychology, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Uniklinik RWTH Aachen, Aachen, Germany. [5]Institute of Medical Psychology and Medical Sociology, Uniklinik RWTH Aachen, Aachen, Germany. [6]Department of Psychology, School of Biological and Chemical Sciences, Queen Mary University of London, London, UK. [7]Department of Psychology, University of Cambridge, Cambridge, UK. [8]Department of Pediatrics, University of Cambridge, Cambridge, UK. e-mail: victorialeong@ntu.edu.sg

(learning and responding) are fundamental processes in social decision-making that are conserved across species[5]. In adult models, well-described neural circuits are known to subserve these stages[5–7]: social perception is supported by activity in the medial temporal lobes and fusiform gyrus whilst inferring intentions is known to engage the posterior superior temporal sulcus, temporoparietal junction, anterior cingulate cortex (ACC) and medial prefrontal cortex (mPFC). Subjective value calculations are encoded in the mPFC, orbitofrontal cortex, amygdala, and ventral striatum, and social learning implicates the dopaminergic midbrain, striatum (including the nucleus accumbens, NAc), ACC and dorsolateral PFC. An exceptional example of social decision-making is the process of pair bond formation, which involves a seismic change in the perception and valuation of another individual. In prairie voles, a monogamous species of rodents, a crucial step in pair-bond formation is thought to involve PFC coordination of inter-regional neuronal oscillatory coupling with the NAc and amygdala to direct the flow of social information, linking neural representations of the social partner with reward[8,9]. However, little is known about social decision-making processes in early human development. Although human infants can perform social perception and valuation using ostensive cues as key inputs, the neural processes that underpin early social valuation and choice-making are largely unknown.

Direct gaze is thought to be one of the most salient ostensive cues for conveying communicative intent[1] and enhancing learning (see review[10]) and also elicits communicative behaviour from infants[11,12]. From birth, infants prefer to look at pictures of faces with direct gaze over averted gaze[13], and by 4 months, heightened brain responses to direct gaze suggest that direct gaze modulates neural circuits involved in social learning and communication[14,15]. Changes in gaze direction can convey structural information, such as signalling turn taking or event boundaries[16,17]. As well as indicating communicative intent, looking directly at a social partner gives access to information from their facial signals, such as blinks and eyebrow movements which can convey both social information (such as indicating understanding or initiating repair) and structural information, as their occurrence is often linked to the start or end of conversational turns or units of language[18–20].

Direct gaze is undoubtedly a highly salient social cue, but debate exists about whether ostensive cues merely modulate social attention or whether they truly play a privileged role in referential learning. For example, Szufnarowska and colleagues demonstrated that infants showed gaze-following to both an ostensive cue (direct gaze) as well as a salient non-ostensive cue (shivering)[21], although learning was not assessed. A replication study[22] confirmed that gaze-following effects were not unique to ostensive cues. Infants showed gaze-following responses to infant-directed speech, shivering and a non-verbal beep. However, only infant-directed speech cues subsequently elicited referential learning. For both shivering and beeps, infants' gaze following behaviour was not associated with significant learning. A similar pattern between attention and learning has been observed in more naturalistic learning contexts. A previous study[23] found that 9-month-old American infants learned Mandarin phonetic properties from a live teacher but showed no significant learning when the same teaching was presented via video, despite infants paying equal visual attention in live and video contexts. These results have led the authors to propose the social gating hypothesis which suggests that contingent social interaction is key to language learning in real-world settings, with the social brain acting as a "gate" on the computational mechanisms involved in language learning[24]. Collectively, these data point to a larger role for ostensive cues in early decision-making–beyond social perception (i.e., modulating attention), they may in fact weight social valuation computations and thereby influence learning and action.

Social brain imaging may shed light on the neural processes that underpin these decision stages. Interpersonal neural coupling, wherein neural activity between infants and caregivers becomes aligned during interactions, has been proposed as a mechanism supporting social learning[25,26]. Such proposals have prompted a move towards more embodied and ecological multi-modal approaches to social development[27], with a number of models emphasising the key role of co-ordination or alignment between partners in the development of social cognition. For example, the enactive concept of participatory sense-making[28] argues for a shift towards understanding social cognition as a participatory process, while the interactive alignment model[29] emphasises the interdependence of production and comprehension processes.

When comparing communication in a live context and via an online video platform, adult-infant interpersonal neural coupling has been shown to be stronger in the live context, and also to have stronger relationships with video coded measures of child empathic social engagement, including measures of the child's gaze, involvement, empathy and cooperation[30]. Interpersonal neural coupling has also been shown to be sensitive to ostensive gaze cues, as direct gaze leads to stronger bi-directional adult-infant neural influence (compared to averted gaze)[12], and interpersonal coupling is more strongly influenced by mutual gaze than by joint attention to an object[31]. Indeed, the mPFC (a key region for inferring intention and value calculation) is activated in infants by direct gaze and by infant-directed speech[32,33], and PFC activity has been shown to be synchronised between a mother and child during verbal conversations[34]. Accordingly, it is plausible that the observed interpersonal neural coupling in regions such as the PFC could indicate estimation of the partner's intent and reliability, informing value calculations for selective learning and decision-making in social contexts.

Here, we use concurrent measurement of interpersonal (here, from adult-speaker to infant-listener in one direction) and intrapersonal (within-infant) neural activity during a statistical learning paradigm to identify how speaker gaze, an ostensive cue, modulates infants' selection of language stimuli for learning from a social partner in a pre-recorded video. Since the adult speaker is pre-recorded, two-way natural interaction is not possible. Nonetheless, this design permits assessment of unidirectional neural effects from speaker to infant listener, while carefully controlling the speaker's delivery (e.g., facial movements) across gaze conditions to isolate the effects of gaze availability. We also measure attention and neural entrainment to speech as potential contributors to the process. Entrainment, the temporal alignment of neural oscillations to rhythmic stimuli such as speech or gestures, has also been proposed as a key contributor to interpersonal neural coupling[35]. Some researchers argue that entrainment facilitates information exchange by enhancing attention and prediction[36], while others suggest that interpersonal neural coupling is simply an epiphenomenon of individual entrainment to common sensory inputs[37,38]. Accordingly, here we ask whether speaker-to-listener neural coupling explains infant behaviour over and above effects of neural entrainment to the speech stimulus, thereby distinguishing the contributions of these processes. In highly controlled experimental settings using artificial languages delivered without ostensive cues, infants of 8 months have shown remarkable abilities to learn transitional probabilities, and to use these statistical regularities to segment words from continuous speech[39]. Electroencephalography (EEG) studies further confirm that even newborns are able to track transitional probabilities in a similar way[40]. Therefore, we expected that the presentation of a social ostensive cue (gaze)–even in a non-contingent video context–would be effective in modulating infants' learning. Further, although eye-contact is considered a key communicative cue, evidence regarding cultural differences in mutual gaze during face-to-face interactions is mixed[19,20]. While some researchers have found evidence that cultures differ in their interpretation of direct gaze[41] and in looking patterns to faces during interactions[42,43], other results challenge the idea of cultural differences between Asian

https://doi.org/10.1038/s41467-026-75831-x

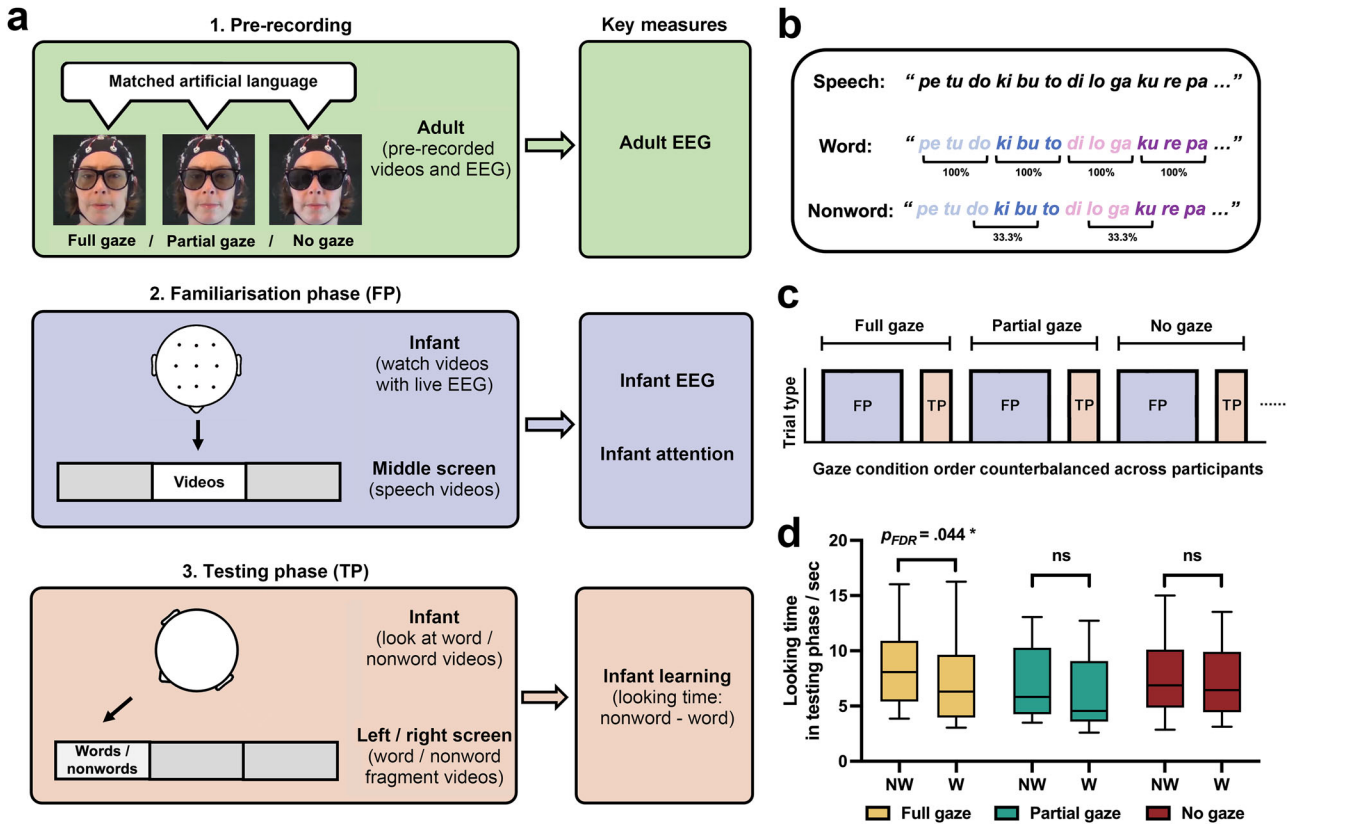


Fig. 1 | Overview of experimental design and learning results. a Summary of the experimental protocol: 1. Pre-recording: Videos were recorded of a British female adult speaking a continuous artificial language, with EEG recorded simultaneously. Three different but phonologically-matched language sets were recorded and paired with different gaze conditions. 2. Familiarisation phase FP: Infants watched pre-recorded video in one gaze condition, displayed in the centre of the screen. Infant attention and EEG were recorded. 3. Testing phase TP: Infants watched word / nonword fragments of videos from the same gaze condition, pseudo-randomly displayed on the left or right of the screen. Looking time differences between nonwords and words (in seconds) were calculated as a measure of learning. Familiarisation and Testing phase were then repeated for different gaze conditions in a counterbalanced order. **b** Example of artificial language speech, words, and nonwords. Speech: Continuous sequence of syllables without pauses between syllables or words. Word: A set of three syllables with 100% transition probability, indicated in the same colour (e.g., "pe-tu-do"). Nonword: A set of three syllables spanning a word boundary, with 33.3% transition probability (e.g., "do ki-bu" which crosses two words). **c** Illustration of experimental temporal sequence within a block. A block includes three different gaze conditions, each with a Familiarisation phase (FP; 60-sec of video) and a Testing phase (TP; 2 words and 2 nonwords, each repeated 12 times). The block sequence shown in (**c**) was repeated three times in the full experiment. The pairings between language-set and gaze condition and the presentation order of gaze conditions were counterbalanced across infants. **d** Box and whisker plots of learning performance. Learning was significant during the Full gaze condition, as indicated by one-tailed paired *t*-tests (FDR-corrected) comparing the looking time difference between nonwords (NW) and words (W) measured during the Testing phase. Full gaze: $t(46) = 2.25$, $p_{FDR} = 0.044$, Cohen's $d = 0.33$; Partial: $t(46) = 1.81$, $p_{FDR} = 0.058$; No: $t(46) = 0.46$, $p_{FDR} = 0.325$. The box represents the median and quartiles, while the whiskers indicate the 10th and 90th percentiles. "ns" indicates not significant and "*" indicates $p < 0.05$. N = 47 independent infants (29 UK, 18 Singapore). Source data are provided as a Source Data file. The authors affirm that informed consent was provided for the publication of the human images in Fig. 1.

and Western people leading to differences in levels of mutual gaze[44]. To assess the consistency of ostensive gaze effects across different cultural contexts, we conducted identical measurements with infants growing up in the UK and Singapore.

Here, we modified a standard infant statistical learning paradigm[39] to include a manipulation of the adult speaker's gaze (see Fig. 1) while EEG signals were collected (infant recorded live, adult pre-recorded). We predicted that infants would show selective learning for languages accompanied by Full gaze and no statistically significant learning when speaker gaze was occluded. However, it was also of interest to assess whether infants would show any learning during Partial gaze occlusion. Second, based on previous literature, we predicted that the gaze ostensive effect on learning would be mediated by speaker-to-listener neural connectivity, over and above the effects of neural entrainment to the speech signal. Finally, whilst we expected to uncover cultural differences in ostensive gaze processing between UK and Singapore (SG) cohorts, it was also of interest to assess which effects were consistent. Our results demonstrate that infants in both country cohorts show selective learning based on speaker gaze, and further that this ostensive learning effect is mediated through adult-to-infant unidirectional neural coupling.

Results

Forty-seven infants aged 9.4 months on average from Singapore and the UK took part in the study, matched on age and sex (see "Methods"). During the experiment (see Fig. 1a), infants were presented with pre-recorded videos of a female adult speaking three artificial languages (modelled on a standard paradigm[39]) under three conditions where the speaker's gaze was parametrically manipulated: Full gaze (wearing clear glasses, eyes visible), Partial gaze (wearing dark glasses) and No gaze (wearing fully opaque glasses). The adult speaker's EEG was pre-recorded during recording of the stimulus videos. For each gaze condition, infants first watched the videos while their EEG and attention to the screen were recorded during this Familiarisation phase. In the subsequent Testing phase infants were presented with individual words and nonwords from the relevant language. Learning was assessed for each condition as the looking time difference between nonwords and words, given their expected novelty preference (longer

looking time) for less familiar nonwords[39,45]. The experiment was presented in a repeated measures design in blocks, where each block included Familiarisation and Testing phases for all three conditions. Blocks were repeated up to three times per participant.

Speaker gaze modulates selection of a socially relevant language stimulus for learning

Following established practice for infant statistical learning paradigms[39,46,47], learning was assessed by comparing looking time for nonwords versus words (i.e., learning occurred when the looking time difference significantly exceeded zero). Consistent with the theoretical framework of natural pedagogy[1], here we found that fully visible speaker gaze acted as an ostensive cue for selection of the most relevant stimulus for learning. As shown in Fig. 1d and Table 1, learning was significant for the artificial language which was paired with Full speaker gaze ($t(46) = 2.25$, $p_{FDR} = 0.044$ (one-tailed), Cohen's $d = 0.33$, 95% CI [0.04, 0.62]), whereas no statistically significant learning was detected for languages that were paired with Partial ($t(46) = 1.81$, $p_{FDR} = 0.058$ (one-tailed), Cohen's $d = 0.26$, 95% CI [−0.03, 0.55]) or No speaker gaze ($t(46) = 0.46$, $p_{FDR} = 0.325$ (one-tailed), Cohen's $d = 0.07$, 95% CI [−0.22, 0.35]), after adjusting for age, sex, and country. Hierarchical omnibus testing assessing relative learning across gaze conditions also supported this conclusion (Supplementary Note 1), with Full and Partial gaze showing stronger learning than No gaze in planned contrasts. Supplementary Note 1 further illustrates that individual infants with the highest overall learning also showed strongest learning in the Full gaze condition, whereas learning for Partial and No speaker gaze did not reach statistical significance at the group level.

Table 1 | Performance metrics for the three gaze conditions (mean value ± standard error of the mean, block-level observations pooled across 47 infants)

	Full gaze	Partial gaze	No gaze
Learning / sec	1.22 ± 0.46	0.76 ± 0.51	0.34 ± 0.42
Total visual attention / sec	22.91 ± 1.50	21.48 ± 1.36	23.34 ± 1.39

Learning was measured as the looking time difference between nonwords and words, in seconds.

Infants' learning was not statistically significantly associated with their total visual attention (Fig. 2a), which did not statistically significantly differ across gaze conditions (Omnibus ANOVA $F(2,128) = 0.64$, $p = 0.531$, $\eta^2 = 0.01$, 95% CI [0.00, 0.06]; Full vs. Partial gaze $t(128) = 0.55$, $p = 0.585$, Cohen's $d = 0.11$, 95% CI [−0.30, 0.52]; Full vs. No gaze $t(128) = 1.13$, $p = 0.262$, Cohen's $d = 0.01$, 95% CI [−0.40,

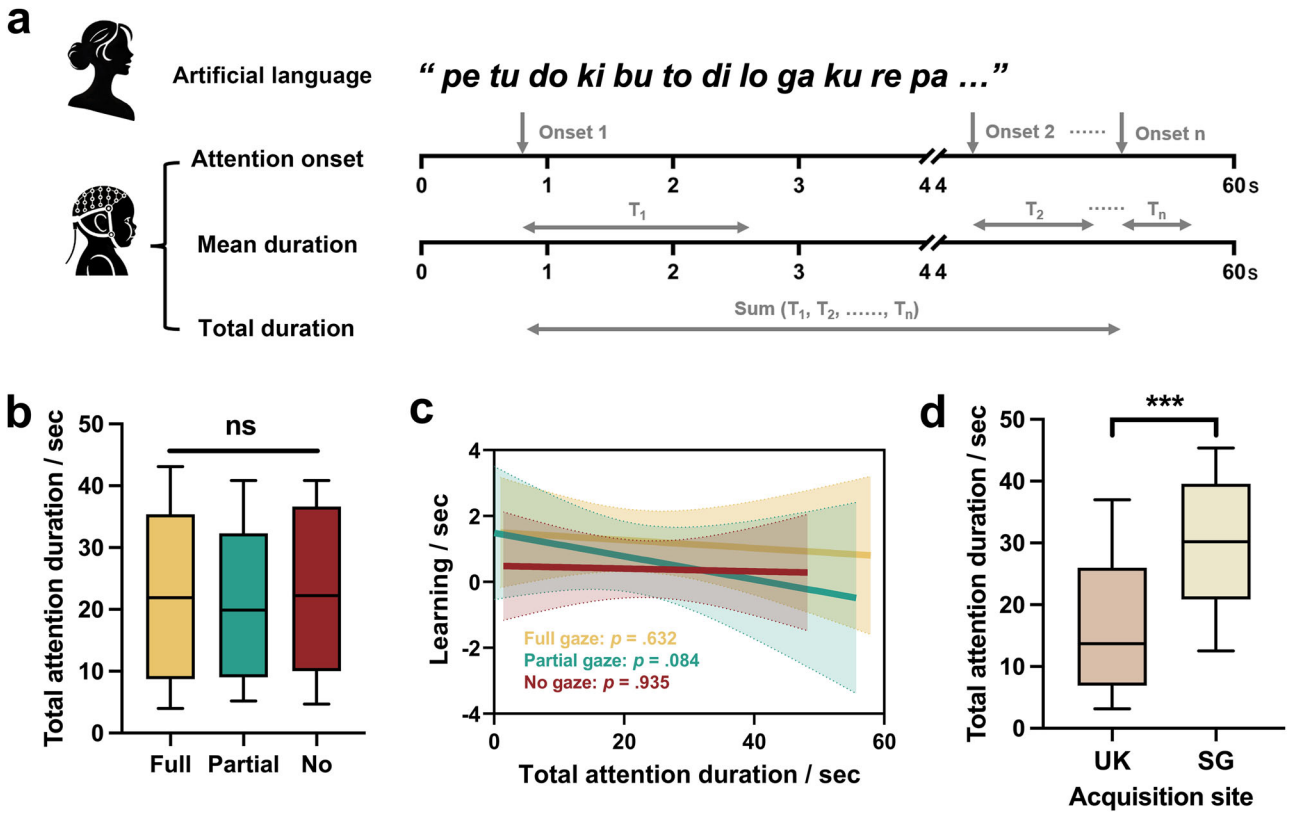


Fig. 2 | Relationship between infant visual attention and learning. a Three measures of infant visual attention were assessed during the Familiarisation phase: attention onsets (the number of times infants shifted their gaze to the screen), average attention duration (average length of each look), and total attention duration (sum of all looking times). Here, only the results for total attention duration are shown. Results for the other attention measures are provided in Supplementary Note 2. **b** Box and whisker plots show total attention duration across Full, Partial, and No gaze conditions. Omnibus ANOVA: $F(2128) = 0.64$, $p = 0.531$, $\eta^2 = 0.01$. Data represent block-wise repeated measurements. **c** Linear mixed-effects analysis of the relationship between learning and total attention duration showed no statistically significant relationship for all gaze conditions. (LME: $t(135) = -1.82$, $p = 0.070$, partial $r = -0.15$, 95% CI [−0.31, 0.02]). Shaded band represents the 95% CI. Colours represent gaze conditions: yellow for Full gaze, green for Partial gaze, and red for No gaze. Condition-specific slopes were also not statistically significant (Full gaze: $p = 0.632$; Partial gaze: $p = 0.084$; No gaze: $p = 0.935$). **d** Box and whisker plots of total attention duration during the Familiarisation phase differed significantly between United Kingdom (UK) and Singaporean (SG) infants (LME: $t(136) = 5.46$, $p < 0.001$, Cohen's $d = 1.22$, 95% CI [0.85, 1.59]). Lines indicate median (centre) and quartiles (25th/75th percentiles), while the whiskers indicate the 10th and 90th percentiles. "ns" indicates not significant; "***" indicates $p < 0.001$. N = 47 independent infants (29 UK, 18 Singapore). Box plots in (**b**, **d**) are based on block-level observations (up to 3 per infant per condition); linear mixed-effects models with participant as random intercept account for repeated measures. Source data are provided as a Source Data file.

0.41]; Partial vs. No gaze $t(128) = -0.58$, $p = 0.563$, Cohen's $d = -0.11$, 95% CI [−0.52, 0.29], see Fig. 2b and Table 1) and did not statistically significantly correlate with learning ($t(135) = -1.82$, $p = 0.070$, partial $r = -0.15$, 95% CI [−0.31, 0.02]) across all gaze conditions, see Fig. 2c. Statistical analyses of two other visual attention measures showed similar patterns (see Supplementary Note 2). Notably, total attention duration did differ between cohorts, with SG infants attending to stimuli significantly longer overall compared to the UK infants ($t(136) = 5.46$, $p < 0.001$, Cohen's $d = 1.22$, 95% CI [0.85, 1.59], see Fig. 2d). Despite this, there was no statistically significant difference in learning overall between the SG and UK cohorts ($t(137) = 0.74$, $p = 0.458$, Cohen's $d = 0.14$, 95% CI [−0.20, 0.48]). We also assessed infant's language development using the MacArthur Bates Communicative Development Inventory (CDI), focusing on gesture scores as the primary measure of expressive language at this age. There was no statistically significant difference between scores of the SG and UK cohorts ($t(39) = 0.63$, $p = 0.533$, Cohen's $d = 0.14$, 95% CI [−0.48, 0.75]) and infants' CDI scores were not statistically significantly associated with learning ($t(124) = 0.85$, $p = 0.398$, partial $r = 0.07$, 95% CI [−0.11, 0.24]).

In summary, here three viable and distinct language sets were offered by the same speaker, yet infants only learned one of the three proffered languages–marked by Full gaze availability. This effect of ostensive gaze was observed across both country cohorts. Visual attention measures did not statistically significantly differ across gaze conditions, although learning did. Conversely, visual attention differed between SG and UK cohorts, whereas no statistically significant cohort difference in learning was detected.

Adult-speaker to infant-listener neural connectivity predicts selective learning whereas within-infant connectivity associates with expressive language

To assess neural substrates of infant learning, we performed partial least squares (PLS) regression analysis on the adult and infant neural connectivity measures in the 6–9 Hz infant alpha band (see Supplementary Note 3 for identical analysis in the delta and theta bands). A PLS approach was chosen to effectively capture and contrast the summative contribution of entire interpersonal and intrapersonal neural networks toward infant learning and language outcomes. Neural connectivity was initially computed as a fully-connected directed connectivity matrix using generalised partial directed coherence (GPDC, see "Methods") and then partitioned into the within-adult, within-infant, adult-to-infant, and infant-to-adult submatrices (Fig. 3a).

To identify significant connections, GPDC values were compared to a surrogate distribution. After correcting for multiple comparisons, no infant-to-adult connections exceeded the 95% confidence interval (CI) upper bound of the surrogate distribution, which aligned with our expectations given that adult EEG was pre-recorded. In contrast, significant connections were detected in all three of the other submatrices across all gaze conditions (Fig. 3b). Only connections that were significantly above chance in at least one of the three gaze conditions were entered into the subsequent PLS analyses.

As shown in Fig. 4a, the PLS analyses revealed that only adult-to-infant GPDC connectivity significantly predicted infant learning above chance (overall surrogate $p = 0.031$), with the first adult-to-infant GPDC component explaining 24.6% of variance (95% CI [20.7%, 37.0%], surrogate mean = 19.5%, SD = 2.8%) in infant learning. By contrast, within-infant GPDC connectivity did not statistically significantly predict infant learning (overall surrogate $p = 0.751$; first component: $R^2 = 14.0\%$, 95% CI [13.2%, 30.6%], surrogate mean = 17.2%, SD = 2.6%). However, within-infant GPDC components did significantly predict infants' CDI gesture scores (see Fig. 4b), with the first within-infant component explaining 33.7% of variance (95% CI [22.6%, 44.1%], surrogate mean = 21.7%, SD = 3.2%, surrogate $p = 0.003$) in CDI gesture scores. The relevant component loadings and scalp topographies are shown in Fig. 4c, d. For the first adult-to-infant GPDC component, peak loading was observed on certain key connections, including adult Fz to infant F4. Supplementary Note 4 presents additional analyses performed at the single-channel-wise connection level, showing that this connection is significantly modulated by speaker gaze, (Full gaze > Partial or No gaze, $t(221) = 3.48$ and $p_{FDR} = 0.048$, Cohen's $d = 0.49$, 95% CI [0.21, 0.77]), consistent with previous published research[12].

To confirm the observed double dissociation between adult-to-infant and within-infant connectivity on learning and CDI gesture scores respectively, we performed a 10-fold cross-validation estimated through 1000 nonparametric bootstrap resampling iterations (Figs. 4e, f). This cross-validation confirmed that for PLS prediction of learning, adult-to-infant GPDC performance was significantly higher than that of within-infant GPDC ($t(1998) = 27.7$, $p < 0.001$, Cohen's $d = 1.24$, 95% CI [1.14, 1.34]), whereas for PLS prediction of CDI gesture scores, within-infant GPDC performance was significantly higher than that of adult-to-infant GPDC ($t(1998) = 44.7$, $p < 0.001$, Cohen's $d = 2.00$, 95% CI [1.89, 2.11]).

Neural entrainment to the speech amplitude envelope is modulated by speaker gaze but is not a statistically significant predictor of learning

In previous studies, neural entrainment to the speech amplitude envelope has been proposed as a mechanism of interest for phonological processing[48,49] as well as a potential contributor to cross-brain connectivity[50,51]. Accordingly, here we examined whether speaker gaze modulated infants' neural-speech entrainment (NSE) during the Familiarisation phase, measured by the peak cross-correlation between infants' EEG signals and the amplitude envelope of the speaker's audio waveform (illustrated in Fig. 5a). Compared to a surrogate distribution (95% CI upper bound), significant NSE was observed only in the Full gaze condition. Entrainment strength was tested against a surrogate distribution across three frequency bands—delta, theta, and alpha—and nine EEG channels (Fig. 5b), after correcting for multiple comparisons. Specifically, significant NSE was identified in the delta band at C3, theta band at F4 and Pz, and alpha band at C3 and Cz (surrogate $p_{FDR} = 0.027$ for alpha band at C3, and surrogate $p_{FDR} = 0.049$ for others). A visualisation of this analysis (delta band at C3) is provided in Fig. 5c, illustrating the real data against the surrogate distribution for each gaze condition. We further assessed whether NSE was a predictor of infant learning using the PLS procedure conducted previously for adult-to-infant and within-infant connectivity. This PLS analysis revealed that NSE did not statistically significantly explain infant learning (see Supplementary Note 5), even though NSE levels were sensitive to speaker gaze.

Adult-speaker to infant-listener connectivity mediates gaze-selective learning

Recall that thus far, an ostensive effect of gaze on learning has been observed, and a PLS component of adult-to-infant connectivity (adult-to-infant GPDC) has been identified to predict learning. Separately, NSE sensitivity to gaze was detected, with significant entrainment observed only in the Full gaze condition. To assess whether neural entrainment (indirectly) explains the previously demonstrated effect of adult-to-infant connectivity on learning, or whether these represent separate mechanisms by which gaze may affect learning, we conducted a mediation analysis incorporating both variables as alternative pathways for the effect of gaze on infant learning. Specifically, infant learning was entered as the dependent variable with gaze condition as the independent variable. NSE and adult-to-infant GPDC (PLS) measures were assessed as potential mediators separately. Further details of this analysis are given in the Methods section. The significance of each pathway was evaluated using LME modelling. No statistically significant effects of country were detected on the key measures of

https://doi.org/10.1038/s41467-026-75831-x

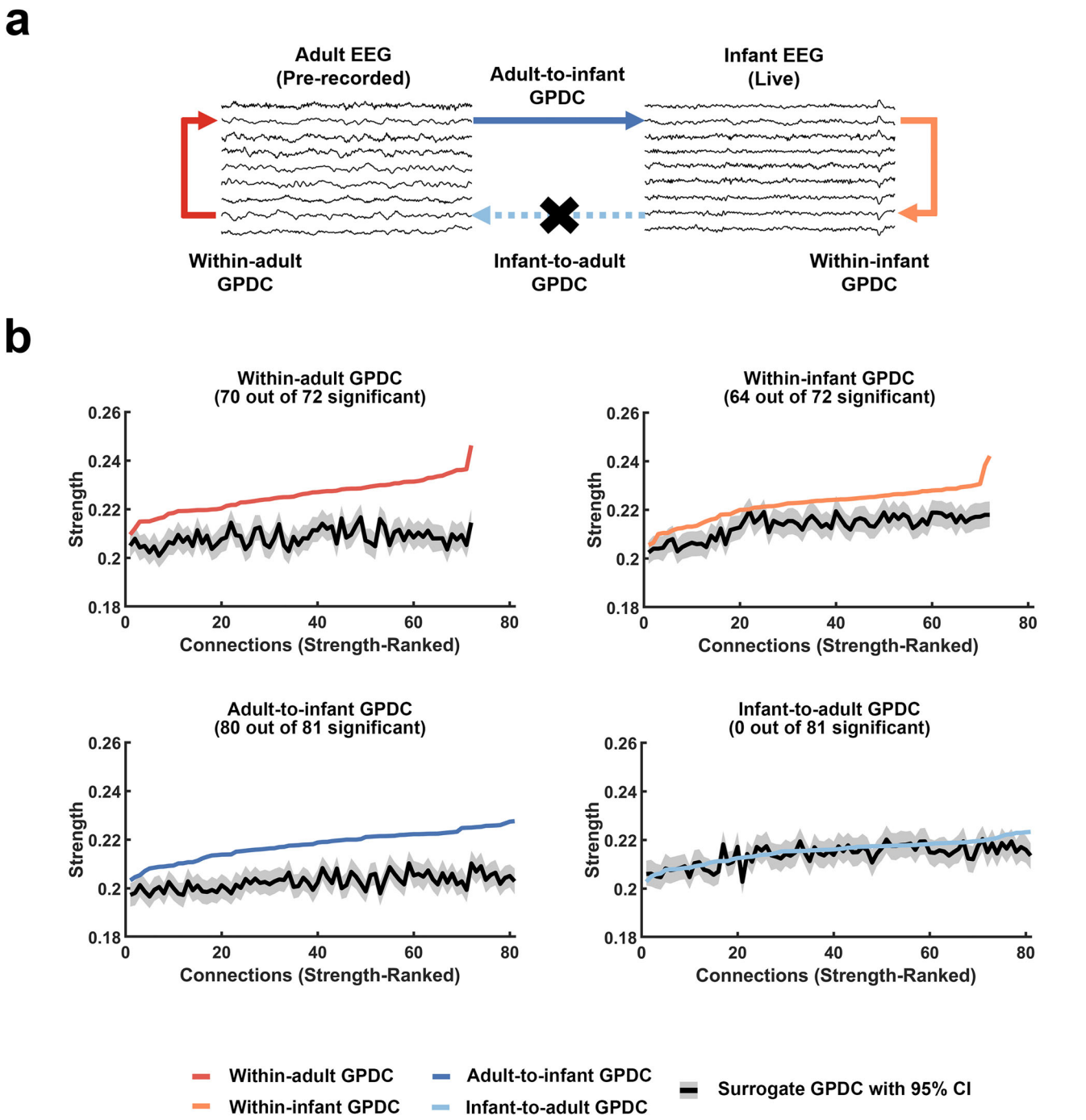
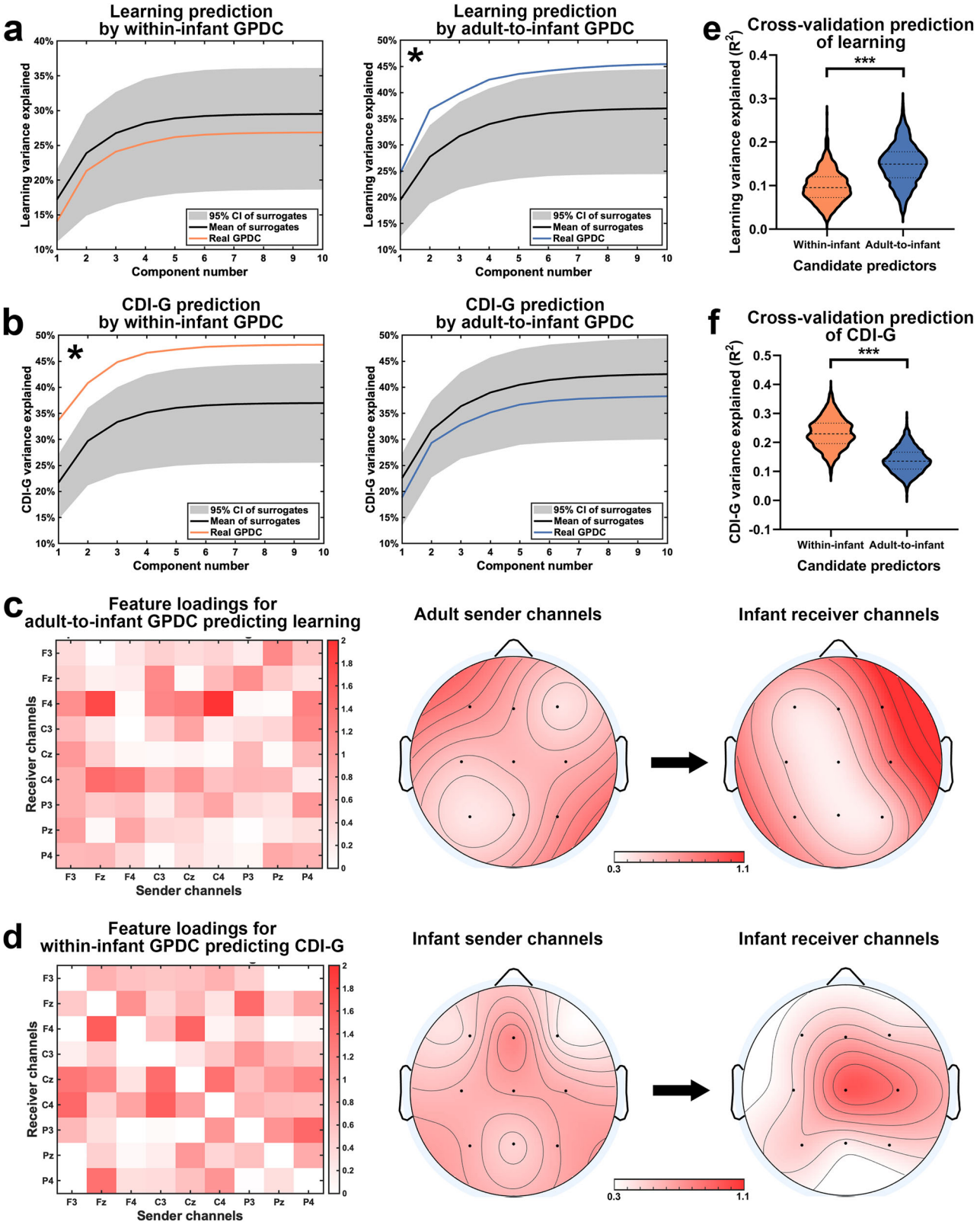


Fig. 3 | Quantification of dyadic neural connectivity using generalised partial directed coherence (GPDC). a Illustration of concurrent estimation of directed connectivity within the adult-to-infant network: within-adult (red), within-infant (orange), adult-to-infant (dark blue), and infant-to-adult (light blue). #: Since infants watched pre-recorded videos of the adult speaker, any detected infant-to-adult connections should be spurious and not significantly exceed the surrogate infant-to-adult GPDC distribution. **b** Ranked GPDC strength for individual real connections (coloured lines) as compared to their surrogate data (black lines: mean; grey areas: 95% CI). Connections were deemed significant if the real data exceeded the 95th percentile of the surrogate distribution, with Benjamini-Hochberg False Discovery Rate (BHFDR) correction applied. Across all gaze conditions, significant connections were detected in the within-adult, within-infant, and adult-to-infant directions, whereas no statistically significant infant-to-adult connections were detected. One-sided surrogate permutation tests (1000 iterations) were used. N = 42 infants contributed valid EEG data. Source data are provided as a Source Data file.

interest in this analysis (see Supplementary Note 6). Adult-speaker to infant-listener connectivity significantly mediated the relationship between speaker gaze and learning (indirect effect by bootstrapping: $\beta = 0.52 \pm 0.23$ (SD), 95% CI [0.10, 1.03], $p = 0.014$; Fig. 6). Specifically, Full gaze was associated with significantly higher levels of adult-to-infant connectivity than No gaze (Full vs. No gaze $\beta = 1.01 \pm 0.40$, 95% CI [0.22, 1.80], $t(111) = 2.53$, $p = 0.013$); while adult-to-infant connectivity was not statistically significantly different between Full and Partial gaze ($\beta = 0.37 \pm 0.44$, 95% CI [−0.48, 1.23], $t(111) = 0.85$, $p = 0.396$; Fig. 6a). Importantly, higher adult-to-infant connectivity was associated with increased learning by infants ($\beta = 0.50 \pm 0.06$, 95% CI [0.39, 0.61], $t(224) = 8.58$, $p < 0.001$; partial $r = 0.49$, $p < 0.001$, 95% CI [0.39, 0.58], see Fig. 6b). By contrast, there was no statistically significant direct effect of speaker gaze on learning (by bootstrapping,

https://doi.org/10.1038/s41467-026-75831-x



$\beta = 0.06 \pm 0.12$, 95% CI [−0.18, 0.30], $p = 0.650$). To ensure the mediation results did not arise from biased selection of the adult-to-infant connectivity index, we repeated the mediation analysis using a different adult-to-infant metric that had been independently identified at the single connection level (see Supplementary Note 7.3). This model again showed a significant mediation effect of adult-to-infant coupling, consistent with the main analysis.

For NSE, all five features previously identified (Fig. 5) as being significant in Full gaze were independently tested as potential mediators of the gaze effect on learning (see Supplementary Note 7). In Fig. 6 we present results for delta C3 NSE since this was the only feature for which NSE was significantly higher in Full as compared to Partial or No gaze ($\beta = 0.40 \pm 0.18$, 95% CI [0.05, 0.75], $t(112) = 2.28$, $p = 0.025$; Fig. 6c). However, this NSE feature was not statistically significantly

Fig. 4 | Identifying neural predictors of infant learning and language development. a, **b** Results of partial least squares (PLS) regression conducted for (**a**) learning and (**b**) language development (CDI gesture scores, CDI-G) using within-infant connectivity (within-infant GPDC, orange line) or adult-to-infant connectivity (adult-to-infant GPDC, blue line), respectively. Only adult-to-infant GPDC statistically significantly explained infants' learning (surrogate $p = 0.031$), whereas within-infant GPDC predicted language development scores (surrogate $p = 0.003$) but did not statistically significantly explain learning (surrogate $p = 0.751$). Significance is determined by comparison with a surrogate distribution (grey area: 95% CI, black line: mean). ("*" indicates $p < 0.05$, one-tailed surrogate permutation tests). **c**, **d** Absolute PLS loadings for the first component of the (**c**) adult-to-infant GPDC prediction model for learning and (**d**) within-infant GPDC prediction model for CDI gesture scores, highlighting key connections that load on each component, as estimated by bootstrapping. Respective topographical scalp plots are shown in the right panel, separated by sender and receiver maps. **e**, **f** Cross-validation performance (R^2) of within-infant and adult-to-infant GPDC features for (**e**) learning and (**f**) language development respectively, using 10-fold cross-validation. Violin plot distributions are based on 1000 bootstrapping samples (individual data points omitted for clarity). The three horizontal lines within each violin represent the 25th percentile, median, and 75th percentile. Two-tailed two-sample *t*-tests on bootstrapped R^2 distributions were used. ("***" indicates $p < 0.001$). N = 42 infants contributed valid EEG data. Source data are provided as a Source Data file.

associated with learning ($\beta = -0.12 \pm 0.10$, 95% CI [−0.31, 0.07], $t(112) = -1.22$, $p = 0.226$; partial $r = -0.11$, 95% CI [−0.28, 0.07], $p = 0.236$ in Fig. 6d). Further, none of the NSE features acted either directly or indirectly to mediate the effect of gaze on learning (e.g., by bootstrapping, $\beta = -0.05 \pm 0.04$, 95% CI [−0.15, 0.02], $p = 0.152$; $\beta = 0.35 \pm 0.20$, 95% CI [−0.05, 0.72], $p = 0.084$ for delta C3). Interestingly, the interaction analysis revealed that gaze had a modulatory effect on the relationship between NSE and speaker-to-listener connectivity (Fig. 6e). Specifically, increased delta C3 NSE was associated with stronger adult-to-infant connectivity, but only in the No gaze condition ($\beta = 0.34 \pm 0.17$, 95% CI [0.01, 0.67], $t(111) = 2.02$, $p = 0.045$). Accordingly, although NSE features were sensitive to gaze, none of them statistically significantly mediated the effect of Full gaze on infant learning.

Finally, within-infant connectivity (within-infant GPDC) did not show statistically significant mediation effects of gaze on learning (by bootstrapping, $\beta = 0.06 \pm 0.25$, 95% CI [−0.38, 0.57], $p = 0.820$), as described in further detail in Supplementary Note 7.

Discussion

For infants, learning typically occurs in social contexts, and adult partners play a crucial role in providing ostensive cues for infants' decision processes of social perception, valuation (of relevance and reliability) and selection for learning. Based on the social gating hypothesis[24], we predicted that in our unidirectional learning paradigm infants would show selective learning, with statistically significant learning for languages accompanied by Full gaze and no statistically significant learning when speaker gaze was occluded. Consistent with this prediction, learning was statistically significant in the Full gaze condition and significantly greater than in the No gaze condition (Supplementary Note 1). Second, we predicted that gaze ostensive effects on learning would be mediated by speaker-to-listener neural connectivity. This prediction was again supported. Specifically, mediation analysis demonstrated that only adult-to-infant neural connectivity significantly mediated the relationship between speaker gaze and learning, with Full gaze associated with higher levels of adult-to-infant connectivity, and higher adult-to-infant connectivity associated with increased learning by infants. To confirm that these mediation effects did not arise from PLS optimisation artefacts, we conducted two independent validations (Supplementary Note 7). This finding replicates and extends earlier work demonstrating that the availability of speaker gaze is associated with increased interpersonal neural connectivity in the context of adult-infant communication[12] and supports the view that speaker-listener neural connectivity should be considered as part of any embodied account of social interaction or social learning.

Entrainment and learning

Crucially, neural-speech entrainment (NSE), which reflects phonological encoding of temporal properties of the speech signal and its neural representation in auditory regions[48], did not statistically significantly predict infant learning, although NSE itself was modulated by gaze. Significant NSE was only observed in the Full gaze condition (specifically, at C3 in the delta band, at F4 and Pz in the theta band, and at C3 and Cz in the alpha band). However, NSE was not a statistically significant predictor of infant learning. Interestingly, NSE was positively correlated with adult-to-infant connectivity in the No gaze condition, which may reflect a strategy to optimise sensory information that is available in the absence of ostensive cues, such as acoustic patterns. For example, phonological encoding or rhythmic prediction may be enhanced by NSE[52], but these outcomes were not measured in this study. Similar to findings with early blind individuals[53], this strategy could suggest a compensatory mechanism where the synchronisation of available sensory inputs along with brain functional connectivity is enhanced when visual information is limited.

Adult-to-infant and within-infant connectivity

When comparing adult-to-infant and within-infant connectivity, we noted that infants' intrapersonal connectivity predicted their early communicative gestures scores but did not statistically significantly relate to stimulus selection for learning in this task. Whether within-infant connectivity metrics are stronger indicators of developmental maturity than of transient learning dynamics remains to be tested. This is consistent with findings that within-brain functional connectivity patterns in the left temporal lobe during infancy predict language skills in later childhood[54].

Attention and learning

Recall that infants selectively learned the artificial language which was delivered with Full speaker gaze, whereas no statistically significant learning was detected for languages delivered with Partial or No speaker gaze. This ostensive effect of gaze differed from the pattern for infant visual attention in two ways. First, while learning differed across gaze conditions, measures of visual attention showed no statistically significant gaze-condition differences. Second, UK and Singapore cohorts exhibited differences in visual attention, whereas learning performance showed no statistically significant cohort difference. This suggests that whilst attention is required for learning, gaze-mediated learning is significantly related to speaker-to-listener neural coupling in this paradigm. These patterns are consistent with results from previous studies: Findings by Okumura et al.[22] parallel those of the current study as an overt communicative cue (infant-directed speech in their study) enhanced learning by 9-month-old infants, but there were no statistically significant differences in measures of infant visual attention between conditions with and without this ostensive cue. In addition, Kuhl and colleagues have shown that naturalistic language learning (measured by phonetic discrimination) occurs in live interaction settings, but not when similar stimuli are presented via pre-recorded video, despite infants paying equal attention to video and live presentations[23]. Taken together, these results support the view that ostensive cues play a distinct role in infant learning, in addition to attentional salience and orienting effects. Here, we identify speaker-to-listener neural coupling as a potential marker of this ostensive learning effect.

https://doi.org/10.1038/s41467-026-75831-x

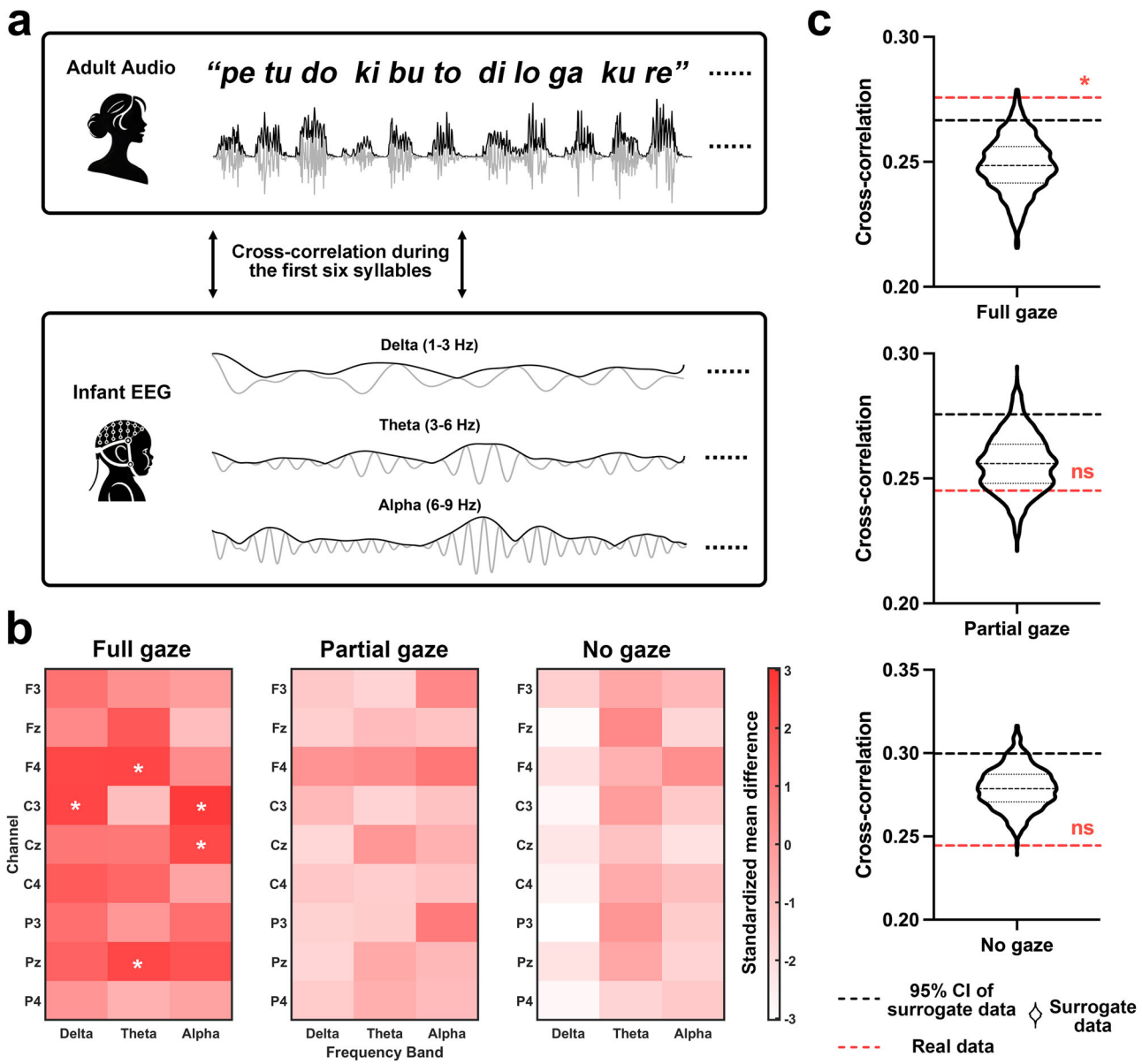


Fig. 5 | Infant neural-speech entrainment (NSE). a Schematic illustrating the measurement of infant neural entrainment to adult speech. Entrainment strength was quantified as the cross-correlation between the Hilbert envelope of adult speech audio waveform and the amplitude envelope of the infant EEG signal in delta, theta and alpha bands during the first six syllables of the Familiarisation phase. **b** Standardised mean difference plots comparing real entrainment strength with surrogate values. Surrogate distributions were derived by randomly shuffling the temporal alignment between infant EEG and the corresponding speech envelope. Significant entrainment was only observed in the Full gaze condition, at C3 in the delta band, at F4 and Pz in the theta band, and at C3 and Cz in the alpha band. "*" indicates $p < 0.05$. *P*-values were corrected by BHFDR for multiple comparisons. Delta C3 ($p_{FDR} = 0.049$), theta F4 ($p_{FDR} = 0.049$), theta Pz ($p_{FDR} = 0.049$), alpha C3 ($p_{FDR} = 0.027$), alpha Cz ($p_{FDR} = 0.049$). **c** Example of surrogate tests for significant entrainment, shown here for the C3 channel in the delta EEG band, for each gaze condition. Only in the Full gaze condition did the real data exceed the 95% CI upper bound of the surrogate data distribution. Violin plot distributions are based on 1000 surrogate samples (individual data points omitted for clarity, one-tailed surrogate permutation tests). The three horizontal lines within each violin represent the 25th percentile, median, and 75th percentile. "ns" indicates not significant. N = 42 infants contributed valid EEG data. Source data are provided as a Source Data file.

Between-cohort effects

Finally, it was of interest to examine between-cohort similarities and differences in ostensive gaze processing between UK and SG cohorts. We did not detect statistically significant differences between infants from the two countries in their response to the gaze manipulation, in either measured learning behaviour or neural connectivity metrics. Responses may nonetheless differ in cultures where parent-child interactions are characterised by physical proximity and touch rather than direct gaze, such as some Arab cultures[55]. Future research with a broader range of cultural contexts is needed to assess the generalisability of these effects. However, we observed a between-cohort difference: Singaporean infants looked significantly longer at the adult on the screen than their UK counterparts during the Familiarisation phase. We propose that this is due to an outgroup effect, as the adult speaker was native English, which may have attracted the attention of the Singaporean infants due to the novelty of less familiar Western facial features. Nonetheless, this novelty effect did not statistically significantly increase learning or modulate the measured speaker-to-listener neural connectivity metrics.

https://doi.org/10.1038/s41467-026-75831-x

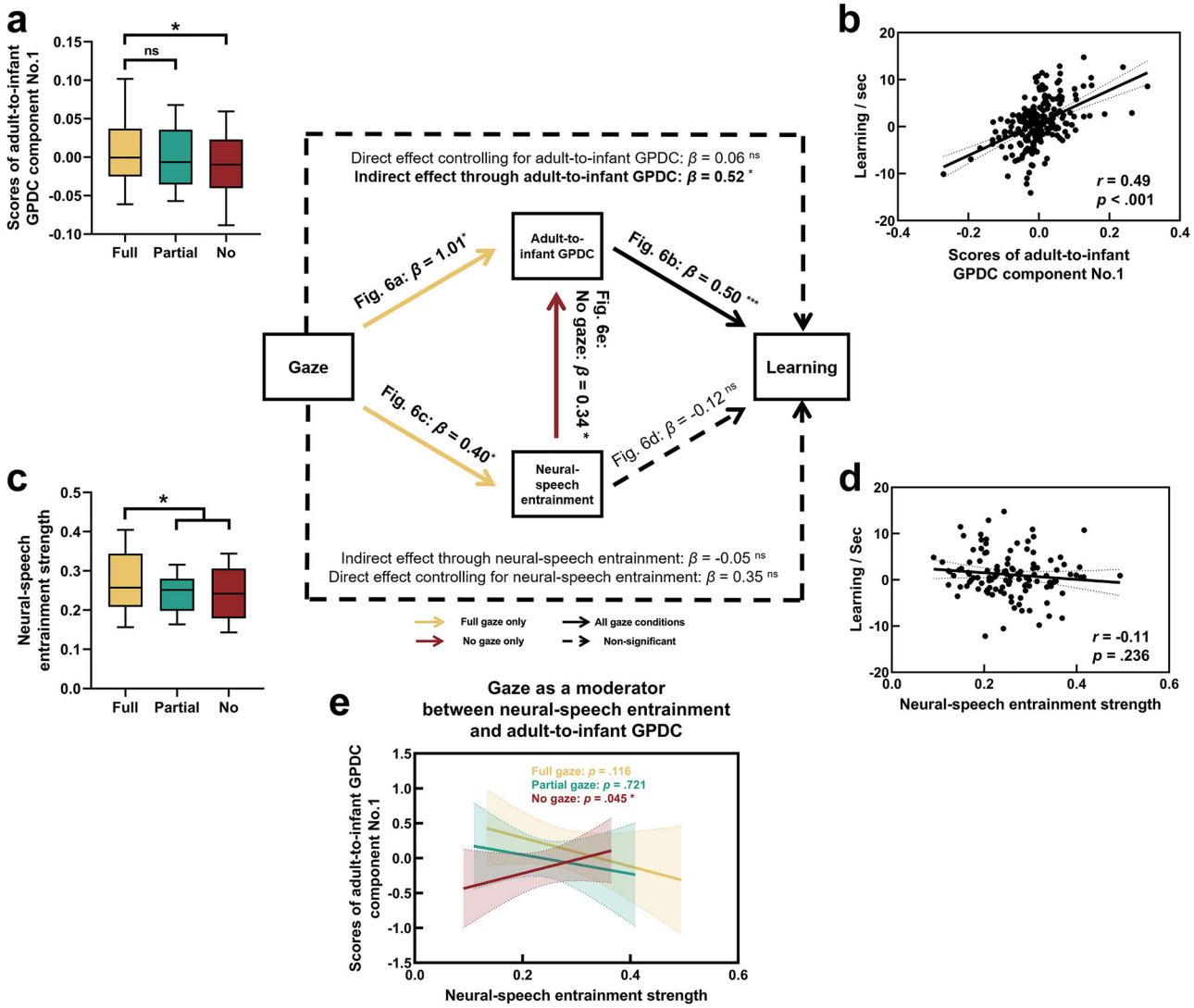


Fig. 6 | Mediation analysis assessing pathways linking ostensive gaze to learning. The main mediation model (middle panel) illustrating direct and indirect pathways by which gaze status might influence learning, including neural-speech entrainment (NSE, represented by delta C3) and adult-to-infant neural connectivity (adult-to-infant GPDC, first PLS component) as potential mediators. Coloured lines represent effects that are significant in one specific gaze condition (yellow for Full gaze, red for No gaze), while black lines indicate effects that are significant across all gaze conditions. Solid arrows represent significant pathways, while dashed arrows indicate statistically non-significant ones, annotated with standardised coefficients and *p*-values. The following subplots illustrate individual pathways revealed in the mediation model. The box represents the median and quartiles, while the whiskers indicate the 10th and 90th percentiles. "ns" indicates not significant, "*" indicates $p < 0.05$ and "***" indicates $p < 0.001$. **a** Box and whisker plots showing effect of Full gaze on adult-to-infant GPDC. Full gaze is associated with significantly higher values of adult-to-infant GPDC Component 1 scores compared to No gaze (LME: $\beta = 1.01$, $p = 0.013$). **b** Correlation between adult-to-infant GPDC and infant learning (LME: $\beta = 0.50$, $p < 0.001$). Shaded band represents 95% CI. **c** Box and whisker plots showing effect of speaker gaze on NSE. Only Full gaze is associated with statistically significantly higher NSE strength (delta C3 entrainment identified in Fig. 5b) compared to Partial and No gaze conditions (LME: $\beta = 0.40$, $p = 0.025$). Similar analyses of other NSE features did not yield significant results, as detailed in Supplementary Note 5. **d** Assessing the direct relationship between NSE (Delta C3) and learning. No statistically significant correlation was observed (LME: $\beta = -0.12$, $p = 0.226$). Shaded band represents the 95% CI. **e** Moderating effect of gaze on the relationship between NSE (delta C3) and adult-to-infant GPDC. Gaze changes the relationship between entrainment and adult-to-infant connectivity, with a significant positive relationship detected only in the No gaze condition ($p = 0.045$). N = 42 infants contributed valid EEG data. Box plots in (**a**, **c**) are based on block-level observations (up to 3 per infant per condition); linear mixed-effects models with participant as random intercept account for repeated measures. Shaded band represents the 95% CI. Source data are provided as a Source Data file.

Limitations

One limitation of the current study is its relatively small sample size (N = 47 for behaviour and N = 42 for EEG) and a related constraint in detecting small effect sizes. Although sensitivity analyses confirmed that our design is adequately sensitive to medium-to-large effects across all procedures (see Supplementary Note 14), replication in larger samples is required. Specifically, while we detected significant learning in the Full gaze condition with a small-to-moderate effect size (Cohen's $d = 0.33$), the Partial gaze condition showed borderline effects (Cohen's $d = 0.26$, $p_{FDR} = 0.058$) that might be more reliably observed with larger samples. Further, the mediation analyses have inherent limitations due to the non-independence of variables entered, thus conclusions should be considered as preliminary, and requiring independent causal validation. Our use of shared frequency bands (6–9 Hz for both adult and infant) is methodologically required for MVAR-based GPDC estimation but may not fully capture the adult alpha frequency range, though key findings replicate across infant-specific theta (3–6 Hz) and alpha (6–9 Hz) bands (Supplementary

Note 13). Further, as the aim of the study was to carefully isolate changes in gaze availability and measure learning, it was necessary to standardise the delivery of the novel languages across infants through the use of pre-recorded videos, leading to a unidirectional design. Thus, we were unable to observe learning in a truly ecological bi-directional interactive context and the extent to which these findings can be generalised to real-world language learning is limited. However, it is interesting to note that although the adult speaker delivered the verbal stimuli in a highly controlled manner, avoiding the communicative cues used in natural conversation such as eye-movements, facial expressions and changes in tone, learning still occurred. Further studies are needed to assess the generalisability of these effects in more naturalistic and interactive settings. In conclusion, the current study provides evidence that speaker-to-listener neural coupling may explain the effect of ostensive gaze in infants' stimulus selection for learning in social contexts. Further, speaker-to-listener neural coupling is a better predictor of selective learning than infants' own neural entrainment activity and attention. This suggests that the adult-to-infant connectivity effectively captures social influences on early cognition, and may provide insights into social valuation decision-processes that are involved in early social learning.

Methods

This research complies with all relevant ethical regulations. The study protocol was approved by the Cambridge Psychology Research Ethics Committee (reference PRE.2016.029) and the Nanyang Technological University Institutional Review Board (reference IRB2019-06-030). Parents provided informed consent for the collection of their own data and on behalf of their infants in accordance with the Declaration of Helsinki. As compensation, UK families received travel and parking reimbursement and a small gift. SG families received cash compensation for their time and a small gift.

Participants

The inclusion criteria were as follows: 1. Infants aged between 8 and 10.5 months. 2. Infants with no diagnosed developmental or neurological problems as assessed by maternal report. 3. Infants raised in English-speaking households, as oscillatory brain activity is thought to be influenced by native language properties[56]. 4. Mothers screened to exclude mental health issues or learning difficulties that could affect their infants' development.

Infant participants were tested at two sites in different countries: the University of Cambridge in the United Kingdom (UK) and Nanyang Technological University, SG. The same experimental protocol and EEG acquisition setup were implemented at both sites. UK infants were from English speaking households, and SG infants were from English-Chinese bilingual families.

Fifty-five infants were recruited, but eight did not contribute data, due to either infant fussiness or accidental software failure. The remaining 47 infants (29 from UK and 18 from SG) had a mean age of 286 days (9.4 months) with a female ratio of 40.4% (19/47) (SG mean age = 294 days, range 251 – 323, 7/18 female; UK mean age = 282, range 245 – 315, 12/29 female). No statistically significant country differences were detected for age ($t(35) = -2.00$, $p = 0.054$, $d = -0.61$, 95% CI [−1.23, 0.02]) or sex ($\chi^2(1) = 0.03$, $p = 0.866$, $\varphi = 0.025$), but age, sex, and country were entered as covariates in subsequent statistical analyses where applicable (see Statistical analyses section for details).

Note that analysis of behavioural results (learning) was based on data from all 47 participants who completed the experiment, however, analyses involving EEG were carried out on a subset of 42 participants who had both usable EEG and behavioural data. Five infants were excluded from EEG analysis due to insufficient valid epochs, and remaining EEG data length did not statistically significantly differ across gaze conditions ($F(2,369) = 0.01$, $p = 0.991$, $\eta^2 < 0.001$, 95% CI [0.00, 0.02]; Supplementary Note 12).

The sample size was estimated to be sufficient based on effect sizes reported in similar studies on infant word learning[57]. A prospective power calculation (for *t*-test analysis) indicated that N = 45 participants would be required to achieve a statistical power of 0.8 at an alpha level of 0.05. A detailed summary of participant demographics, including infant age and sex, maternal age and education, and data exclusion criteria, is provided in Supplementary Note 8.

Study protocol

Experimental design and manipulation of speaker gaze. The study followed a repeated measures design whereby each infant watched videos showing the head of a female native British English speaker presenting three different artificial languages paired with three gaze conditions respectively (Fig. 1a): Full gaze, partially occluded gaze (Partial gaze), and fully occluded gaze (No gaze). The authors affirm that informed consent was provided for the publication of the images in Fig. 1. In each condition the speaker maintained a still head position, looking directly forward (into the camera). She showed no emotional expression and avoided any form of movement beyond speaking. This allowed control for the possible communication of structural information via eye-movements, or facial movements. Any recordings with excessive blinking or facial movements were re-recorded. The three gaze conditions were achieved by placing polarising filters on both the speaker's glasses and the camera, with the speaker wearing the same glasses throughout. By adjusting only the camera-side polarising filter's angle, we controlled eye area visibility in the videos while maintaining consistent recording conditions for the adult speaker (i.e., the speaker's view was not occluded). This ensured that any differences in infant behaviour and brain activity could be attributed solely to differences in infants' perception of the speaker's gaze availability, rather than differences in the speaker's experience whilst recording the stimuli. Pairings of language and gaze condition, and the order of presentation of gaze conditions were counterbalanced across participants. Infants were sequentially assigned to one of three presentation orders: Order 1 (N = 16, Full, Partial, No gaze), Order 2 (N = 16, Partial, No, Full gaze), or Order 3 (N = 15, No, Full, Partial gaze), ensuring each gaze condition was experienced first, second, and third by approximately equal numbers of participants (Fig. 1c). Analysis of order effects revealed no statistically significant effects on (1) infant attention during familiarisation, (2) usable test trials, and (3) learning performance (Supplementary Note 9).

Design of artificial language stimuli. Three different but phonologically equivalent artificial language sets were created so as to ensure that a language that had already been heard was not re-used on the same infant in a different gaze condition. Each language set consisted of 12 unique syllables, each formed by a consonant-vowel combination, and these syllables were organised into four trisyllabic words, following the artificial language structure commonly used in previous statistical learning research[39]. For example, one set might include words like "pe-tu-do", "ki-bu-to", "di-lo-ga", and "ku-re-pa". The syllables for each language were distinct. Syllables were produced at a continuous rate of one syllable every 333 ms (3 Hz), with no intonation, prosody or facial expression, and with shorter recordings concatenated to avoid the speaker taking a breath. Since the stimuli were produced by a human speaker, any naturally-occurring differences in intensity were removed by digital equalisation, with loudness equalised to a playback volume of 61 dB. We further selected recordings that were closely matched for pitch, to ensure no statistically significant differences in mean pitch across gaze conditions.

Infants' learning of the artificial language was assessed using "words" versus "nonwords" from each language set. "Words" were valid words from the language, with 100% transitional probability between syllables (e.g., "di-lo-ga"; hyphens indicate within-word boundaries). The "nonwords" were made of three-syllable strings

spanning a word boundary. As shown in Fig. 1b, these nonword syllable strings had a one-third likelihood of occurring together, as they only occurred when one particular word followed another (e.g., "ga ku-re" occurred when "di-lo-ga" was randomly followed by "ku-re-pa"). The syllable sets, video recordings, and stimulus quality control procedures are detailed in Supplementary Note 9. We also provide four Supplementary Movies to illustrate our experimental stimuli. The first movie demonstrates how we implemented the three different gaze conditions with three different artificial languages during the Familiarisation phase. The other three movies show complete experimental procedures for three different presentation orders (see Supplementary Note 9), including a one-block example of the Familiarisation and Testing phases.

Task structure. The experiment was designed with three identical repeated blocks. Each block consisted of three gaze conditions (Full gaze, Partial gaze, and No gaze), with each condition including a Familiarisation phase followed by a Testing phase. Blocks were constructed this way so that even an infant who only completed one block contributed data to all three gaze conditions, and to minimise unbalanced effects of fatigue or non-completion. Each block lasted for approximately ten minutes. Prior to each Familiarisation and Testing phase, brief cartoon clips were played until the infant's attention was captured, at which point the experimenter manually initiated the phase. Further details of the experimental block structure are provided in Supplementary Note 10.

Familiarisation phase. The Familiarisation phase exposed infants to the artificial language, allowing them to calculate transitional probabilities for defining word boundaries. In each gaze condition within a block, infants watched a 60-sec video of the female adult speaker delivering the artificial language. To maximise infant attention, the 60-sec exposure was divided into three 20-second segments with short (1–3 secs) cartoon clips presented between segments to attract the infant's attention. During each 60 s video, 180 syllables in total were presented at a rate of 3 syllables per second without inter-word pauses, with each of the four words occurring 15 times and each word-pair transition occurring five times. For example, the speech might sound like: "pe-tu-do ki-bu-to di-lo-ga ku-re-pa di-lo-ga pe-tu-do ……" (the hyphen indicates the three syllables within the same word, see Fig. 1b). The infants' gaze at the adult speaker videos was monitored, and visual attention to the screen was measured. Additionally, infant EEG was recorded throughout.

Testing phase. To assess word learning, infants viewed the same speaker producing isolated words or nonwords. Learning was assessed by comparing the looking times towards the screen for words and nonwords per condition. The premise for using looking times is that, provided sufficient familiarisation with the language, the infant will show a novelty preference and look for longer at the less familiar nonwords[39,45]. Thus, it was expected that the words, whose syllables co-occurred more frequently (i.e., had higher transitional probabilities), would be more familiar and thus yield shorter looking times than nonwords.

During the Testing phase, videos of the adult speaker appeared pseudo-randomly on either side of the screen, with a single test item (a word or nonword from the language used in the preceding Familiarisation phase) repeated 12 times, with 900 ms of silence accompanied by a black screen between each repetition. In total, two words and two nonwords were presented in a single Testing phase in pseudo-random order. Infant looking times were recorded based on their head-turning toward the side of the screen and analysed using both automatic video coding (in-house Matlab scripts) and manual checking. Further details of Testing phase implementation are provided in the Supplementary Note 11.

Measures of language development. Mothers assessed their child's receptive and spoken language via self-report using the MacArthur-Bates Communicative Development Inventories (CDI)[58]. Here, we focused on scores for the use of nonverbal communicative actions and gestures (rather than words or phrases), as these were most representative of the language development stage for the young infants participating in the study, who had limited verbal skills.

Data acquisition, preprocessing and analyses

EEG acquisition and preprocessing. EEG was recorded separately from infants (during testing) and from the female adult experimenter (during recording of the video stimulus prior to infant testing) using a BIOPAC Mobita mobile amplifier with a 32-channel Easycap (Brain Products) electrode system following the international 10–20 placement. Data were acquired using Acqknowledge 5.0 software, at a 500 Hz sampling rate. The ground electrode was affixed to the back of the neck, as this location is the least invasive. Channels were online referenced to the average.

Since the adult data were pre-recorded, we only used recordings where all the EEG data were of a high quality and fully usable. During the concurrent recording of adult video and EEG data, we employed active monitoring for eye blinks. Segments with excessive eye blinks were re-recorded. Supplementary Table S12 provides an analysis of the number of blinks remaining in the video stimuli for each language set and gaze condition. The corresponding EEG blink artefacts were removed from the adult data. For infant recordings, bad EEG channels were identified by visual inspection and removed from further analyses. No interpolation was performed and missing data was marked for exclusion in the subsequent GPDC connectivity analyses. Following this, both infant and adult EEG data were average re-referenced, downsampled to 200 Hz and bandpass filtered in the range of 0.1 Hz −45 Hz. Filtering was applied using an inverse FFT filter via the eegfiltfft function in EEGLAB toolbox[59] with default settings. All preprocessing steps were conducted on the full Familiarisation data EEG registration.

Artefact rejection was performed on infant EEG data using both manual and automated steps, and adult EEG data corresponding to the same time periods were also excluded for each participant. For manual rejection, we video coded infants' behaviour at the time precision of individual frames (25 fps, 40 ms per frame) to identify time periods during the experiment when the infant was moving or inattentive (not looking at the screen). These data were excluded entirely from the analysis. An average of 36.4% of attended data were retained per infant after this first stage of manual rejection. Next, each attended data segment was subjected to an automatic artefact rejection algorithm. For this purpose, the attended segments were divided into 1-sec (200 samples) epochs, and baseline correction (mean removal) and linear detrending were applied to each epoch. Epochs with minimum or maximum values outside the predetermined boundaries of −150 and +150 μV were automatically rejected to remove artefacts like ocular movements and eye blinks. Finally, all retained epochs were visually inspected for additional artefacts that were not detected by the preceding steps. In totality, this resulted in an average of 32.9% of data per participant retained for analysis, see also Supplementary Note 12, which provides details of data retained at each pre-processing stage. Critically, artefact rejection rates did not statistically significantly differ across the three gaze conditions in both country datasets (UK: $F(2,78) = 0.20$, $p = 0.817$, $\eta^2 = 0.005$, 95% CI [0.00, 0.06]; Singapore: $F(2,42) = 1.22$, $p = 0.303$, $\eta^2 = 0.05$, 95% CI [0.00, 0.21]).

Measuring neural connectivity using generalised partial directed coherence. Partial directed coherence (PDC)[60] is a validated method for assessing brain connectivity in adult-infant dyads[12]. This technique provides a direct, directed measure of information flow between network nodes, estimated using a Granger Causal[61] framework. Generalised partial directed coherence (GPDC) quantifies the directed flow of

information from channel *j* (the 'Sender') to channel *i* (the 'Receiver') relative to the total predictive contribution of *j* across all channels in the network. Here, we used GPDC[62] calculated using the eMVAR (Extended Multivariate Autoregressive Modelling) Toolbox[63] in Matlab. GPDC is an adapted version of PDC that uses weighted averaging across receivers, providing better variance stabilisation and scale invariance[64]. GPDC was estimated from MVAR models with order determined via the Bayesian Information Criterion (BIC) across candidate orders 2-15. Order 7 minimised BIC for both alpha and theta bands. Model adequacy was confirmed through two diagnostics: (1) variance explained: 58.0% for infant channels and 52.3% for adult channels, with 88.1% of subjects exceeding the 30% threshold for adequate fit[65]; (2) stability[66]: all models showed stable dynamics (maximum eigenvalue = 0.9957 < 1.0). The detailed implementation and robustness analysis of parameter selection for the GPDC analysis are provided in Supplementary Note 13.

We chose to focus our GPDC analysis on 9 channels, comprising F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4 (representing bilateral frontal, central, and parietal regions). This smaller subset of channels was used as the GPDC metric cannot be applied to data with missing nodes (e.g., rejected channels), and our previous analyses suggested that interpolation techniques are unsatisfactory as they may introduce bias. Accordingly, we chose a minimal 9-channel array, representing a balanced spatial topography, for which sufficient usable data (after pre-processing for artefacts) was available that has been previously used in similar infant/children EEG research[67,68]. This vertex-focused topography shows fewer ocular artefacts[69], and also shows lower contamination by speech artefacts[70], which were of potential concern since the adult was producing speech during the EEG recording.

GPDC estimates were obtained for each 1.5 s EEG epoch with 50% overlap for each channel. This balances temporal resolution with frequency stability, capturing around ten cycles of alpha oscillation per window. The windowing procedure and GPDC calculation were applied only on the continuous cleaned EEG data, with no concatenation of separate segments of clean data (i.e., any windows that contained previously identified artefacts or windows at the end of the recording that were less than 1.5 s were excluded from analysis). Finally, the mean GPDC value for each participant pair was calculated by averaging across windows, separately for each gaze condition and block.

This study focused on GPDC within the infant alpha band, defined as the range from 6–9 Hz, as this range has been identified to be functionally equivalent in infants to the adult alpha rhythm[71]. Focusing on a common 6–9 Hz band (used for both adult and infant) is informed by previous adult-infant dyad EEG research, namely: 1. Current practice in adult-infant brain connectivity research supports the use of infant alpha frequency range when calculating connectivity strength (see reviews[72,73]). 2. In a previous study[12], adult-infant GPDC strength within the 6–9 Hz range was found to be influenced by adult speaker gaze. 3. Evidence that adult-infant neural synchrony occurs within overlapping alpha frequencies during social interaction[12,74]. 4. The alpha band is less affected by facial myogenic and speech artefacts and exhibits high test-retest reliability in infants[75]. 5. The MVAR-based GPDC framework evaluates directional connectivity using a unified frequency parameter across all channel pairs[63]. Therefore, cross-frequency analysis (e.g., infant 6–9 Hz, adult 8–12 Hz) falls outside this methodological framework. To verify robustness across frequency bands, we replicated key analyses for theta (3–6 Hz) and alpha (6–9 Hz) in Supplementary Notes 3 and 13. Similar significance across theta and alpha bands demonstrates that findings are not specific to a single narrow frequency range and validates the shared-band approach for GPDC estimation.

Measuring neural-speech entrainment (NSE) using cross-correlation. Cross-correlation analysis is often used to capture entrainment strength between EEG activity and auditory speech stimuli[76,77], assessing temporal dependencies between EEG signals and speech envelopes.

The Hilbert transform was used to compute the speech amplitude envelope, downsampled to match the EEG sampling rate for alignment. EEG amplitude envelope in the delta (1–3 Hz), theta (3–6 Hz), and alpha (6–9 Hz) bands was likewise computed using the Hilbert transform. Cleaned data of EEG segments corresponding to the first six syllables of each phrase were selected to capture neural entrainment, separately for each gaze condition and block, as this early phase typically exhibits strong neural oscillatory entrainment to the speech envelope[78,79].

To account for potential non-linear relationships, both EEG amplitude envelope and the speech envelope were rank-transformed for a Spearman correlation. Cross-correlation analysis (via Matlab's xcov function with normalised coefficients) was applied to calculate the strength of entrainment. For each six-syllable window, we computed cross-correlations within a frequency-band-specific lag (lag range = ±sampling rate / upper bound of the band frequency, e.g., lag range for alpha band = ±200 / 9 ≈ ± 22 time points). The absolute peak correlation value within this window was taken as the measure of entrainment strength.

Generation of surrogate data for significance testing. To assess whether the computed neural GPDC and NSE values were spurious, we performed significance testing against a surrogate dataset for both indices, corrected for multiple comparisons across different channels and frequency bands. The surrogate dataset was generated by randomly disrupting temporal dependencies between signals, providing a chance-level benchmark, a procedure commonly used in previous studies[36,76,80,81].

GPDC. For each participant and each channel, the infant and adults' EEG epochs were randomly shuffled in time, thus preserving spectral content but disrupting global temporal alignment. The same GPDC computation pipeline was then applied to the shuffled data. This process was repeated 1000 times per participant and then averaged to obtain a group level surrogate distribution of GPDC values. The 95% CI upper bound of this surrogate distribution was taken as the threshold value (equivalent to $p < 0.05$) for determining significance for each connection. Correction for multiple comparisons was performed using the Benjamini-Hochberg False Discovery Rate (BHFDR) procedure[82].

NSE. For each participant and channel, surrogates were created by replacing the original speech segments with random segments of equal length drawn from a different language set, and then recombined with the infants' EEG. The same NSE computation pipeline was then applied to this recombined data. A total of 1000 surrogate datasets were generated at the group level, and the resulting group mean values were used to establish a 95% confidence interval for each EEG channel and frequency band. The observed NSE data means were then compared to this confidence interval, with those exceeding the upper bound considered significant (equivalent to $p < 0.05$). Correction for multiple comparisons was performed using the BHFDR procedure.

Statistical analyses. Learning was defined as the looking time difference between nonwords and words, following established infant statistical learning methodology[39,46,47]. All statistical models controlled for infant age, sex, and country as fixed-effect covariates. When applicable, subject was regarded as a random intercept to account for repeated measures. To assess our primary hypothesis of infant learning selectivity, separate paired t-tests were used for each gaze condition to assess subject-level learning performance against chance, corrected for multiple comparisons using the Benjamini-Hochberg FDR procedure (p_{FDR}). Directional (one-tailed) tests were used given

the established protocol that longer looking times for nonword compared to word stimuli indicate a learning effect[39,46,47]. Surrogate-based permutation tests for GPDC connectivity, PLS prediction, and neural-speech entrainment were also one-tailed (real > chance-level distribution). All remaining tests were two-sided, and multiple comparisons were corrected using the Benjamini-Hochberg FDR procedure. Supplementary Note 1 provides alternative complementary analyses of relative learning effects across gaze conditions.

To assess relationships between behavioural and EEG variables, statistical analyses employed linear mixed-effects (LME) models for measures with repeated observations (learning, visual attention, EEG connectivity, and NSE), with age, sex, and country as covariates and participant ID as a random intercept. LME models were fitted using the fitlme function in Matlab R2024b (Statistics and Machine Learning Toolbox) with maximum likelihood (ML) estimation.

Welch t-tests were used for between-country comparisons of single-measurement variables (e.g., CDI scores). Partial Pearson correlations with covariates quantified potential attention-learning and CDI-learning relationships.

It should be noted that degrees of freedom vary across analyses reflecting different observation units and potential missing data: subject-level testing (df ≈ 35–46), behavioural analyses with subjects nested within 3 conditions (df ≈ 111–137), and EEG-behavioural analyses with subjects × conditions × 3 repeated blocks (df ≈ 221–369). We have listed all df settings in Supplementary Note 8 (Table S5). Power and sensitivity analyses[83] are reported in Supplementary Note 14. Effect sizes were calculated from raw descriptive statistics with 95% confidence intervals based on non-central distributions. Normality was assessed using Shapiro-Wilk tests on the key outcome variables. Learning scores were normally distributed in all three gaze conditions (all $W > 0.95$, $p > 0.080$). Levene's tests confirmed homogeneity of variances across conditions (all $p > 0.480$). Where normality was not met (attention duration, GPDC connectivity scores), analyses employed linear mixed-effects models and non-parametric surrogate permutation tests, both of which are robust to violations of normality.

Prediction analyses with partial least squares regression

Partial least squares (PLS) regression was used to determine whether specific neural connectivity (e.g., adult-to-infant GPDC and within-infant GPDC) or NSE features could predict learning outcomes or CDI gesture scores while controlling for covariates such as age, sex, and country. PLS was chosen because it combines principal component analysis (PCA)-like dimensionality reduction with linear regression, identifying components in the predictor variables (e.g., neural measures) that best maximise covariance with the response variables (e.g., learning performance), with each component represented by loadings that indicate the linear contribution of each original variable to the predictive pattern. This method is particularly suitable for handling features prone to high collinearity, such as GPDC from multiple EEG channels, while providing interpretable components that linearly correspond to behavioural outcomes[84]. For adult-to-infant GPDC analysis, we included 80 connections (out of 81 possible adult-to-infant connections) that passed FDR-corrected surrogate testing. For within-infant GPDC analysis, we included 64 connections (out of 72 within-infant connections, excluding self-connections) meeting the same significance threshold. Age, sex, and country were included as covariates in all models. PLS components are ordered by the amount of variance they explain. For interpretability, reporting is focused on the first component as this captures the most covariance between predictors and outcomes. To evaluate the reliability and generalisability of the PLS models, we performed three validation analyses (based on prior literature[85,86]):

1. Goodness-of-fit evaluation: Predictive performance (R^2 between real and predicted learning outcome values) was compared between real and shuffled surrogate datasets for each modality to identify significant predictors. Real prediction performance (R^2) was assessed against the upper 95th percentile of the surrogate distribution to determine significance (equivalent to one-tailed $p < 0.05$).

2. Component loading estimation: Bootstrapping (1000 iterations) was applied to estimate the loadings of key components within the PLS model, providing robust feature evaluations suited to small sample sizes[87]. Specifically, we used the 1000 surrogate datasets (GPDC or NSE) generated previously to perform identical surrogate PLS regression analyses.

3. Cross-validation: To assess generalisability and account for inter-individual variability, we employed nested bootstrap cross-validation[88]: each of 1000 bootstrap iterations randomly resampled the dataset with replacement, then applied 10-fold cross-validation to the resampled data. Performance metrics (correlation between predicted and observed outcomes) were averaged across folds within each iteration, yielding a distribution of 1000 cross-validated estimates.

Mediation analyses

In the last section of the results, mediation analysis[89] was applied to identify and quantify the relative strengths of direct and indirect pathways through which gaze status could influence learning outcomes. Specifically, we wished to test whether the effects of gaze on infant learning were indirectly mediated through neural features such as adult-to-infant GPDC or NSE. Mediation analysis was implemented using linear mixed-effects (LME) models instead of general linear models, as LME models provide a suitable framework for analyzing pathways in mediated relationships while accounting for the correlation structure of repeated measurements[90]. To quantify and evaluate the significance of each pathway within the mediation model, we performed two analysis steps: (1) LME modelling to estimate the standardised coefficient value and significance, and (2) nonparametric bootstrapping with 1000 resamples to estimate the effect size distributions and confidence intervals of the direct and indirect effects[91]. The direct effect was calculated as the coefficient for gaze when both gaze and the mediator were included as predictors of learning (thus controlling for the mediation effect). The indirect effect was calculated as the product of: (a) the coefficient of gaze predicting the mediator and (b) the coefficient of the mediator predicting learning while controlling for gaze.

The adult-to-infant neural connectivity feature entered into the model corresponded to the loadings of the first component of adult-to-infant GPDC, identified through the previous PLS analysis reported in the second section of results. We acknowledge that using PLS-derived components (optimised for learning prediction) as mediators introduces analytical dependencies, precluding confirmatory causal inference. Therefore, we position these analyses as preliminary and pending independent causal validation. Two additional tests using an independently-derived measure of adult-to-infant neural coupling and substituting within-infant connectivity in the mediation analyses are provided in Supplementary Note 7.3 and 7.2, respectively. Similarly, we identified five possible NSE features (channel C3 for delta band, Pz and F4 for theta and C3 and Cz for alpha band) based on the NSE results section. These were tested in separate mediation models (shown in Supplementary Note 7), and the delta C3 model was reported in the main manuscript as this was the only one with a significant moderation effect.

Finally, to assess whether within-infant connectivity also mediated gaze effects on learning, we replaced adult-to-infant GPDC with its within-infant GPDC analogue (the loadings of the first PLS component) in a separate mediation model reported in Supplementary Note 7.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data are provided with this paper. The raw and pre-processed EEG, behavioural data, and video stimuli that support the findings of this study are openly available in DR-NTU (Data) - Nanyang Technological University's research data repository at https://researchdata.ntu.edu.sg. The specific dataset access links: Raw EEG data: https://doi.org/10.21979/N9/BQLIB9. Preprocessed EEG data: https://doi.org/10.21979/N9/F9N5BE. Behavioural data: https://doi.org/10.21979/N9/4EBTKT. Video stimuli: https://doi.org/10.21979/N9/NJ1KJA Source data are provided with this paper.

Code availability

The analysis code used in this study is openly available on GitHub at https://github.com/Baby-Linc-Singapore/BABBLE_CODE and archived on Zenodo[92]: https://doi.org/10.5281/zenodo.19203884.

References

1. Csibra, G. & Gergely, G. Natural pedagogy. *Trends Cogn. Sci.* **13**, 148–153 (2009).
2. Kovács, Á. M., Téglás, E. & Endress, A. D. The social sense: susceptibility to others' beliefs in human infants and adults. *Science* **330**, 1830–1834 (2010).
3. Senju, A. & Csibra, G. Gaze following in human infants depends on communicative signals. *Curr. Biol.* **18**, 668–671 (2008).
4. Parise, E. & Csibra, G. Neural responses to multimodal ostensive signals in 5-month-old infants. *PloS One* **8**, e72360 (2013).
5. Gangopadhyay, P., Chawla, M., Dal Monte, O. & Chang, S. W. Prefrontal–amygdala circuits in social decision-making. *Nat. Neurosci.* **24**, 5–18 (2021).
6. Tremblay, S., Sharika, K. M. & Platt, M. L. Social decision-making and the brain: a comparative perspective. *Trends Cogn. Sci.* **21**, 265–276 (2017).
7. Suzuki, S. & O'Doherty, J. P. Breaking human social decision making into multiple components and then putting them together again. *Cortex* **127**, 221–230 (2020).
8. Amadei, E. A. et al. Dynamic corticostriatal activity biases social bonding in monogamous female prairie voles. *Nature* **546**, 297–301 (2017).
9. Walum, H. & Young, L. J. The neural mechanisms and circuitry of the pair bond. *Nat. Rev. Neurosci.* **19**, 643–654 (2018).
10. Çetinçelik, M., Rowland, C. F. & Snijders, T. M. Do the eyes have it? a systematic review on the role of eye gaze in infant language development. *Front. Psychol.* **11**, 589096 (2021).
11. Bloom, K. Social elicitation of infant vocal behavior. *J. Exp. Child Psychol.* **20**, 51–58 (1975).
12. Leong, V. et al. Speaker gaze increases information coupling between infant and adult brains. *Proc. Natl. Acad. Sci. USA.* **114**, 13290–13295 (2017).
13. Farroni, T., Csibra, G., Simion, F. & Johnson, M. H. Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. USA.* **99**, 9602–9605 (2002).
14. Farroni, T., Massaccesi, S., Menon, E. & Johnson, M. H. Direct gaze modulates face recognition in young infants. *Cognition* **102**, 396–404 (2007).
15. Hoehl, S. et al. Looking at eye gaze processing and its neural correlates in infancy—implications for social development and autism spectrum disorder. *Child Dev.* **80**, 968–985 (2009).
16. Kliesch, C., Parise, E., Reid, V. & Hoehl, S. The role of social signals in segmenting observed actions in 18-month-old children. *Dev. Sci.* **25**, e13198 (2022).
17. Degutyte, Z. & Astell, A. The role of eye gaze in regulating turn taking in conversations: a systematized review of methods and findings. *Front Psychol.* **12**, 616471 (2021).
18. Holler, J. Facial clues to conversational intentions. *Trends Cogn. Sci.* **29**, 750–762 (2025).
19. Hömke, P., Holler, J. & Levinson, S. C. Eye blinking as addressee feedback in face-to-face conversation. *Res. Lang. Soc. Interact.* **50**, 54–70 (2017).
20. Hömke, P., Levinson, S. C., Emmendorfer, A. K. & Holler, J. Eyebrow movements as signals of communicative problems in human face-to-face interaction. *R. Soc. Open Sci.* **12**, 241632 (2025).
21. Szufnarowska, J., Rohlfing, K. J., Fawcett, C. & Gredebäck, G. Is ostension any more than attention?. *Sci. Rep.* **4**, 5304 (2014).
22. Okumura, Y., Kanakogi, Y., Kobayashi, T. & Itakura, S. Ostension affects infant learning more than attention. *Cognition* **195**, 104082 (2020).
23. Kuhl, P. K., Tsao, F.-M. & Liu, H.-M. Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. USA.* **100**, 9096–9101 (2003).
24. Kuhl, P. K. Is speech learning 'gated' by the social brain?. *Dev. Sci.* **10**, 110–120 (2007).
25. Wass, S. V., Whitehorn, M., Haresign, I. M., Phillips, E. & Leong, V. Interpersonal neural entrainment during early social interaction. *Trends Cogn. Sci.* **24**, 329–342 (2020).
26. Reindl, V., Konrad, K., Poon, K. K. & Leong, V. Classroom-based learning dynamics: the role of interbrain synchrony. *Trends Cogn. Sci.* **28**, 1063–1065 (2024).
27. Kliesch, C. Postnatal dependency as the foundation of social learning in humans. *Proc. R. Soc. B.* **292**, 20242818 (2025).
28. De Jaegher, H. & Di Paolo, E. Participatory sense-making: an enactive approach to social cognition. *Phenom. Cogn. Sci.* **6**, 485–507 (2007).
29. Pickering, M. J. & Garrod, S. An integrated theory of language production and comprehension. *Behav. Brain Sci.* **36**, 329–347 (2013).
30. Schwartz, L. et al. Technologically-assisted communication attenuates inter-brain synchrony. *Neuroimage* **264**, 119677 (2022).
31. Piazza, E. A., Hasenfratz, L., Hasson, U. & Lew-Williams, C. Infant and adult brains are coupled to the dynamics of natural communication. *Psychol. Sci.* **31**, 6–17 (2020).
32. Grossmann, T. et al. Early cortical specialization for face-to-face communication in human infants. *Proc. R. Soc. B.* **275**, 2803–2811 (2008).
33. Urakawa, S., Takamoto, K., Ishikawa, A., Ono, T. & Nishijo, H. Selective medial prefrontal cortex responses during live mutual gaze interactions in human infants: an fNIRS study. *Brain Topogr.* **28**, 691–701 (2015).
34. Nguyen, T. et al. The effects of interaction quality on neural synchrony during mother-child problem solving. *Cortex* **124**, 235–249 (2020).
35. Hoehl, S., Fairhurst, M. & Schirmer, A. Interactional synchrony: signals, mechanisms and benefits. *Soc. Cogn. Affect. Neurosci.* **16**, 5–18 (2021).
36. Pérez, A., Carreiras, M. & Duñabeitia, J. A. Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Sci. Rep.* **7**, 4190 (2017).
37. Nozaradan, S., Peretz, I. & Mouraux, A. Steady-state evoked potentials as an index of multisensory temporal binding. *NeuroImage* **60**, 21–28 (2012).
38. Novembre, G. & Iannetti, G. D. Hyperscanning alone cannot prove causality. multibrain stimulation can. *Trends Cogn. Sci.* **25**, 96–99 (2021).
39. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).

40. Teinonen, T., Fellman, V., Näätänen, R., Alku, P. & Huotilainen, M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* **10**, 21 (2009).
41. Akechi, H. et al. Attention to eye contact in the West and East: autonomic responses and evaluative ratings. *PloS One* **8**, e59312 (2013).
42. Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G. & Caldara, R. Cultural confusions show that facial expressions are not universal. *Curr. Biol.* **19**, 1543–1548 (2009).
43. Gobel, M. S., Chen, A. & Richardson, D. C. How different cultures look at faces depends on the interpersonal context. *Can. J. exp. Psychol.* **71**, 258 (2017).
44. Haensel, J. X., Smith, T. J. & Senju, A. Cultural differences in mutual gaze during face-to-face interactions: A dual head-mounted eye-tracking study. *Vis. Cogn.* **30**, 100–115 (2022).
45. Fantz, R. L. Visual experience in infants: decreased attention to familiar patterns relative to novel ones. *Science* **146**, 668–670 (1964).
46. Pelucchi, B., Hay, J. F. & Saffran, J. R. Statistical learning in a natural language by 8-month-old infants. *Child Dev.* **80**, 674–685 (2009).
47. Benitez, V. L., Bulgarelli, F., Byers-Heinlein, K., Saffran, J. R. & Weiss, D. J. Statistical learning of multiple speech streams: A challenge for monolingual infants. *Dev. Sci.* **23**, e12896 (2020).
48. Obleser, J. & Kayser, C. Neural entrainment and attentional selection in the listening brain. *Trends Cogn. Sci.* **23**, 913–926 (2019).
49. Kösem, A. et al. Neural entrainment determines the words we hear. *Curr. Biol.* **28**, 2867–2875 (2018).
50. Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S. & Keysers, C. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* **16**, 114–121 (2012).
51. Wilson, M. & Wilson, T. P. An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* **12**, 957–968 (2005).
52. Cason, N. & Schön, D. Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* **50**, 2652–2658 (2012).
53. Van Ackeren, M. J., Barbero, F. M., Mattioni, S., Bottini, R. & Collignon, O. Neuronal populations in the occipital cortex of the blind synchronize to the temporal dynamics of speech. *eLife* **7**, e31640 (2018).
54. Yu, X. et al. Functional connectivity in infancy and toddlerhood predicts long-term language and preliteracy outcomes. *Cereb. Cortex* **32**, 725–736 (2022).
55. Feldman, R., Masalha, S. & Alony, D. Microregulatory patterns of family interactions: cultural pathways to toddlers' self-regulation. *J. Fam. Psychol.* **20**, 614 (2006).
56. Pérez, A., Carreiras, M., Dowens, M. G. & Duñabeitia, J. A. Differential oscillatory encoding of foreign speech. *Brain Lang.* **147**, 51–57 (2015).
57. Choi, D., Batterink, L. J., Black, A. K., Paller, K. A. & Werker, J. F. Preverbal infants discover statistical word patterns at similar rates as adults: evidence from neural entrainment. *Psychol. Sci.* **31**, 1161–1173 (2020).
58. Fenson, L. et al. *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (Paul H. Brookes Publishing Co., Baltimore, 2007).
59. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
60. Baccalá, L. A. & Sameshima, K. Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* **84**, 463–474 (2001).
61. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: J. Econom. Soc.* **37**, 424–438 (1969).
62. Baccalá, L. A., Sameshima, K. & Takahashi, D. Y. Generalized Partial Directed Coherence. In *2007 15th International Conference on Digital Signal Processing* 163–166 (IEEE, Cardiff, 2007).
63. Faes, L., Erla, S., Porta, A. & Nollo, G. A framework for assessing frequency domain causality in physiological time series with instantaneous effects. *Philos. Trans. R. Soc. A.* **371**, 20110618 (2013).
64. Hoerzer, G. M., Liebe, S., Schloegl, A., Logothetis, N. K. & Rainer, G. Directed coupling in local field potentials of macaque v4 during visual short-term memory revealed by multivariate autoregressive models. *Front. Comput. Neurosci.* **4**, 1214 (2010).
65. Seth, A. K. A MATLAB toolbox for Granger causal connectivity analysis. *J. Neurosci. Methods* **186**, 262–273 (2010).
66. Ding, M., Bressler, S. L., Yang, W. & Liang, H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biol. Cybern.* **83**, 35–45 (2000).
67. Háden, G. P., Mády, K., Török, M. & Winkler, I. Newborn infants differently process adult directed and infant directed speech. *Int. J. Psychophysiol.* **147**, 107–112 (2020).
68. Kuipers, J. R. & Thierry, G. Bilingualism and increased attention to speech: Evidence from event-related potentials. *Brain Lang.* **149**, 27–32 (2015).
69. Romero, S., Mañanas, M. A. & Barbanoj, M. J. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: a simulation case. *Comput. Biol. Med.* **38**, 348–360 (2008).
70. Brooker, B. H. & Donald, M. W. Contribution of the speech musculature to apparent human EEG asymmetries prior to vocalization. *Brain Lang.* **9**, 226–245 (1980).
71. Stroganova, T. A., Orekhova, E. V. & Posikera, I. N. EEG alpha rhythm in infants. *Clin. Neurophysiol.* **110**, 997–1012 (1999).
72. Turk, E., Vroomen, J., Fonken, Y., Levy, J. & Van Den Heuvel, M. I. In sync with your child: The potential of parent–child electroencephalography in developmental research. *Dev. Psychobiol.* **64**, e22221 (2022).
73. Markova, G., Nguyen, T. & Hoehl, S. Neurobehavioral interpersonal synchrony in early development: The role of interactional rhythms. *Front. Psychol.* **10**, 2078 (2019).
74. Santamaria, L. et al. Emotional valence modulates the topology of the parent-infant inter-brain network. *NeuroImage* **207**, 116341 (2020).
75. Georgieva, S. et al. Toward the understanding of topographical and spectral signatures of infant movement artifacts in naturalistic EEG. *Front. Neurosci.* **14**, 352 (2020).
76. Zoefel, B. & VanRullen, R. EEG oscillations entrain their phase to high-level features of speech sound. *Neuroimage* **124**, 16–23 (2016).
77. Lalor, E. C., Power, A. J., Reilly, R. B. & Foxe, J. J. Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiol.* **102**, 349–359 (2009).
78. Luo, H. & Poeppel, D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* **54**, 1001–1010 (2007).
79. Peelle, J. E., Gross, J. & Davis, M. H. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* **23**, 1378–1387 (2013).
80. Chiarion, G., Sparacino, L., Antonacci, Y., Faes, L. & Mesin, L. Connectivity analysis in EEG data: a tutorial review of the state of the art and emerging trends. *Bioengineering* **10**, 372 (2023).
81. Çetinçelik, M., Rowland, C. F. & Snijders, T. M. Ten-month-old infants' neural tracking of naturalistic speech is not facilitated by the speaker's eye gaze. *Dev. Cogn. Neurosci.* **64**, 101297 (2023).
82. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.: Ser. B (Methodol.)* **57**, 289–300 (1995).

83. Lakens, D. Sample size justification. *Collabra: Psychol.* **8**, 33267 (2022).
84. Krishnan, A., Williams, L. J., McIntosh, A. R. & Abdi, H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56**, 455–475 (2011).
85. Seidlitz, J. et al. Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. *Neuron* **97**, 231–247 (2018).
86. Vértes, P. E. et al. Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks. *Philos. Trans. R. Soc. B: Biol. Sci.* **371**, 20150362 (2016).
87. Fu, W. J., Carroll, R. J. & Wang, S. Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics* **21**, 1979–1986 (2005).
88. Chin, W. W. Bootstrap Cross-Validation Indices for PLS Path Model Assessment. In *Handbook of Partial Least Squares* (eds Esposito Vinzi, V., Chin, W. W., Henseler, J. & Wang, H.) 83–97 (Springer Berlin Heidelberg, 2010).
89. Hayes, A. F. Beyond baron and kenny: statistical mediation analysis in the new millennium. *Commun. Monogr.* **76**, 408–420 (2009).
90. Blood, E. A., Cabral, H., Heeren, T. & Cheng, D. M. Performance of mixed effects models in the analysis of mediated longitudinal data. *BMC Med. Res. Methodol.* **10**, 16 (2010).
91. Preacher, K. J. & Hayes, A. F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* **40**, 879–891 (2008).
92. Zhang, W. Baby-linc-singapore/BABBLE_CODE: adult-infant neural coupling R1.0. *Zenodo* https://doi.org/10.5281/zenodo.19203884 (2026).

Acknowledgements

Special thanks to Eszter and Diarmid Campbell for invaluable practical advice on the use of polarising filters to modulate visibility of the speaker's eyes while still allowing the speaker to read syllable strings from the autocue. Thanks also to Melis Çetinçelik, Marina Wenzl, Winnie Wee, Ivfy Foong, Teo Kai Xin, Dorcas Keow, Yvonne Chia, Jamie Lee, Sim Jia Yi, Lois Timothy, Arya Bhomick, Nastassja Fischer, and Lee Kean Mun for assistance with data collection and analysis. We thank and acknowledge the Cognitive Neuroimaging Centre, Nanyang Technological University, Singapore for computational resources used in data analysis.

Author contributions

V.L. supervised project administration and the study. V.L., K.C., and S.G. designed and conceptualized the experiments. K.C., S.G., N.D., and V.V. collected the data. W.Z., S.G., V.R., and P.S. performed the formal analysis. W.Z. performed data visualisation and developed software/resources. V.L., K.C., and W.Z. drafted the manuscript. All authors (W.Z., K.C., S.G., L.S., V.R., V.N., N.D., V.V., P.S., and V.L.) revised and approved the manuscript.

Funding

This research is supported by RIE2025 Human Potential Programme Prenatal/Early Childhood Grants (H22P0M0002/H24P2M0008), administered by A*STAR. VL is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (MOE-T2EP40121-0001) and Science of Learning grant (MOESOL2021-0001) and by a Social Science & Humanities Research Fellowship (MOE2020-SSHR-008).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-75831-x.

Correspondence and requests for materials should be addressed to Victoria Leong.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.