Endogenous oscillatory rhythms and interactive contingencies jointly influence infant attention during early infant-caregiver interaction

Emily A.M. Phillips(*)(1) , Louise Goupil(1), Megan Whitehorn(1), Emma Bruce-Gardyne(1), Florian A. Csolsim(1), Navsheen Kaur (1), Emily Greenwood(1), Ira Marriott-Haresign(1) & Sam V. Wass(1)

1 - Department of Psychology, University of East London, London, UK, E15 4LZ

2 - LPNC / CNRS, Grenoble Alpes University, Grenoble, FR

***Corresponding author:** Ms Emily Phillips. Address: University of East London, London E15 4LZ. Telephone: +44(0)7885838069

Email: u1920558@uel.ac.uk

Abstract

Almost all early cognitive development takes place in social contexts. At the moment, however, we know little about the neural and cognitive mechanisms that drive infant attention during social interactions. Recording EEG during naturalistic caregiver-infant interactions (N=66), we compare two different accounts. Attentional scaffolding perspectives emphasise the role of the caregiver in structuring the child's behaviour, whilst active learning models focus on motivational factors, endogenous to the infant, that guide their attention. Our results show that, already by 12-months, intrinsic cognitive processes control infants' attention: fluctuations in endogenous oscillatory neural activity associated with changes in infant attentiveness, and predicted the length of infant attention episodes towards objects. In comparison, infant attention was not forwards-predicted by caregiver gaze, or modulations in the spectral and temporal properties of their caregiver's speech. Instead, caregivers rapidly modulated their behaviours in response to changes in infant attention and cognitive engagement, and greater reactive changes associated with longer infant attention. Our findings suggest that shared attention develops through interactive but asymmetric, infant-led processes that operate across the caregiver-child dyad.

Introduction

Almost all early cognitive development and learning takes place in social contexts [1]. We know that caregiver behaviours influence where, how, and for how long children allocate their attention in real-world settings [2], and that individual differences in how caregivers behave while interacting with their child can predict later language learning and socio-cognitive development [3–5]. But we currently understand little about the intrapersonal and bidirectional neural mechanisms that influence how infants allocate their attention to learn from their environment during naturalistic, free-flowing interactions.

A number of different theoretical models try to explain how social partners influence infants' attention. The first, and probably the oldest, proposes that caregivers directly and didactically scaffold their infant's attention, for example by building a structure of how they pay attention, and when, and encouraging the child to follow their attentional focus (a process sometimes known as 'attentional scaffolding' [6]. This might take place through children copying where caregivers are paying attention, second by second, while they complete a shared task [2]. Or, it might happen through adults organisedly and actively using ostensive signalling to guide infant attention [7,8]. To do this, the adult partner might be using salient behaviours (e.g., eye gaze, high pitched speech, etc) to exogenously influence where children allocate their attention. In either of these cases, infant attention is *reactive* to changes in the behaviour of the caregiver [9,10].

Recent micro-behavioural analyses of caregiver and infant gaze behaviour during joint table-top interactions support this perspective, to some extent. Multimodal behavioural inputs by the caregiver are known to support episodes of sustained attention towards objects: for

example, infant attention durations lasting over 3 seconds are directly predicted by the amount and timing of caregiver speech and touch to objects [11]. More indirectly, other research has shown that infant attention is more fast-changing in joint compared to solo play, despite infant attention durations being, overall, longer in joint play [12] - suggesting that endogenous cognitive processes such as attentional inertia (the finding that, the longer a look lasts, the less likely it is to end [13]) have less of an influence on infant attention in social contexts. Further research suggests that, rather than following the focus of the adults' gaze, infants most often co-ordinate their attention with the adult through attending towards their partners' object manipulations, which corresponds to the idea that adults use exogenous attention capture to drive infant attention [10]. Other salient behaviours might also be important, but are under-investigated. For example, infant-directed speech is known to contain more variability in amplitude and pitch [14], which increases its auditory salience [15]; but although it is known that children generally pay more attention to infant-directed speech [15], no previous research has examined whether caregivers use moment-by-moment variability in the salience of their voice to influence how children allocate attention.

Within this framework, it is possible that, rather than repeated and reactive contingent responsivity to isolated behaviours, temporal dependency between infant and caregiver attention is driven by infant behaviour becoming periodically coupled to the behavioural modulations of their partner [16,17]. Similar to inter-dyadic patterns of vocalisations in adults and marmoset monkeys [18,19], in early infant-caregiver interactions, vocal pauses in one partner's vocalisations can be predicted from those of the other [20,21], and, during face-to-face interactions at the end of the first year, caregiver-infant facial affect becomes temporally aligned [22]. Oscillatory entrainment, that is, consistent temporal alignment between fluctuations in caregiver and infant behaviour, could be particularly important in ensuring

that salient sensory and information-rich inputs by the caregiver occur at moments infant are most receptive to receiving information [17].

An alternative interpretation of these micro-behavioural findings, however, is that, rather than structuring infant behaviour through leading infant attention, caregivers instead scaffold how infants pay attention by following and responding to re-orientations in their infant's attention. This second model suggests that, rather than considering unidirectional caregiver->child influences we should instead be considering bidirectional child<->caregiver influences. In following the focus of their infants' attention at moments that they reorient towards a new object, the caregiver 'catches' and extends infant attention with reactive and dynamic change in their salient ostensive behaviours, to which infants are responsive [2]. The contingent adaptation of the caregiver to modulations in infant attention serves to maintain and extend infant attention, and provides inputs at points where infants anticipate to receive new information [23]. Indeed, from early infancy, caregivers are contingently responsive to modulations in their infant's behaviour. From 2-3 months, caregivers respond differentially to distinct facial affects produced by the infant [24], modulate their vocal feedback to infant babbling [25–27]; and, towards the end of the first year, provide more labelling responses relative to infant's pointing than to their object-directed vocalisations [28].

According to the first model, then, caregivers drive and actively control infants' attention during joint interaction. According to the alternative model, caregivers influence infants' attention by reactively and contingently responding to the infant's attention shifts. But according to the latter model, what drives how infants initially allocate their attention in the first place? In caregivers, the timing of attention shifts can be partially described using an oscillatory structure, reflecting rhythmic attention reorientations that possibly correspond to

fluctuations in the central nervous system [29–31]. Research with infants has also suggested that, even during early life, infants' attention shifting is not purely stochastic [32]. In free-viewing paradigms, infant gaze exhibits a fractal structure: becoming more periodic and less stochastic over the course of the first year [33,34], and periodic structure in 12-month old attention patterns has been associated with increased cognitive control [35]. Regulatory mechanisms endogenous to the infant could therefore be one mechanism that influences when infants reorient their attention during real-world naturalistic interactions.

By the end of the first year, however, as well as periodic attention reorientations, fluctuations in top-down attentional control processes, thought to be driven by the executive attention system, begin to influence where and when infants shift their attention. For example, research has shown that infants routinely deploy active and effortful information-sampling strategies to maximise their opportunities for learning [36–40]. For example, infants aged 8-9 months optimise information gain by directing their attention towards stimuli that are neither too complex, nor too predictable [38,41] and disengage from stimuli that are less informative compared to past observations [40]. Corresponding to developments in intentionally-mediated forms of joint communication [42], infants are also thought to begin to use active strategies to directly elicit information from a social partner about their environment. For example, infants aged 12-14 months point in an interrogative manner [43,44], and look towards their caregiver to ask for help when uncertain [45,46].

These approaches suggest that infants' endogenous engagement or interest forward-predicts their attention patterns. In addition, though, there is an alternative, complementary possibility. Infants' attention shifts may initially happen as random, foraging-type behaviours [32,47] (i.e. not forward-predicted by fluctuations in infants' endogenous engagement or

interest); processes *after* the attention shift (determined by what information is present at the attended-to location) may drive increases in infants' endogenous engagement or interest which prolong that attention episode. (This distinction is similar to that we discussed above, about whether caregiver behaviours forwards-predict infant attention, or whether caregivers influence infants by reactively responding to their attention shifts, but operates at the individual level.) Consistent with this possibility, dynamic, generative models based on this framework can accurately predict attention patterns at least in younger infants [32,48]. Dynamic, amplificatory processes that take place after an attention shift can also explain patterns of attention inertia observed in naturalistic settings [13].

To examine how fluctuations in endogenous engagement or interest drive and/or maintain infant attention during naturalistic interactions, we can measure theta activity (3-6Hz), which is an oscillatory rhythm associated with intrinsically guided cognitive process in early infancy [49]. In particular EEG activity in the theta range has been found to increase over fronto-central electrodes during episodes of endogenously controlled attention. For example, theta activity over fronto-central electrodes increases where infants anticipate the next actions of an experimenter, and theta activity occurring in the time before infants look towards an object has been found to predict the length of time infants pay attention to that object during solitary play [50,51]. Recent work has also showed dynamic fluctuations in theta activity over the course of sustained attention episodes: Xie and colleagues found that, whilst 10-12 month-old infants viewed cartoon videos, theta activity increased during heart-rate defined periods of attentional engagement [52] (see also [53]).

In summary, therefore, research has examined two separate influences that could support how infants pay attention in social settings. The first type of influence is endogenous engagement

or interest. The second is caregivers' exogenous behaviour. But for both of them, it is unclear whether the influences are forwards-predictive or reactive. Does infants' endogenous attention engagement forwards-predict attention, or do fluctuations in engagement that take place after an attention shift predict how long that episode lasts? And do caregivers drive infant attention using salience cues, or do they reactively change their behaviours in response to infant behaviours?

Here, recording EEG from infants during naturalistic interactions with their caregiver, we examined the (inter)-dependent influences of infants' endogenous oscillatory neural activity, and inter-dyadic behavioural contingencies in organising infant attention. First, we examined processes endogenous to the infant that determine the timing of their attention during the interaction (part 2). Second, we examine caregiver behaviours (part 3).

In part 1, we first test whether oscillatory structures can be derived from the patterns of infant and caregiver looking behaviour at an individual level, by computing the partial auto-correlation function (PACF) for caregiver and infant attention durations. We then test whether infant and caregiver behaviours act as coupled oscillators, by examining the time-course of the cross-correlation function between infant and adult gaze [18]. If true, this would point to the existence of mechanisms of influence between infant and adult gaze that our other analyses, examining forwards- and backwards-predictive relationships (see below), would be unable to detect.

In part 2 we then assess whether infants' endogenous cognitive processing forward-predicts infant attention, by using cross-correlations to estimate the forwards- and backwards-predictive associations between infant theta activity, recorded over fronto-central electrodes

[54], and look durations. In addition, we further examined reactive changes in infant endogenous oscillatory neural activity that take place after the onset of an attention episode. To do so, we used two analyses: first, using linear-mixed effects models, we examined the direct temporal associations between the infant attention durations and the average levels of infant theta activity during that look. Next, we examined how theta activity changes dynamically across the course of individual looks.

In part 3 we examine the (inter)-dependent relationships between caregiver behaviours and infant attention. We examine two aspects of caregiver behaviour in particular. First (part 3.1), we examine caregiver gaze behaviour, using cross-correlations to test whether increases in caregiver attention towards objects forwards- or backwards-predicted changes in infant attention. In order to test whether any association between infant attention and caregiver behaviour was independent of the relationship between infant attention and their endogenous oscillatory neural activity, we also conducted cross-correlations to examine the associations between caregiver attention and infant theta activity. And we used the same two analyses as used in part 2 to examine how caregiver gaze behaviour changes reactively following the onset of an infant attention episode.

Second (part 3.2) we examined saliency in the caregiver's speech signal by computing the rate of change in the fundamental frequency of their voice [15]. For this, we used the same analysis approach. First, we conducted cross-correlations to examine whether changes in caregiver vocal behaviour forwards- or backwards-predict changes in infant attention. Second, we examined how caregiver vocal behaviour changes reactively following the onset of an infant attention episode. Allostatic attentional-structuring models predict reactive

change in caregiver behaviour at the onset of infant attention, and over the duration of the look, that associate with the length of infant looking.

Results

The results section is divided into three parts. In part 1, we first conduct descriptive statistics of infant attention durations, and test for oscillatory structures in caregiver and infant attention. Then, in part 2, we examine whether endogenous infant neural activity forwards-predicts fluctuations in infant attention, and/or reactively changes in the time after the onset of an attention episode. In part 3, we assess whether modulations in caregiver gaze and vocal behaviour forwards-predict fluctuations in infant attention, and/or reactively change in the time after infants shift their attention.

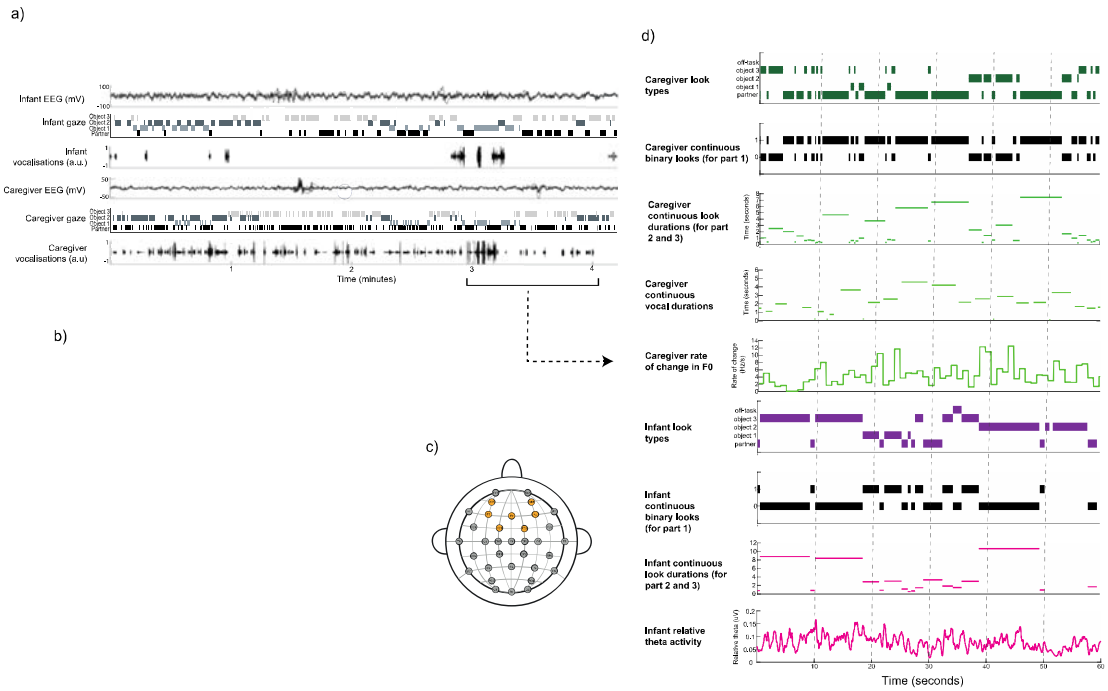

Figure 1. Experimental set-up and example of continuous variables. *a) Raw data sample, showing (from top) infant EEG over fronto-central electrodes, after pre-processing, infant gaze behaviour, infant vocalisations, caregiver EEG over fronto-central electrodes, caregiver gaze behaviour, caregiver vocalisations. b) Example camera angles for caregiver and infant (right and left), as well as zoomed-in images of caregiver and infant faces, used for coding. c) Topographical map showing electrode locations on the bio-semi 32-cap; fronto-central electrodes included in the theta activity analysis are highlighted in orange (AF3, AF4, FC1, FC2, F3, F4, Fz). d) Continuous behaviour and EEG variables extracted from the caregiver and infant time-series, showing (from top), caregiver looks to objects, the partner, and off-task looks, caregiver binary attention durations (for part 1), caregiver continuous attention durations (for part 2 and 3), caregiver vocalisation durations, rate of change in caregiver F0, infant looks to objects, their partner, and off-task looks, infant*

binary attention durations (for part 1), infant continuous look durations (for part 2 and 3), infant relative theta activity.

1 | Oscillatory structures in caregiver and infant attention

First, as descriptive statistics, we report on the frequency distribution of caregiver and infant attention durations towards objects, the partner, and periods of off-task attention, dividing attention durations into 100ms bins. Histograms showing the distribution of caregiver and infant attention durations towards objects, partners, and non-targets are displayed in Fig. 2a. In both distributions the mode is greater than the minimum value, consistent with previous observations that attention shifting is periodic [55]. The caregiver's distribution is also more left-skewed compared to the infants' distribution, reflecting the shorter and more frequent attention durations by the caregiver (c.f [23,51]). Finally, consistent with previous reports [10], caregivers tended to look towards their partner more frequently than infants, with infants attending most frequently to the objects (Fig. 2a).

Next, to investigate whether there was an oscillatory component in the caregiver and infant gaze time series, we computed the PACF of a binary attention variable separately for caregiver and infant using 100, 200, 500 and 1000ms lags (see Methods and Fig. 1d for more detail). In order to explore whether the PACF reflected the temporal interdependencies between infant/caregiver attention episodes (i.e. how likely an attention episode of a given length was to be followed by another of a similar length), or, more simply, the overall distributions of attention episodes (i.e. how common attention episodes of a given length are overall), the PACF was repeated after shuffling the infant and caregiver attention durations in time (see Methods for permutation procedure).

Fig. 2b shows that for the 100ms, 200ms time bins (infant and caregiver) and 500ms time bin (infant only), the lag 1 terms are negative, indicating that attention at time t is negatively predictive of attention at time t+x where x is a short time interval. This pattern is also observed in the baseline data (in which looks have been randomly shuffled in time). It reflects therefore, the overall pattern already shown in the histograms in Fig. 2a, that short looks (e.g. 100-200ms) are less frequent than longer looks (e.g. 500ms). It can also be seen that, at higher time lags, the observed PACF values are above the baseline rate. This indicates temporal interdependencies between look durations (i.e. that an attention episode of a given length is likely to be followed by another of a similar length), which are not present when the look durations are randomly shuffled to generate the baseline data.

Finally, we replicated previous analyses [18] to explore whether the inter-dyadic dynamics of caregiver and infant looking behaviour could be modelled as coupled oscillators. To do so, we first computed a binary attention variable, separately, for the infant and the caregiver. For this binary attention variable, we coded each look alternatively as a 0 or 1 from the first look of the interaction to the last, irrespective of where the infant or caregiver was looking (e.g. objects/ the partner; see Fig. 1d). To examine whether caregiver and infant attention changes at consistent temporal latencies, as would be the case if they were acting as entrained oscillators [17], we calculated the cross-correlation function between the infant and caregiver binary attention variables. If caregiver and infant gaze behaviour act as coupled oscillators, then the cross-correlation function should display significant peaks at regular intervals, reflecting these consistent latencies between attention shifts [18]. In order to identify where peaks in the cross-correlation function exceeded chance, we computed Poisson point process timeseries with look duration lengths matching the average look duration in the actual data

(see Methods for more details). Fig. 2c shows the results of this. Cluster-based permutation analysis revealed no significant peaks in the cross-correlation function, compared to baselines created through poisson process.

In summary, oscillatory mechanisms appear to govern both caregiver and infant attention durations; with infant attention durations centring around 1-2s in length, and adults around 200- 500ms. The cross-correlation analysis, however, suggested that caregiver and infant attention shifts do not act as coupled oscillators across the dyad (Fig. 2c).

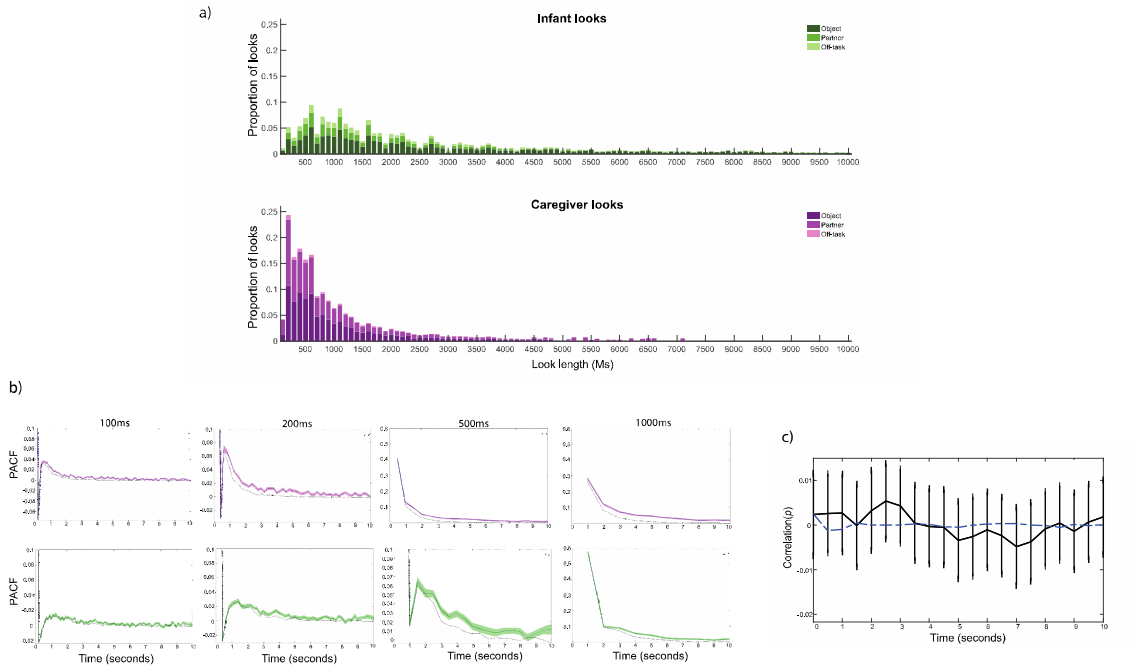

Figure 2. Testing for oscillatory patterns of attention behaviour in infants and caregivers. *a) Histogram of caregiver and infant attention episodes to objects, their partner and off-task episodes. Stacked bars show the number of episodes in each category for each 100ms bin for all episodes up to 10s in duration. b) PACF computed at different time lags for caregiver and*

infant gaze time series. Coloured lines show the PACF for infants (green) and caregivers (pink); shaded areas show the SEM. Dashed black lines show the PACF of shuffled attention duration data. c) Cross-correlation between caregiver and infant binary gaze variables. Black line shows the Spearman correlation coefficient at time-lags ranging from 0-10s; error bars indicate the SEM. Blue dashed line shows the permutation cross-correlation between two time series of poisson point process; one matching the average look rate of caregivers and the other, the average look rate of infants.

2 | Does endogenous infant neural activity forwards-predict infant attention, or reactively change following the onset of a new infant attention episode?

In this section, we investigate the relationship between infant endogenous oscillatory neural activity and infant attention durations, considering both forwards-predictive relationships and reactive changes in infants' endogenous oscillatory activity after the onsets of attention episodes.

2.1 Forwards predictive relationship between infant attention and infant theta activity

To examine whether infant endogenous neural activity significantly forwards-predicted infant attentiveness, we calculated a cross-correlation between the continuous infant attention duration time-series (see Fig. 1d), including all infant attention episodes to objects, the partner and looks elsewhere, and infant theta activity. Fig. 3a shows the results of the cross-correlation analysis. This analysis revealed a significant, positive association between the two variables at time-lags ranging from -2 to +6s ($p = 0.004$).

This indicates that infant theta power significantly forwards-predicted infant attention durations at lags up to 2 seconds, as well as that infant attention durations significantly forwards-predicted infant theta at lags of up to 6 seconds. Interpreting the exact time intervals over which a cross-correlation is significant is challenging due to the auto-correlation in the data [56,57], but there are two points of significance here. The first is the fact that the peak cross-correlation is observed not at time 0 but at time t+1.5 seconds (i.e. between looking behaviour at time t and theta power at time t+1.5 seconds). The second is that the significance window is asymmetric around time 0. Neither of these points can be attributable to residual auto-correlation. Overall, then, we can conclude that there is a temporally specific relationship between infant attention durations and theta power; and that attention durations forwards-predict theta power more than *vice versa*.

2.2 Reactive change in infant theta activity following look onset

In addition to the cross-correlation, we also conducted two further analyses to investigate the relationship between theta activity and the duration of attention episodes. First, we calculated a linear mixed effects model to examine the relationship between the lengths of infant attention episodes and average theta activity during that episode. This showed a significant, positive association between the two variables ($\beta = 0.33$; $p<0.001$); scatter plot between the two variables is shown in Fig. 3b. This indicates that higher average theta power across the attention episode associates with longer attention durations. Second, we explored dynamic change in theta activity relative to the onset of infant attention episodes towards objects. The modulation analysis (Fig. 3c), examining average infant theta activity during each third of a continuous look, showed that there was little change in infant theta activity over the duration

of infant attention episodes, for any duration time-bin: a series of Wilcoxon signed rank tests indicated decreases in infant theta activity for attention episodes lasting 1-3s, but this did not survive Benjamini-Hochberg correction.

2.3 Summary

In summary, there is a temporally specific relationship between infant attention durations and theta power, with attention durations forwards-predicting theta power more than vice versa (Fig 3a). Longer attention episodes are associated with increased average theta activity over the length of the episode (Fig. 3b), but little dynamical change in theta activity is observed over the course of an attention episode.

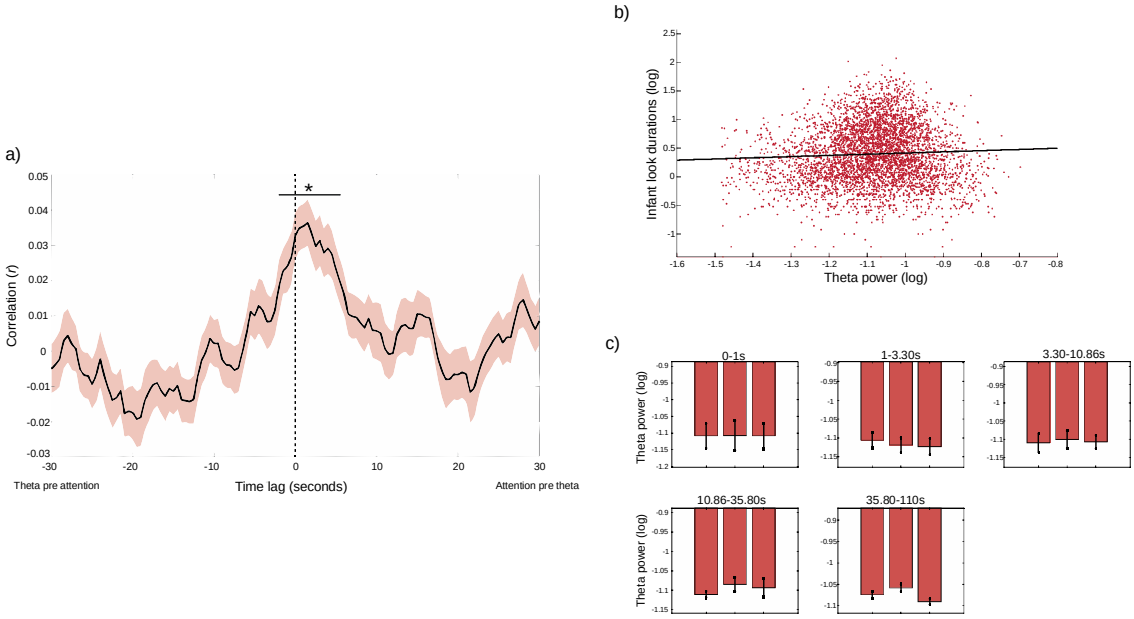

Figure 3: Relationship between infant attention durations and infant theta activity. a) *Cross-correlation between infant theta activity and infant attention durations. Black line shows the Pearson correlation at each time lag, coloured shaded areas indicate the SEM.*

*Significant time lags identified by the cluster-based permutation analysis are indicated by black horizontal lines (*p <0.05). Cluster-based permutation analysis revealed a significant cluster of time points ranging from -2 to +6 seconds (p = 0.004). b) Linear mixed effects model, predicting infant object attention durations from infant theta activity. The model reveals a significant positive association between infant theta activity and attention durations (β=0.33; p<0.001). c) Infant theta activity split into 3 attention chunks across the duration of attention episodes, binned according to episode length. Wilcoxon signed ranks tests explored significant differences between attention chunks, for each duration bin (*p < 0.05).*

3 | Does caregiver behaviour forwards-predict infant attention, or reactively change following the onset of a new infant attention episode?

First, we examine whether caregiver gaze behaviour associates with infant attentiveness (section 3.1). Second, we examine whether caregiver vocal behaviour associates with infant attentiveness, focusing on the rate of change of caregiver F0 as an index of auditory salience (section 3.2). In each case, we examine both forwards-predictive relationships and reactive changes in caregiver behaviour relative to the onsets of infant attention episodes.

3.1 Caregiver gaze behaviour

3.1.1 Forwards-predictive relationships between infant attention durations and caregiver attention durations

To examine whether caregiver attention forwards-predicts infant attentiveness, we conducted cross-correlation analyses between the continuous infant and caregiver attention durations towards the objects (see Fig. 1d). In order to test whether any association between infant attentiveness and caregiver attentiveness was independent of the relationship between infant attentiveness and their endogenous oscillatory neural activity shown in Fig. 3a, we also repeated these analyses relative to infant theta activity. Results are reported in Fig. 4. The cross-correlation between caregiver and infant attention durations peaks after lag zero (t+2.5 seconds), but cluster-based permutation analysis revealed no significant clusters of time points, though one cluster verged on significance (p=0.10). The cross-correlation function between caregiver attention durations and infant theta activity revealed a similar pattern (Fig.4b), peaking in the period after time 0, and the cluster-based permutation analysis revealed a significant cluster ranging from -*1 to 5s (p=0.012*). Although it is likely that the association between caregiver attention durations and infant theta shown in Fig. 4b is mediated by the association between caregiver attention durations and infant attention durations shown in Fig. 4a, the latter association is significant whereas the former is not. As for the analyses described in part 2, the exact time window over which the cross-correlation is significant cannot be interpreted due to autocorrelation in the data; but the fact that the peak cross-correlation is observed, again, at t+1.5 seconds, and that the significance window is asymmetric around time 0, both indicate that, overall, infant theta predicts caregiver attention durations more than *vice versa*.

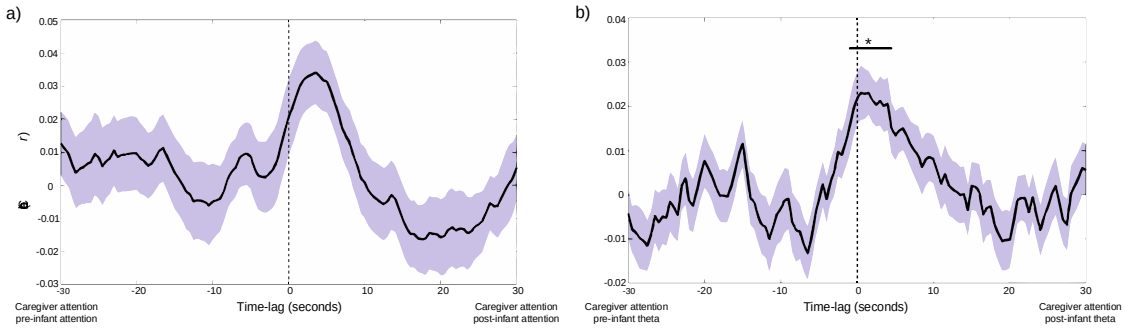

Figure 4. Assessing forwards-predictive associations between caregiver attention durations, infant attention durations, and infant theta activity. *Black lines show the Pearson cross-correlation between two variables; coloured shaded areas indicate the SEM. Black horizontal lines show significant clusters of time lags (*p < 0.05). a) Infant and caregiver continuous attention durations. Cluster based permutation analysis revealed no significant clusters of time points, although one cluster verged on significance (p=0.10). b) Infant theta activity and caregiver continuous attention durations. Cluster-based permutation analysis indicated one significant cluster ranging from -1 to 5s (p=0.012).*

3.1.2 Reactive change in caregiver look durations following infant look onset

To examine reactive change in caregiver attention to objects following the onsets of infant attention episodes to objects, we time-locked caregiver attention durations to infant attention onsets towards objects. Fig. 5a shows changes in caregiver attention durations around the onset of infant attention towards an object. Cluster-based permutation analysis revealed a significant cluster of time points 0 to 4 seconds post attention onset (p = 0.009), indicating that caregiver attention durations significantly decreased after the onset of a new infant

attention episode. Fig. 5b shows the same event-related analysis subdivided by infant attention duration. This revealed that the decrease in caregiver attention durations after infant attention onsets was significant for attention episodes lasting over 3s.

To investigate how caregiver behaviour changed over the course of infant object looks, we next employed the same modulation analysis as described in part 2.2, computing differences in mean caregiver attention durations between 3 equal-spaced chunks over the course of an infant object look. This analysis revealed that, in contrast to the first 4 seconds of an infant attention episode during which caregiver attention durations decreased, caregiver attention durations actually significantly increased over the course of the entire attention episode, with a Wilcoxon signed ranks test indicating a significant difference between the first chunk of an attention episode and the third (Fig. 5c). Dividing infant attention durations into log-spaced bins again revealed that this effect was driven by attention episodes lasting over 3s (Fig. 5d). Finally, we computed a linear mixed effect model to examine the relationship between infant object attention durations and caregiver object attention durations. Corresponding to the modulation analyses reported above, when we averaged over the course of the entire infant object attention episode, we found that longer infant object attention durations associated with longer average caregiver attention durations ($\beta = 0.16$, $p < 0.001$). Fig. 5e shows the scatter plot of the association between infant look durations and averaged caregiver look durations over the length of each individual infant look duration.

3.1.3 Summary

In summary, both the continuous and event-related analyses revealed that caregivers dynamically adapted their gaze behaviour in response to changes in infant attentiveness

during the interaction. Infant theta activity significantly forwards-predicted caregiver attention durations, suggesting that caregivers dynamically adapt their behaviour according to infant engagement (Fig 4b). Caregiver attention durations to objects decreased around the start of a new infant attention episode (Fig 5a); but overall, longer infant attention durations associated with longer attention durations by the caregiver towards objects (Fig 5e). These analyses demonstrate immediate, reactive, change in caregiver behaviour at the onset of infant attention towards an object, as well as slower-changing modulations in their behaviour over the length of an attention episode.

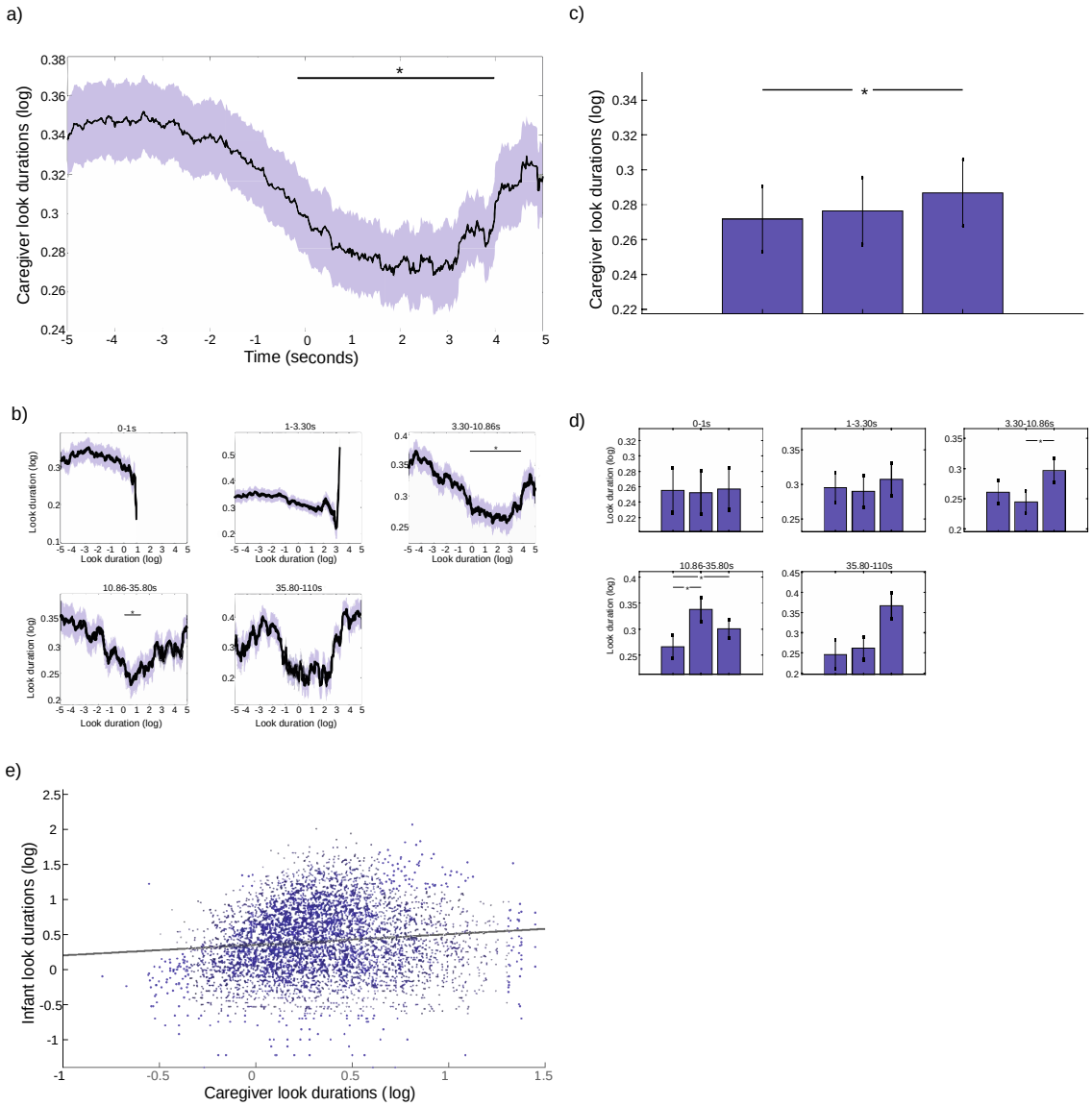

Figure 5. Dynamic event-locked association between infant object attention and caregiver attention durations. *a) Event-related analysis showing change in caregiver attention durations around infant attention onsets to objects: black line shows average caregiver attention durations (log); coloured shaded areas indicate the SEM. Black horizontal line shows areas of significance revealed by the cluster-based permutation analysis (p < 0.05). Cluster-based permutation analysis reveals a significant cluster of time points 0 to 4 seconds after attention onset (p = 0.009). b) event-related analysis split by infant object attention-*

duration time bins. Black lines show average caregiver attention durations (log); coloured shaded areas indicate the SEM. Black horizontal lines shows areas of significance revealed by the cluster-based permutation analysis ($p < 0.05$). Permutation analysis again revealed a decrease in caregiver attention durations in the time after attention onset for looks 3-35s long. c) Modulation analysis: each bar shows the median caregiver attention duration, across participants, for each chunk, averaged across all infant object attention durations. Wilcoxon signed ranks tests investigated significant differences between chunks ($p < 0.05$). d) same as c), for each infant object attention duration time bin. e) Scatter-plot showing the association between infant object attention durations and caregiver continuous attention durations. Coloured dots show each individual object attention duration; black line shows the linear line of best fit. Linear mixed effects modelling revealed a significant positive association between the two variables* ($\beta = 0.16$, $p < 0.001$).

3.2 Caregiver vocal behaviour

Next, we used an identical analysis approach to examine forwards-predictive and reactive associations between infant attention and caregiver vocal behaviours. Here, we concentrate on the rate of change in F0 as a marker of auditory saliency in the caregiver's voice. In additional analyses presented in the SM, we also examine caregiver vocal durations, and caregiver amplitude modulations (Figures S2 and S3).

3.2.1 Forwards-predictive relationships between infant look durations and caregiver vocal behaviour

First, we computed the cross-correlations between rate of change of caregiver F0, infant attention, and infant endogenous neural activity. Results are shown in Fig.6. Cluster-based permutation analysis revealed that the time-lagged associations between infant attention and rate of change in caregiver F0 did not exceed chance (Fig. 6a). To test whether there was any direct influence of caregiver behaviour on modulations in infant endogenous neural activity, the same analyses were subsequently repeated relative to infant theta activity (Fig. 6b): cluster-based permutation analysis again suggested no significant association between rate of change in caregiver F0 and infant endogenous neural activity. The same analyses are presented relative to caregiver vocal durations and amplitude modulations in Fig.S2, which showed a similar pattern of results.

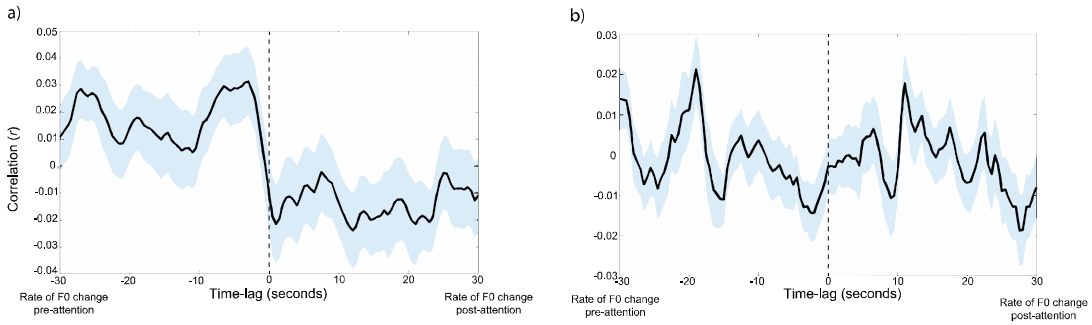

Figure 6: Assessing forwards-predictive associations between caregiver vocal behaviour, and infant attention, and infant theta activity. *Black lines show the Pearson cross-correlation between two variables; coloured shaded areas indicate the SEM. Black horizontal lines show significant clusters of time lags (*p<0.05). a) Rate of change in*

caregiver F0 and continuous infant attention durations. b) Rate of change in caregiver F0 and infant theta activity.

3.2.2 Reactive change in caregiver vocal behaviour following infant look onset

To examine whether caregivers reactively adapted their vocal behaviour to changes in infant attention, we repeated the same analysis presented in section 3.1.2, with rate of change of caregiver F0 as the dependent variable. The event-related analysis revealed no increase in the rate of change in caregiver F0 relative to infant attention onsets: cluster-based permutation analysis revealed no change above chance levels (Fig. 7b). This suggests that modulations in caregiver's speech were not immediately reactive to infant attention onsets towards objects. Over the length of individual attention episodes towards objects, however, linear mixed effects models revealed that longer object looks associated with a greater rate of change in caregiver F0 (β=0.13; p< 0.001; Fig. 7a), and, for looks lasting between 3-10 seconds, caregivers' decreased the rate of change in the fundamental frequency of their voice, over the course of a look (Fig. 7c,d). The exact same analysis relative to caregiver vocal durations and amplitude modulations showed a similar pattern of findings, which is presented in Fig.S3.

3.2.3 Summary

In summary, longer infant object look durations associated with a greater rate of change in caregiver F0, overall. Caregiver vocal behaviour showed no event-related change relative to infant attention onsets, but longer attention durations were associated with a decrease in the rate of change in F0.

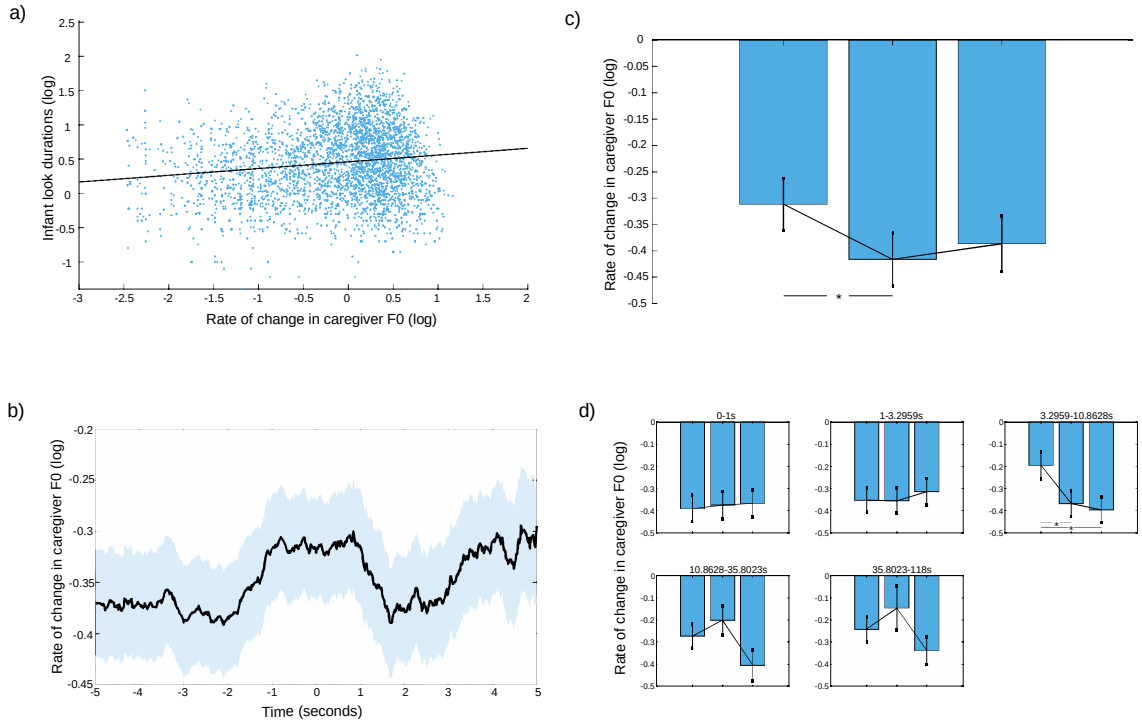

Figure 7. Reactive change in caregiver vocal behaviour relative to infant attention onsets. *a) Scatter plot of the association between infant attention durations and rate of change in caregiver F0. A linear mixed effects model revealed a significant positive association between the two variables* (β=0.13; *p*< 0.001), *b) Event related analysis examining reactive change in caregiver F0 in the time after the onset of an infant object look. Black line indicates the average across participants; coloured shaded area indicates SEM. c) Modulation analysis: each bar shows the median for each chunk across participants; errors bars show the SEM. Wilcoxon signed ranks tests explored significant differences between attention chunks (*p < 0.05), d) Same as c), binned by infant attention durations.*

Discussion

Recording EEG activity from infants whilst they engaged in shared, naturalistic interactions with their caregiver, we examined the endogenous mechanisms and bi-directional, interactive, contingencies that control the allocation of infant attention during social interaction. To do so, we conducted three sets of analyses. First, we examined whether caregiver and infant attention patterns act as coupled oscillators (part 1). Second, we examined how infants' endogenous neural activity forwards-predicted attention durations, and how it changed reactively relative to the onsets of infant attention episodes towards objects (part 2). Third, we examined how caregiver gaze and vocal behaviour forwards-predicted infant attention durations, and how it changed reactively to the onsets of infant object looks (parts 3.1 and 3.2).

When we examined whether and how endogenous cognitive processes predict infant attention, we found evidence for two distinct mechanisms. First, oscillatory mechanisms predict infant attention durations (part 1, Fig. 2), with a period centring around 1-2 seconds in length. Second, independently, fluctuations in neural markers of infants' engagement or interest forward-predict their attentiveness towards objects (part 2). Cross-correlation analyses revealed associations between infant theta activity and infant attention durations, such that increases in infant attention durations forwards-predicted increases in infant theta activity more than *vice versa* (Fig. 3a). Overall, average theta power during an attention episode correlated with the duration of that episode (Fig 3b). Infant theta activity did not, however, show any immediate change at the beginning of an attention episode, or modulate over the length of longer episodes (Fig. 3c; S1). This last result may appear inconsistent with other previous findings that theta activity increases over the course of a sustained attention

episode [52]. The reason for this is likely to be methodological, as Xie and colleagues measured infants' attention while they were alone watching unfolding events on TV, while infants in our task were playing with the same toy over the course of an attention episode, and were engaged in a social interaction with their caregiver.

Overall these findings suggest, partially consistent with the predictions of active learning models [58,59], that infants' own endogenous cognitive processing is one mechanism that drives and maintains infant attention during online interactions. Strikingly, however, we found that attention durations forward-predicted theta power more than *vice versa*. One possible interpretation of this finding is that longer attention durations by the infant drive incremental increases in infants' endogenous control over the allocation of their attention, through self-sustaining, bidirectional interactions between their own exploratory behaviours and information gain from the environment [38,39,41,55]. Whilst infants' attention shifts may often be initiated as random, foraging-type behaviours [32,47], at times, these self-sustaining interactions drive increases in infants' endogenous attention control, over the course of consecutive attention onsets.

Next, in order to evaluate the hypothesis that caregivers actively scaffold their infants' attention, we examined the association between caregiver behaviours and infant attention. Consistent with previous research [2] we found that oscillatory mechanisms govern both caregiver and infant attention durations, but that the oscillatory period of infant attention durations is shorter (centring around 1-2 seconds in length) compared with caregivers (centring around 200- 500ms in length) (part 1, Fig. 2a, 2b). However, when we examined whether infant and caregiver attention patterns act as coupled oscillators, which is one mechanism through which caregiver gaze behaviour might support infant gaze behaviour [30],

we found no evidence to support this (Fig. 2c). This suggests that mechanisms of influence between infant and caregiver attention are more likely to operate as lagged, forwards- or backwards- predictive relationships, as we investigated in part 2.

On the one hand, we found little to no evidence in support of the hypothesis that adult gaze and vocal behaviours forwards-predict infant attention (part 2). Against adult-led attentional structuring perspectives of early interaction, the cross-correlation analyses showed that, overall, fluctuations in infant look durations were not forwards-predicted by changes in caregiver look durations (Fig. 4a); rather, changes in infant neural engagement largely forward-predicted changes in caregiver attention durations (Fig. 4b). This association was likely partially mediated by the weaker and non-significant associations observed between infant attention and caregiver attention (Fig 4a). We also found no evidence for co-fluctuations between the rate of change of caregiver F0 (a marker of auditory salience) and infant attention durations (Fig.6), suggesting that increases in caregiver vocal saliency did not forward-predict changes in infant attention.

On the other hand, we did find evidence that caregivers rapidly modulated their behaviours in response to shifts in infant attention. This was particularly evident in adult gaze behaviour, where in addition to the cross-correlation findings (Fig. 4) our event-locked analyses showed that caregiver attention durations significantly decreased after the onset of a new infant attention episode (Fig. 5a). Over the duration of longer attention episodes, however, caregiver attention durations significantly increased (Fig.5c), so that, overall, the linear mixed effects model revealed that longer infant object looks were associated with longer looks by the adult partner (Fig.5d). A series of linear mixed effects models also revealed that longer infant attention durations co-occurred with a greater rate of change in caregiver F0, as well as

longer caregiver vocalisations, and an increase in caregiver amplitude (Fig.7; S3). The modulation analysis further showed that longer infant looks (those lasting between 3-10 seconds) were associated with a decrease in the rate of change in caregiver F0 over the course of the look. In contrast to caregiver gaze behaviour, however, there was little dynamic change in caregiver vocal behaviour immediately after attention onset (Fig.7, Fig.S3).

Overall, the caregiver behaviours we studied were largely reactive to changes in infant attention. The rapid change in caregiver gaze in response to the onset of infant attention towards objects, beginning just before attention onset, suggests that it is unlikely that caregivers are responding to active attention sharing cues produced by the infant [42,44,60]. Indeed, similar to previous micro-behavioural studies of 12-month-old infants in shared interactions [10], infants rarely looked towards their caregiver's face (Fig.2a), and, in a previous analysis of this data, infants did not increase looks to their partner's face in the time before leading an episode of joint attention [61]. It seems therefore unlikely that the relationship between infant attention and fluctuations in their own endogenous cognitive processing is related to intentionally mediated forms of communication by the infant, with the goal of directly eliciting information from their caregiver [10,61].

Instead, caregivers are anticipating shifts in infant attention, and, in line with an allostatic model of inter-personal interaction, 'catching' infants' attention, and monitoring their behaviour [2]. This increase in the rate of caregiver behaviour after look onsets could reflect dynamic up-regulatory processes that serve to maintain infant attention: though not reflected in their vocalisations; other fast-changing salient cues such as hand movements and facial affect could also increase in variability [62]. The down-regulation of caregiver attention over the course of longer attention episodes by the infant might subsequently index decoupling of

caregivers' regulatory processes from infant attention; this is also reflected in the decreased rate of change in caregiver F0 (Fig.6f). Combined, therefore, our findings suggest that, during interactions at the end of the first year, infant attention is structured through joint but independent influences of caregiver responsivity and regulation, and their own intrinsically motivated engagement.

In this perspective, our results can be interpreted relative to neurocomputational, associative accounts of active learning in early infancy [39,59]. These accounts postulate that contingent changes in the environment in response to actions produced by the infant improves infants' prediction and control over their own behaviour [63,64]. In the context of shared interaction, consistent and contingent responsiveness by the caregiver to infant attention gives meaning to infants' behaviour, increasing infant engagement and further exploratory behaviours [39,65]. Over time, therefore, infants' experience of repeated interactive contingencies could influence how controlled processes begin to guide their attention, as well as their sensitivity to and engagement in intentionally mediated forms of shared communication [65].

This has implications for how we view and understand the interactive processes that support how infants begin to use and engage with a language system. Previous accounts have emphasised the role of the caregiver in structuring infant learning in joint attentional frames, where they use clear ostensive signals to guide infant attention, and support word-object representations [42,66]. The present study, however found no evidence that increases in salient cues by the caregiver forward-predicted increases in infant attention durations. Increases in infant attentiveness are instead related to inter-dyadic, sensorimotor processes that are independent of the influence of infants' own endogenous cognitive process. How these fast-

acting intra- and inter-individual influences on infant attention support early language acquisition should be a key focus for future research [23,67].

The naturalistic design of our study is a strength as well as a limitation. Of note, we were unable to control how much infants moved during the interaction, which may have contributed to eye movement artifacts time-locked to shifts in infant attention. However, eye-movement artifacts were removed using ICA decomposition, and, though this does not remove all artifact introduced to the EEG signal [68], the relationships observed between infant attention and theta activity suggest that this did not affect our main findings. If eye-movement artifact influenced the association between infant attention duration and theta activity, shorter attention episodes ought to associate with more theta activity, which was not the case.

In future work it will be important to take a more holistic, computational and multi-modal approach to studying how factors intrinsic to the infant, and the inter-personal behavioural contingencies of the dyad, structure infant attention and behaviour [69]. For example, studying how inter-related multi-modal patterns of caregiver behaviour, such as body movement [62], facial affect [24,70] and vocalisations [26] support infants' engagement in joint attention, will build on the work that we reported here. In addition to the micro-dynamic analyses that we present here, it will also be important for future work to employ modelling approaches to further investigate infants' neural entrainment to the unidirectional and inter-dyadic action-generated contingencies of shared interaction [71,72]. A particular focus of this work should be on studying the temporal latencies at which entrainment and/or behavioural responsivity occur; utilising eye tracking methods will help with this. Finally, a limitation of our study is that our findings might reflect a particular caregiving style (that of middle-class mothers living in East

London), and it will be important in future research to study other populations, to investigate whether our results generalize to other populations and caregiving practices[73].

Overall, our findings suggest that infant attention in early interaction is asymmetric, related to their own endogenous cognitive processing and to consistent, reactive contingency to changes in their attention by the caregiver. Active learning strategies operate across the dyad; and are likely foundational to early language acquisition and socio-cognitive learning.

Materials and Methods

Participants

Ninety-four caregiver-infant dyads took part in this study. The final overall sample with usable, coded, gaze data was 66 (17 infants were excluded due to recording error or equipment failure, 4 infants were excluded for fussiness and 6 infants were excluded due to poor quality EEG data, and limited coding resources). Of the infants with usable gaze data, 51 had additional vocal data (15 excluded due to recording error/equipment failure. Of those with gaze data, 60 infants had usable EEG data (a further 6 excluded due to noisy EEG data (see artifact rejection section below). All usable data sets available for each separate analysis were used in the results reported below (e.g. infants with gaze and EEG data but no vocal data are included in analyses exploring the relationship between infant EEG and gaze). The mean age of the final overall sample (n=66) was 11.18 months (SD=1.27); 33 females, 30 males. All caregivers were female. Participants were recruited through baby groups and Childrens' Centers in the Boroughs of Newham and Tower Hamlets, as well as through online platforms such as Facebook, Twitter and Instagram. Written informed consent was obtained from all participants before taking part in the study, and consent to publish was obtained for all identifiable images used. All experimental procedures were reviewed and approved by the University of East London Ethics Committee.

Experimental set-up

Parents and infants were seated facing each other on opposite sides of a 65cm wide table. Infants were seated in a high-chair, within easy reach of the toys (see Fig. 1b). The shared toy play comprised two sections, with a different set of toys in each section, each lasting ~5 minutes each. Two different sets of three small, age-appropriate toys were used in each section; this number was chosen to encourage caregiver and infant attention to move between the objects, whilst leaving the table uncluttered enough for caregiver and infant gaze behaviour to be accurately recorded cf.[10].

At the beginning of the play session, a researcher placed the toys on the table, in the same order for each participant, and asked the caregiver to play with their infant just as they would at home. Both researchers stayed behind a screen out of view of caregiver and infant, except for the short break between play sessions. The mean length of joint toy play recorded for play section 1 was 297.28s (SD=54.93) and 323.18s (SD=83.45) for play section 2.

Equipment

EEG signals were recorded using a 32-chanel BioSemi gel-based ActiveTwo system with a sampling rate of 512Hz with no online filtering using Actiview Software. The interaction was filmed using three Canon LEGRIA HF R806 camcorders recording at 50 fps. Parent and infant vocalisations were also recorded throughout the play session, using a ZOOM H4n Pro Handy Recorder and Sennheiner EW 112P G4-R receiver.

Two cameras faced the infant: one placed on the left of the caregiver, and one on the right (see Fig. 1b). Cameras were placed so that the infant's gaze and the three objects placed on the table were clearly visible, as well as a side-view of the caregiver's torso and head. One camera faced the caregiver, positioned just behind the left or right side of the infant's high-chair (counter-balanced across participants). One microphone was attached to the caregiver's clothing and the other to the infant's high-chair.

Caregiver and infant cameras were synchronised to the EEG via radio frequency (RF) receiver LED boxes attached to each camera. The RF boxes simultaneously received trigger signals from a single source (computer running MATLAB) at the beginning of each play section, and concurrently emitted light impulses, visible in each camera. Microphone data was synchronised with the infants' video stream via a xylophone tone recorded in the infant camera and both microphones, which was hand identified in the recordings by trained coders. All systems were extensively tested and found to be free of latency and drift between EEG, camera and microphone to an accuracy of +/- 20 ms.

Video coding

The visual attention of caregiver and infant was manually coded using custom-built MATLAB scripts that provided a zoomed-in image of parent and infant faces (see Fig. 1b). Coders indicated the start frame (i.e. to the closest 20ms, at 50fps) that caregiver or infant looked to one of the three objects, to their partner, or looked away from the objects or their partner (i.e. became inattentive). Partner attention epsiodes included all looks to the partner's face; looks to any other parts of the body or the cap were coded as inattentive. Periods where the researcher was within camera frame were marked as uncodable, as well as instances

where the caregiver or infant gaze was blocked or obscured by an object, or their eyes were outside the camera frame. Video coding was completed by two coders, who were trained by the first author. Inter-rater reliability analysis on 10% of coded interactions (conducted on either play section 1 or play section 2), dividing data into 20ms bins, indicated strong reliability between coders (kappa=0.9 for caregiver coding and kappa=0.8 for infant coding).

Vocalisation coding

The onset and offset times of caregiver and infant vocalisations were identified using an automatic detector. The algorithm detected voiced segments and compared the volume and fundamental frequency detected in each recorded channel to infer the probable speaker (caregiver vs. infant). Identification of the onset and offset times of the detector then underwent a secondary analysis by trained coders, who identified misidentification of utterances by the automatic decoder, as well as classifying the speaker for each vocalisation. As the decoder did not accurately identify onset and offset times of caregiver and infant during co-vocalisations, and, as these vocalisations could not be included in analyses of the spectral properties of caregiver vocalisations, these were excluded from all analyses. The mean percentage of caregiver vocalisations that were co-vocalisations was less than 20%: 19.43 (SD=12.36; a box plot across all participants is presented in Fig. S4). In a previous analysis conducted on a sub-sample of the data[61], we have shown that there is no significant change in infant vocalisations, relative to the onset of infant attention episodes, and their vocal beahviour did not distinguish between moments that they either led or followed their partners' attention during the interaction. It is therefore unlikely that inclusion of co-vocalisations in the current analyses would affect the main findings, time-locking caregiver vocalisations to infant attention.

Infant EEG artifact rejection and pre-processing

A fully automatic artifact rejection procedure including ICA was adopted, following procedures from commonly used toolboxes for EEG pre-processing in adults [74,75] and infants [76,77], and optimised and tested for use with our naturalistic infant EEG data [78,79]. This was composed of the following steps: first, EEG data were high-pass filtered at 1Hz (FIR filter with a Hamming window applied: order 3381 and 0.25/ 25% transition slope, passband edge of 1Hz and a cut-off frequency at -6dB of 0.75Hz). Although there is debate over the appropriateness of high pass filters when measuring ERPs (see [80], previous work suggests that this approach obtains the best possible ICA decomposition with our data [68,81]. Second, line noise was eliminated using the EEGLAB [74] function *clean_line.m* [75].

Third, the data were referenced to a robust average reference [74]. The robust reference was obtained by rejecting channels using the EEGLAB *clean_channels.m* function with the default settings and averaging the remaining channels. Fourth, noisy channels were rejected, using the EEGLAB function *clean_channels.m*. The function input parameters 'correlation threshold' and 'noise threshold' (inputs one and two) were set at 0.7 and 3 respectively; all other input parameters were set at their default values. Fifth, the channels identified in the previous stage were interpolated back, using the EEGLAB function eeg_interp.m. Interpolation is commonly carried out either before or after ICA cleaning but, in general, has been shown to make little difference to the overall decomposition [70]. Infants with over 21% (7) electrodes interpolated were excluded from analysis. After exclusion, the mean number of electrodes interpolated for infants was 3.37 (*SD*=2.27) for play section 1, and 3 (*SD*=2.16) for play section 2.

Sixth, the data were low-pass filtered at 20Hz, again using an FIR filter with a Hamming window applied identically to the high-pass filter. Seventh, continuous data were automatically rejected in a sliding 1s epoch based on the percentage of channels (set here at 70% of channels) that exceed 5 standard deviations of the mean channel EEG power. For example, if more than 70% of channels in each 1-sec epoch exceed 5 times the standard deviation of the mean power for all channels then this epoch is marked for rejection. This step was applied very coarsely to remove only the very worst sections of data (where almost all channels were affected), which can arise during times when infants fuss or pull the caps. This step was applied at this point in the pipeline so that these sections of data were not inputted into the ICA. The mean percentage of data removed in play The mean percentage of data removed in play section 1 was 11.30 (SD=14.97), and 6.57(SD=6.57) for play section 2.

Data collected from the entire course of the play session (including play section 1 and play section two, as well as two further five minute interactions) were then concatenated and ICAs were computed on the continuous data using the EEGLAB function runica.m. After ICA rejection, data from each play section were re-split.

Pre-processing of continuous variables

Prior to conducting our main analyses, all primary variables of interest were converted into continuous variables, in order to perform time-lagged and event-locked methods of analysis, relative to infant attention (see Fig. 1d). All continuous variables were down sampled to match the sampling rate of the video cameras (50Hz).

Infant theta activity over fronto-central electrodes

First, missing data points were excluded from the continuous time-series. Where one or more of the fronto-central electrodes of an individual infant exceeded 100uV for more than 15% of the interaction, the infant's continuous theta time-series was excluded from analyses. Next, time-frequency decomposition was conducted via continuous morlet wavelet analysis to extract EEG activity occurring at frequencies ranging from 1-16Hz. Specifically, the EEG signal at each channel was convolved with Gaussian-windowed complex sine-waves, ranging from 1-16Hz, in linearly spaced intervals. The width of the guassian was set to 7 cycles. Power was subsequently extracted as the absolute value squared, resulting from the complex signal. After decomposition, to get rid of edge artifacts caused by convolution, the first and last 500ms of the time series were treated as missing data points. Missing data points were then re-inserted into the continuous variable as blank values, and the 500ms before and after these chunks of data also excluded. For each time point, for each frequency, power was expressed as relative power (i.e. the total power at that frequency, divided by the total power over all frequencies). EEG activity was then averaged over frequencies ranging from 3-6Hz, and averaged over fronto-central electrodes (AF3, AF4, FC1, FC2, F3, F4, Fz*;* see Fig. 1). This electrode cluster was chosen based on previous infant literature [54]. This continuous, one-dimensional variable was then downsampled from 512 to 50Hz by taking the median theta activity for every 10 samples of data, and, in each second, taking an extra 1 sample for 3 time points and an extra 2 samples for 1 time point. The spacing of these added samples was shuffled for each second of data.

Continuous attention durations

An attention episode was defined as a discrete period of attention towards one of the play objects on the table, or to the partner. The end of each attention episode was defined as the moment where the participant first looked away from the target towards another object, towards the partner, or towards another location that was not either the object or the partner (coded as non-target attention). See Fig. 1d for an example. Parts of the caregiver/infant gaze coded as uncodable were treated as missing data points, as well as the looks occurring in the time just before and after (in order to account for the fact that we do not know how long these looks last).

For the analyses in parts 2 and 3, which examine the associations between attention durations and other measures, we recoded each look based on the duration in seconds of that look. The durations of each look were then used to produce a continuous look duration variable, irrespective of whether that look was towards the object, partner, or non-target (see Fig 1d). These analyses examine therefore the associations between the durations of attention episodes and, respectively, endogenous infant neural activity (part 2) and caregiver behaviour (part 3).

Binary attention durations

For the analyses in part 1, which examine the temporal oscillatory patterns of attention shifts, we recoded each look alternatively as a 0 or 1 from the first look of the interaction to the last (see Fig. 1d). These analyses examine therefore the temporal inter-dependencies between

attention durations (within an individual and across the dyad), irrespective of where the attention is directed.

Rate of change in the fundamental frequency (F0) of the caregiver's voice

The fundamental frequency of the caregiver's voice was extracted using Praat [82], with floor and ceiling parameters set between 75-600Hz. Caregiver fundamental frequency was placed into the continuous variable only where the coder had identified that section of speech as the caregiver speaking, so that infant vocalisations were not included in the analysis. Due to the caregiver being within variable distance of their microphones, some clipping was identified in a sample of the microphone recordings. A stringent clipping identification algorithm was used (see SM 1.1 and Fig.S5) to remove parts of the microphone data where clipping occurred [83]. Vocalisations where any clipping was identified were set to missing data points. Interactions with more than 30% missing vocalisations were excluded from the analyses. Statistics on the number of vocalisations excluded on this basis is presented in Fig. S5. Co-vocalisations were set to missing data points.

Next, unvoiced sounds and periods between vocalisations were interpolated, using MATLAB's interp1 function. To reduce the likelihood of background noise (e.g. toy clacks) affecting the fundamental frequency, the interpolated variable was low-pass filtered at 20Hz using a 9th order butterworth filter. The rate of change in the caregivers' fundamental frequency was computed by taking the sum of the derivative in 1000ms intervals. The start and end points of each interval were then converted to time in camera frames, and the rate of change values inserted for the 50 corresponding frames.

See SM section 1.2 and 1.3 for description of the computation of caregiver vocal durations and amplitude modulations.

Analysis procedures

Procedures for part 1

Partial autocorrelation function

The partial auto-correlation function (PACF) of the caregiver and infant gaze time series was computed separately, over a range of time intervals, from 100-1000ms. First, the gaze time series was converted to a continuous binary variable, with either a 1 or 0 inserted into the time series for the duration of each attention episode, alternated for each consecutive look. The PACF was then computed by fitting an ordinary least squares regression model, at time-lags ranging from 0 to 10s, in 100ms intervals, controlling for all previous time-lags on each iteration. This analysis was repeated at intervals of 200, 500 and 1000ms.

Shuffled time series

To investigate whether the shape of the PACF reflected the temporal distribution of infant/caregiver attention episodes or more simply the frequency distribution (i.e. infant/caregiver attention episodes frequently last a similar length [84]), we conducted a permutation procedure, whereby, for each infant, their attention duration time series was shuffled randomly in time to produce a binary gaze time series of shuffled attention durations. The PACF was then computed for this time series in exactly the same way

described above. This procedure was subsequently repeated 100 times for each participant, before averaging over all permutations and participants.

Binary cross-correlation

Computation

The cross-correlation between caregiver and infant binary attention time-series was computed at lags 0 to +10s in 500ms intervals. The zero-lagged cross-correlation was first computed between the two binary attention variables using a Spearman correlation. The infant's time series was then moved forwards in time and the Spearman correlation computed between the two time-series at each 500ms interval. The cross-correlations at each time-lag were then averaged over the two interactions for each participant, and then averaged over all participants.

Poisson baselines

Poisson baselines were created by computing time series of the Poisson point process with the length of look durations matching the average length of look durations in the actual data, for caregivers and infants separately [18]. These variables were then converted to binary look duration variables, and the cross-correlation between the two binary time-series computed in exactly the same way described above. This procedure was repeated 100 times in order to create a baseline permutation distribution.

Significance testing

A cluster-based permutation approach was used to investigate whether the binary cross correlation differed significantly from the Poisson baseline distribution over any time-period. This approach controls for family-wise error rate using a non-parametric Monte Carlo method [85]. First, the cross-correlation at each time lag in the observed data was compared with the Poisson baseline distribution at that time lag, and values falling above the 97.5th centile and below the 2.5th centile were accepted as significant (corresponding to a significance level of 0.05). Next, to examine the likelihood of clusters of significant time points in the observed data occurring by chance, a cluster-threshold was computed using a leave-on-out procedure with the Poisson baselines. On each iteration, one baseline was compared with the 99 other baselines, and significant time-points identified using the same method described above. The largest cluster found on each iteration was identified to create a random permutation distribution of cluster sizes. The clusters identified in the observed data were then compared with this permutation distribution of maximum cluster sizes, and clusters falling above the 95th centile were considered significant (corresponding to a significance level of 0.05).

Procedures for parts 2 and 3

Cross-correlation analyses

Cross-correlations were computed between continuous infant attention durations, the continuous caregiver variables and infant theta activity. All analyses for the continuous caregiver variables were subsequently repeated relative to infant theta activity.

Computation

First, the time series of each variable were log transformed, and outliers falling 2 inter-quartile ranges above the upper quartile and two inter-quartile ranges below the lower quartile removed. A detrend was then applied to each variable; linear and quadratic bivariate polynomials were fit to each transformed time-series, and the residuals of the model of best fit computed. The cross-correlation between the two variables was then computed at lags -30 to +30s in 500ms intervals. The zero-lagged cross-correlation was first computed between the two variables using a Pearson correlation. The caregiver's time series (or infant theta activity where this was computed relative to infant attention durations) was then moved backwards in time (to compute negative lag correlations), or forwards in time (to compute positive lag correlations), and the Pearson correlation computed between the two time-series at each 500ms interval. In this way, we estimated how the association between the two variables changed with increasing time lags. The cross-correlations at each time-lag were then averaged over the two interactions for each participant, and then averaged over all participants.

Significance testing

A cluster-based permutation approach was used to investigate whether the time-lagged cross correlation differed significantly from chance over any time period. This approach controls for family-wise error rate using a non-parametric Monte Carlo method [85]. To create a random permutation distribution at each time-lag, each participant was randomly paired with another participant, through a process of derangement, and the cross-correlation between the caregiver and infant variables computed, and averaged over participants in exactly the same way described above. This procedure was then repeated 1000 times, resulting in a random permutation distribution at each time lag. Next, the cross-correlation at each time lag in the observed data was compared with the permutation distribution at that time lag, and values falling above the 97.5th centile and below the 2.5th centile were accepted as significant (corresponding to a significance level of 0.05). To examine the likelihood of clusters of significant time points in the observed data occurring by chance, a cluster-threshold was computed using a leave-on-out procedure on the permutation data. On each iteration, one permutation was compared with the 999 other permutations, and significant time-points identified using the same method described above. The largest cluster found on each iteration was identified to create a random permutation distribution of cluster sizes. The clusters identified in the observed data were then compared with this permutation distribution of maximum cluster sizes, and clusters falling above the 95th centile were considered significant (corresponding to a significance level of 0.05).

Linear mixed effect models

Linear mixed effect models were used to investigate the relationship between caregiver behaviour, infant theta activity and infant attention durations. First, for each participant, for each object attention episode, the continuous caregiver behavioural variable was averaged over the length of the infant attention episode, to obtain one value for caregiver behaviour per infant attention duration. Next, each variable was log-transformed, and outliers 2 inter-quartile ranges above the upper quartile and two inter-quartile ranges below the lower quartile removed. Finally, linear mixed effects models were fitted, with caregiver behaviour, or infant theta activity as the fixed effect, and infant attention durations as the response variable, with a random effect of participant.

Attention onset event-related analysis

Computation

Before event-locking the continuous variables to infant attention, the continuous variable was log-transformed, and outliers removed, applying a similar procedure to that described above. First, the frame of the onset of each infant object look, as well as the duration of that look was extracted from the infant gaze time series. Next, for each continuous variable, the frames occurring five seconds before and five seconds after the onset of each infant look were extracted from the caregiver time series. Given the fact that we were interested in how caregiver behaviour changed around the onset of an attention episode, where the infant shifted gaze again in the 5 second time period after attention onset, the values in the continuous caregiver variable were set to missing data points. The continuous frames

occurring before and after each look were then averaged over looks, for each interaction, resulting in an averaged continuous variable along the time dimension. These values were then averaged over interactions for each participant, before averaging over all participants.

In order to explore the possibility that the length of the infant attention episode might affect how the caregiver's behaviour changed around the onset of that episode, exactly the same analysis was repeated on attention durations of different lengths, in 5 log-spaced intervals, ranging from 0 to the longest attention episode identified across the datasets (118s).

Significance testing

Significance testing followed exactly the same procedures outlined in the cross-correlation analysis section.

Modulation during attention episodes

For this analysis, all continuous data variables (caregiver behaviour / infant theta activity) were log transformed and outliers removed (see above). Then, for each infant object look, the continuous caregiver behaviour / infant theta activity was extracted over the length of that attention episode, and divided into 3 equal-spaced chunks. The continuous data variable occurring in the first half of each chunk was then averaged for each attention episode, before being averaged over all episodes for that interaction. Averaged chunks from play section 1 and play section 2 were then averaged together for each participant, and the mean over all participants, for each chunk, computed. A series of Wilcoxon Signed ranks tests assessed whether the chunks differed to each other, compared to that which would be expected by

chance. The Benjamini-Hochberg false discovery rate procedure was applied to correct for multiple comparisons ($p < 0.05$ [86]).

Similar to the event-related analysis, infant object look durations were divided into 5 log-spaced bins to assess whether modulations in infant endogenous cognitive processing or caregiver behaviour differed for episodes lasting different lengths: exactly the same procedure was repeated for each duration bin.

References

1. Trevarthen, C. Intrinsic motives for companionship in understanding: Their origin, development, and significance for infant mental health. *Infant Mental Health Journal* **22**, 95–131 (2001).
2. Yu, C. & Smith, L. B. The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current Biology* **26**, 1235–1240 (2016).
3. Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E. & Matthews, D. Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Dev Sci* **23**, (2020).
4. Henning, A., Striano, T. & Lieven, E. V. M. Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development* **28**, 519–536 (2005).
5. Murray, L. *et al*. Randomized controlled trial of a book-sharing intervention in a deprived South African community: effects on carer-infant interactions, and their relation to infant cognitive and socioemotional outcome. *Journal of Child Psychology and Psychiatry* **57**, 1370–1379 (2016).
6. Bornstein, M. H. How infant and mother jointly contribute to developing cognitive competence in the child. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7470–7473 (1985).
7. Csibra, G. & Gergely, G. Social learning and social cognition: The case for pedagogy. 15.
8. Csibra, G. & Gergely, G. Natural pedagogy. *Trends in Cognitive Sciences* **13**, 148–153 (2009).
9. Sebanz, N. & Knoblich, G. Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science* **1**, 353–367 (2009).

10. Yu, C. & Smith, L. B. Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE* **8**, e79659 (2013).
11. Suarez-Rivera, C., Smith, L. B. & Yu, C. Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology* **55**, 96–109 (2019).
12. Wass, S. V. *et al.* Infants' visual sustained attention is higher during joint play than solo play: is this due to increased endogenous attention control or exogenous stimulus capture? *Developmental Science* **21**, e12667 (2018).
13. Richards, J. E. & Anderson, D. R. Attentional inertia in children's extended looking at television. in *Advances in Child Development and Behavior* vol. 32 163–212 (Elsevier, 2004).
14. Cooper, R. P. & Aslin, R. N. Developmental Differences in Infant Attention to the Spectral Properties of Infant-Directed Speech. *Child Development* **65**, 1663 (1994).
15. Nencheva, M. L., Piazza, E. A. & Lew-Williams, C. The real-time dynamics of child-directed speech: Using pupillometry to evaluate children's processing of natural pitch contours. *The Journal of the Acoustical Society of America* **145**, 1765–1765 (2019).
16. Meyer, L., Sun, Y. & Martin, A. E. Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience* 1–11 (2019) doi:10.1080/23273798.2019.1693050.
17. Wass, S. V., Perapoch Amadó, M. & Ives, J. Oscillatory entrainment to our early social or physical environment and the emergence of volitional control. *Developmental Cognitive Neuroscience* **54**, 101102 (2022).
18. Takahashi, D. Y., Narayanan, D. Z. & Ghazanfar, A. A. Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. *Current Biology* **23**, 2162–2168 (2013).

19. Wilson, M. & Wilson, T. P. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review* **12**, 957–968 (2005).
20. Gratier, M. *et al.* Early development of turn-taking in vocal interaction between mothers and infants. *Front. Psychol.* **6**, (2015).
21. Jaffe, J., Beebe, B. & Feldstein, S. Rhythms of Dialogue in Infancy: Coordinated Timing in Development. **66**, 1–149 (2001).
22. Feldman, R., Greenbaum, C. W. & Yirmiya, N. Mother-infant affect synchrony as an antecedent of the emergence of self-control. *Developmental Psychology* **35**, 223–231 (1999).
23. Yu, C. & Smith, L. B. Embodied attention and word learning by toddlers. *Cognition* **125**, 244–262 (2012).
24. Murray, L. *et al.* The functional architecture of mother-infant communication, and the development of infant social expressiveness in the first two months. *Scientific Reports* **6**, (2016).
25. Albert, R. R., Schwade, J. A. & Goldstein, M. H. The social functions of babbling: acoustic and contextual characteristics that facilitate maternal responsiveness. *Dev Sci* **21**, (2018).
26. Goldstein, M. H. & Schwade, J. A. Social Feedback to Infants' Babbling Facilitates Rapid Phonological Learning. *Psychological Science* **19**, 515–523 (2008).
27. Yoo, H., Bowman, D. A. & Oller, D. K. The Origin of Protoconversation: An Examination of Caregiver Responses to Cry and Speech-Like Vocalizations. *Front. Psychol.* **9**, 1510 (2018).
28. Wu, Z. & Gros-Louis, J. Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *J. Child Lang.* **42**, 538–561 (2015).

29. Nuthmann, A., Smith, T. J., Engbert, R. & Henderson, J. M. CRISP: A computational model of fixation durations in scene viewing. *Psychological Review* **117**, 382–405 (2010).

30. Nuthmann, A. & Henderson, J. M. Object-based attentional selection in scene viewing. *Journal of Vision* **10**, 20–20 (2010).

31. Nuthmann, A. & Matthias, E. Time course of pseudoneglect in scene viewing. *Cortex* **52**, 113–119 (2014).

32. Robertson, S. S. Empty-headed dynamical model of infant visual foraging: Empty-Headed Model. *Dev Psychobiol* **56**, 1129–1133 (2014).

33. Renswoude, D. R., Visser, I., Raijmakers, M. E. J., Tsang, T. & Johnson, S. P. Real□world scene perception in infants: What factors guide attention allocation? *Infancy* **24**, 693–717 (2019).

34. Stallworthy, I. C. *et al.* Infants' gaze exhibits a fractal structure that varies by age and stimulus salience. *Sci Rep* **10**, 17216 (2020).

35. Feldman, R. & Mayes, L. C. The cyclic organization of attention during habituation is related to infants' information processing. *Infant Behavior and Development* **22**, 37–49 (1999).

36. Gottlieb, J., Oudeyer, P.-Y., Lopes, M. & Baranes, A. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences* **17**, 585–593 (2013).

37. Goupil, L. & Proust, J. Curiosity as a metacognitive feeling. *Cognition* **231**, 105325 (2023).

38. Kidd, C., Piantadosi, S. T. & Aslin, R. N. The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE* **7**, e36399 (2012).

39. Oudeyer, P.-Y. & Smith, L. B. How Evolution May Work Through Curiosity-Driven Developmental Process. *Top Cogn Sci* **8**, 492–502 (2016).

40. Poli, F., Serino, G., Mars, R. B. & Hunnius, S. Infants tailor their attention to maximize learning. *Sci. Adv.* **6**, eabb5053 (2020).

41. Kidd, C., Piantadosi, S. T. & Aslin, R. N. The Goldilocks Effect in Infant Auditory Attention. *Child Dev* n/a-n/a (2014) doi:10.1111/cdev.12263.

42. Tomasello, M., Carpenter, M. & Liszkowski, U. A New Look at Infant Pointing. *Child Development* **78**, 705–722 (2007).

43. Begus, K. & Southgate, V. Infant pointing serves an interrogative function: Infant pointing serves an interrogative function. *Developmental Science* **15**, 611–617 (2012).

44. Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G. & Csibra, G. Pointing as Epistemic Request: 12-month-olds Point to Receive New Information. *Infancy* **19**, 543–557 (2014).

45. Bazhydai, M., Westermann, G. & Parise, E. “I don’t know but I know who to ask”: 12□month□olds actively seek information from knowledgeable adults. *Dev Sci* **23**, (2020).

46. Goupil, L., Romand-Monnier, M. & Kouider, S. Infants ask for help when they know they don’t know. *Proc Natl Acad Sci USA* **113**, 3492–3496 (2016).

47. Stephens, D. W. & Charnov, E. L. Optimal foraging: Some simple stochastic models. *Behav Ecol Sociobiol* **10**, 251–263 (1982).

48. Robertson, S. S., Guckenheimer, J., Masnick, A. M. & Bacher, L. F. The dynamics of infant visual foraging. *Developmental Sci* **7**, 194–200 (2004).

49. Begus, K. & Bonawitz, E. The rhythm of learning: Theta oscillations as an index of active learning in infancy. *Developmental Cognitive Neuroscience* **45**, 100810 (2020).

50. Orekhova, E., Stroganova, T., Posikera, I. & Elam, M. EEG theta rhythm in infants and preschool children. *Clinical Neurophysiology* **117**, 1047–1062 (2006).

51. Wass, S. V. *et al*. Parental neural responsivity to infants' visual attention: How mature brains influence immature brains during social interaction. *PLoS Biol* **16**, e2006328 (2018).

52. Xie, W., Mallin, B. M. & Richards, J. E. Development of infant sustained attention and its relation to EEG oscillations: an EEG and cortical source analysis study. *Developmental Science* **21**, e12562 (2018).

53. Jones, E. J. H. *et al*. Infant EEG theta modulation predicts childhood intelligence. *Sci Rep* **10**, 11232 (2020).

54. Braithwaite, E. K., Jones, E. J. H., Johnson, M. H. & Holmboe, K. Dynamic modulation of frontal theta power predicts cognitive ability in infancy. *Developmental Cognitive Neuroscience* **45**, 100818 (2020).

55. Saez de Urabain, I. R., Nuthmann, A., Johnson, M. H. & Smith, T. J. Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research* **134**, 43–59 (2017).

56. Pickup, M. in *Introduction to time-series analysis* vol. 174 (Sage Publications, 2014).

57. Clifford, P., Richardson, S. & Hemon, D. Assessing the Significance of the Correlation between Two Spatial Processes. *Biometrics* **45**, 123 (1989).

58. Begus, K. & Southgate, V. Curious Learners: How Infants' Motivation to Learn Shapes and Is Shaped by Infants' Interactions with the Social World. in *Active Learning from Infancy to Childhood* (eds. Saylor, M. M. & Ganea, P. A.) 13–37 (Springer International Publishing, 2018). doi:10.1007/978-3-319-77182-3_2.

59. Kidd, C. & Hayden, B. Y. The Psychology and Neuroscience of Curiosity. *Neuron* **88**, 449–460 (2015).

60. Begus, K., Gliga, T. & Southgate, V. Infants Learn What They Want to Learn: Responding to Infant Pointing Leads to Superior Learning. *PLoS ONE* **9**, e108817 (2014).

61. Phillips, E. *et al. Proactive or reactive? Neural oscillatory insight into the leader-follower dynamics of early infant-caregiver interaction*. https://osf.io/cg38a (2021) doi:10.31234/osf.io/cg38a.

62. Meyer, M., Schaik, J. E., Poli, F. & Hunnius, S. How infant directed actions enhance infants' attention, learning, and exploration: Evidence from EEG and computational modeling. *Developmental Science* (2022) doi:10.1111/desc.13259.

63. Friston, K., Adams, R. A., Perrinet, L. & Breakspear, M. Perceptions as Hypotheses: Saccades as Experiments. *Front. Psychology* **3**, (2012).

64. Friston, K. J. Waves of prediction. *PLoS Biol* **17**, e3000426 (2019).

65. Smith, L. B. & Breazeal, C. The dynamic lift of developmental process. *Developmental Science* **10**, 61–68 (2007).

66. Lieven, E. Usage-based approaches to language development: Where do we go from here? *Lang. cogn.* **8**, 346–368 (2016).

67. Yu, C., Zhang, Y., Slone, L. K. & Smith, L. B. The infant's view redefines the problem of referential uncertainty in early word learning. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2107019118 (2021).

68. Marriott Haresign, I. *et al.* Automatic classification of ICA components from infant EEG using MARA. *Developmental Cognitive Neuroscience* **52**, 101024 (2021).

69. Xu, T. L., Abney, D. H. & Yu, C. Discovering Multicausality in the Development of Coordinated Behavior. 6.

70. Rayson, H., Bonaiuto, J. J., Ferrari, P. F. & Murray, L. Early maternal mirroring predicts infant motor system activation during facial expression observation. *Scientific Reports* **7**, (2017).
71. Jessen, S. Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. 10 (2019).
72. Jessen, S., Obleser, J. & Tune, S. Neural tracking in infants – An analytical tool for multisensory social processing in development. *Developmental Cognitive Neuroscience* **52**, 101034 (2021).
73. Keller, H. Universality claim of attachment theory: Children's socioemotional development across cultures. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11414–11419 (2018).
74. Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M. & Robbins, K. A. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* **9**, (2015).
75. Mullen, T. CleanLine EEGLAB plugin. (2012).
76. Debnath, R. *et al.* The Maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology* **57**, (2020).
77. Gabard-Durnam, L. J., Mendez Leal, A. S., Wilkinson, C. L. & Levin, A. R. The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data. *Front. Neurosci.* **12**, 97 (2018).
78. Georgieva, S. *et al*. Toward the Understanding of Topographical and Spectral Signatures of Infant Movement Artifacts in Naturalistic EEG. *Front. Neurosci.* **14**, 352 (2020).
79. Marriott Haresign, I. *et al*. Gaze onsets during naturalistic infant-caregiver interaction associate with 'sender' but not 'receiver' neural responses, and do not lead to changes in

inter-brain synchrony. *bioRxiv* 2022.05.27.493545 (2022) doi:10.1101/2022.05.27.493545.

80. Widmann, A. & Schröger, E. Filter Effects and Filter Artifacts in the Analysis of Electrophysiological Data. *Front. Psychology* **3**, (2012).
81. Dimigen, O. Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage* **207**, 116117 (2020).
82. Borsema, P & Weenik, D. Praat: doing phonetics by computer.
83. Hansen, J. H. L., Stauffer, A. & Xia, W. Nonlinear waveform distortion: Assessment and detection of clipping on speech data and systems. *Speech Communication* **134**, 20–31 (2021).
84. Brookshire, G. Putative rhythms in attentional switching can be explained by aperiodic temporal structure. *Nat Hum Behav* **6**, 1280–1291 (2022).
85. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* **164**, 177–190 (2007).
86. Benjamini Yoav & Hochberg Yosef. Controlling for false discovary rate: a practical and powerful approach to multiple testing. *Journal of the Royal Society. Series B (Methodological)* **57**,.

Acknowledgements

Thank you to Dean Matthews for help with data coding. Thanks to members of the UEL BabyDev Lab for comments and discussions on earlier drafts of this manuscript, and to all participating children and caregivers.

Funding

This research was funded by the Leverhulme Trust, RPG-2018-281.

Author Contributions

EAMP designed experiment; collected data; performed analyses; wrote manuscript; LG co-wrote manuscript; IMH collected data; contributed analyses; EB collected data; FAC collected data; MW collected data; NK collected data; EG collected data; SVW secured funding; co-wrote analyses.

Competing Interests

The authors declare that they have no competing interests.

Data and materials availability

Due to the personally identifiable nature of our data (video and audio recordings of infants), permission to access the data will be given only by contacting the first author, EAMP, direct via email.

Supplementary Materials

Supplementary Text

Materials and Methods

1.1 Clipping identification algorithm

The clipping algorithm was based on that outlined by [1]. First, points in the speech signal reaching the maximum or minimum amplitude were identified. Next, to identify whether each max/min value was the beginning of a clipping event, the algorithm detected whether the value next to this point was 99.5% +/- of the max/min. A clipping event was considered to have ended where 3 consecutive values below/above the 99.5% threshold occurred. All vocalisations involving any clipping were excluded from analyses.

1.2 *Caregiver vocalisation durations*

The length of each caregiver vocalisation was computed in seconds and inserted into the video-frame time series for the duration of that vocalisation. Periods where the caregiver was not vocalising (i.e. vocal pauses) were set to missing data points. Times where co-vocalisations occurred were also set to missing data points.

1.3 Caregiver amplitude modulations

Amplitude modulations in the caregivers' speech were extracted using the NSL toolbox [2]. First, the speech signal was downsampled to 16kHz. The 128-channel auditory spectrogram, with centre frequencies ranging from 180-7246Hz was then computed (frame length=5ms, time constant=8ms, no nonlinear filtering), and the band-specific envelopes summed across frequencies to obtain the broadband envelope of the speech signal. The amplitude envelope was inserted into the continuous variable only where the coder had identified the caregiver as vocalising: all vocal pauses were treated as missing data points. Clipped vocalisations were also identified using the same method described above, and set as missing values. Finally, the continuous amplitude variable was synchronised to the video frames.

Figures

Results

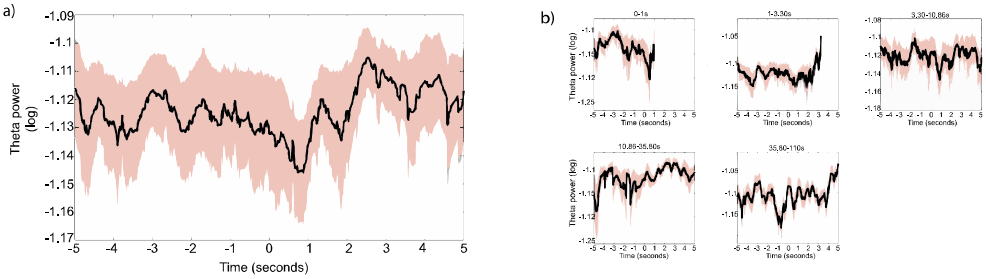

Fig. S1. Event-related analysis for relative infant theta activity. a) *Event-related analysis showing change in infant theta activity around infant attention onsets to objects: black line shows average infant theta activity (log); coloured shaded areas indicate the SEM. Black horizontal line shows areas of significance revealed by the cluster-based permutation analysis ($p < 0.05$). Cluster-based permutation analysis revealed no significant clusters of time points (all $p < 0.05$). b) event-related analysis split by infant object attention-duration time bins. Black lines shows average infant theta activity (log); shaded areas indicate the SEM. Black horizontal lines shows areas of significance revealed by the cluster-based permutation analysis ($p < 0.05$). Permutation analysis again revealed no significant clusters of time points (all $p < 0.05$).*

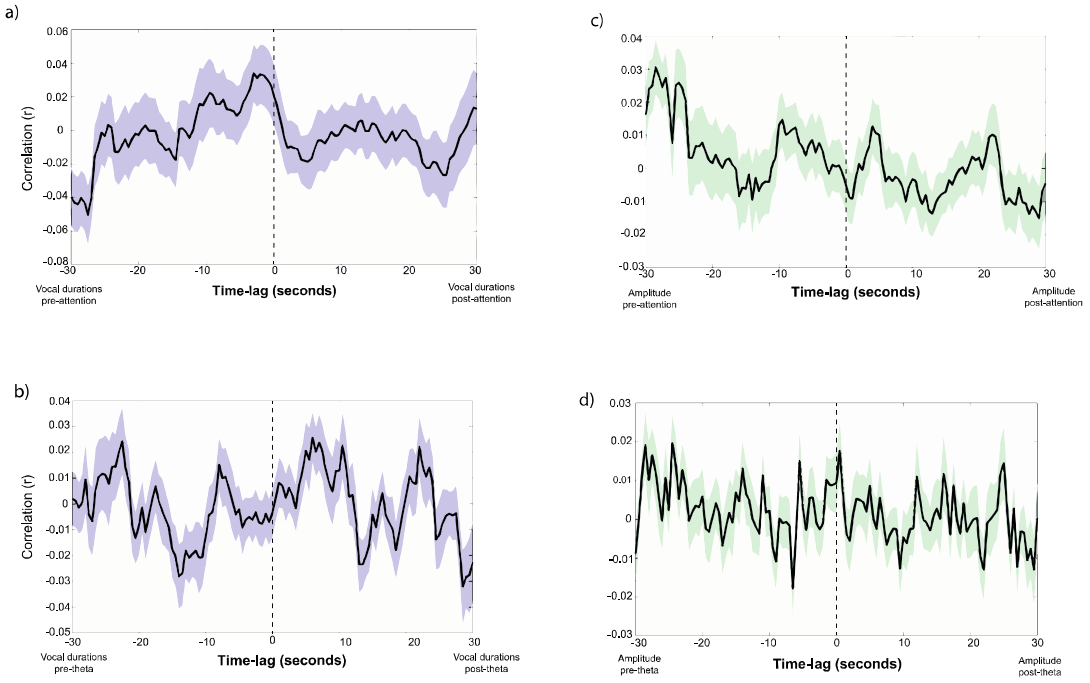

Fig. S2: Assessing forwards-predictive associations between caregiver vocal behaviour, and infant attention, and endogenous oscillatory activity. *Black lines show the Pearson cross-correlation between two variables; shaded areas indicate the SEM. Black horizontal lines show significant clusters of time lags (*p<0.05). First column shows the association between caregiver vocal durations and a) infant attention durations, and b) infant theta activity. Second column shows the association between caregiver amplitude modulations and c) infant attention durations and d) infant theta activity.*

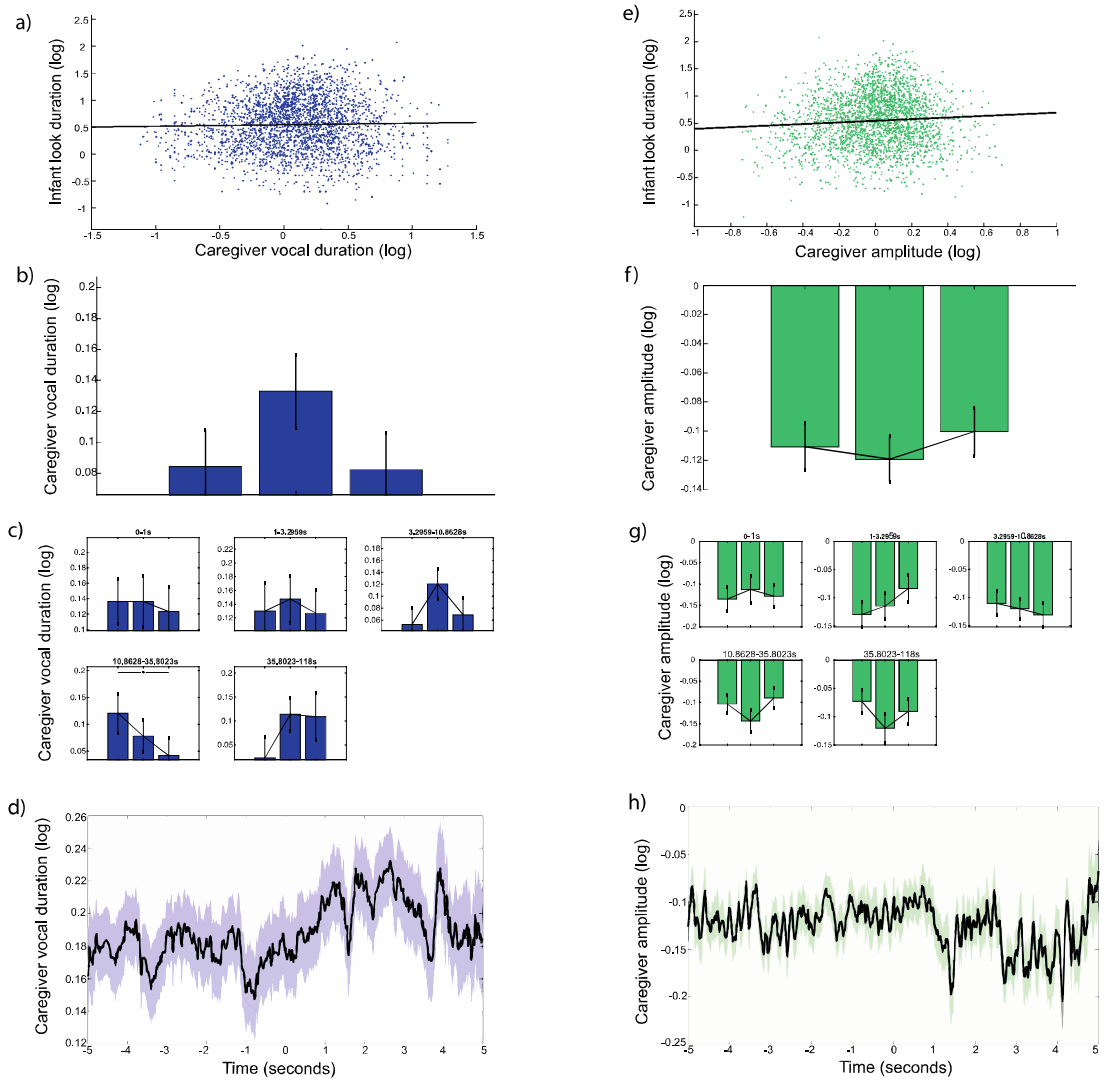

Fig. S3 Reactive change in caregiver vocal behaviour relative to infant attention onsets. *First row: scatter plots of the association between infant attention duration and a) caregiver vocal durations, d) caregiver amplitude modulations. Coloured dots show each infant object look; black line indicates linear line of best fit. Second row: event related analysis to all infant object looks for b) caregiver vocal durations, e) caregiver amplitude modulations. Black line indicates the average across participants; coloured shaded area indicates SEM. Third row: modulation analysis for each infant object look for c) caregiver vocal durations, g) caregiver amplitude modulations. Each bar shows the median for each chunk across participants; errors bars show the SEM. Wilcoxon signed ranks tests explored significant differences between attention chunks (*p < 0.05). Fourth row: same as third row, binned by infant attention durations, for d) caregiver vocal durations, h) rate of change in caregiver F0, l) caregiver amplitude modulations. Bottom row: event-related analysis showing change in caregiver vocal behaviour around infant attention onsets to objects: black line shows average across participants; coloured shaded areas indicate the SEM. Black horizontal lines*

shows areas of significance revealed by the cluster-based permutation analysis ($p < 0.05$). d) caregiver vocal durations, h) caregiver amplitude modulations.

Materials and Methods

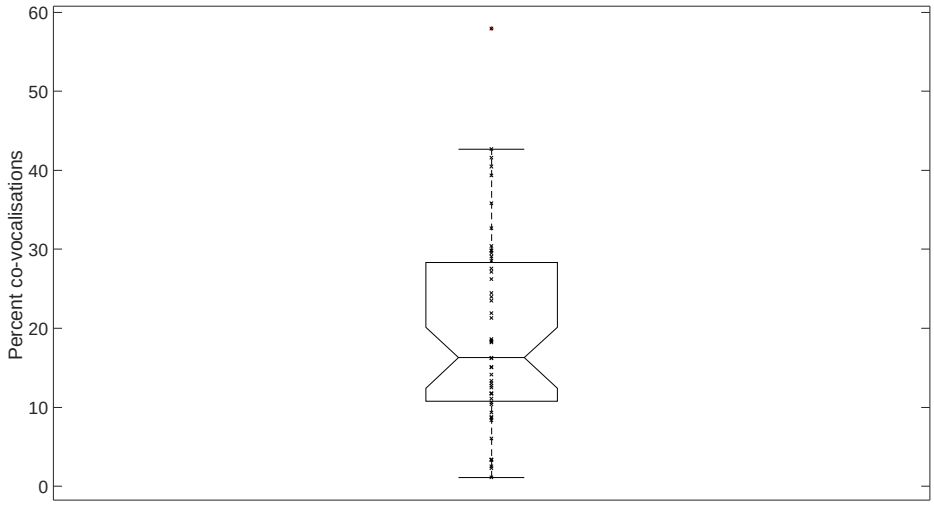

Fig. S4. Percentage of caregiver vocalisations that were co-vocalisations. *Box plot showing the percentage of caregiver vocalisations that were vocalisations across participants.*

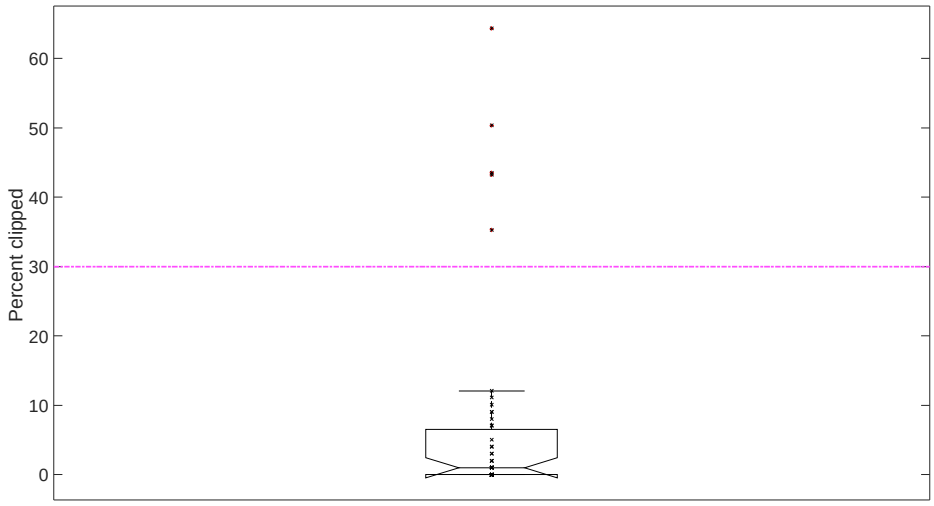

Fig S5. Percentage of clipped vocalisations. *The box plot shows the percentage of clipped vocalisations, across participants. The pink horizontal line indicates the threshold at which participants were excluded from analyses (30%).*

References

1. Hansen, J. H. L., Stauffer, A. & Xia, W. Nonlinear waveform distortion: Assessment and detection of clipping on speech data and systems. *Speech Communication* **134**, 20–31 (2021).
2. Chi, T., Ru, P. & Shamma, S. A. Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America* **118**, 887–906 (2005).