Infants' intentionally communicative vocalisations elicit responses from caregivers and are the best predictors of the transition to language: a longitudinal investigation of infants' vocalisations, gestures, and word production.

In press at Developmental Science (accepted April 2019)

Authors:

Ed Donnellan (University of Sheffield)

Colin Bannard (University of Liverpool)

Michelle L. McGillion (University of Warwick)

Katie E. Slocombe (University of York)

Danielle Matthews (University of Sheffield)

Acknowledgements

We would like to thanks Isobel Dunnett-Orridge for her assistance with coding. This research was supported by a Faculty of Science PhD Scholarship from the University of Sheffield, awarded to Ed Donnellan. The original collection of the longitudinal dataset was supported by British Academy grant SG101641 and Nuffield Foundation grant EDU40447.

Conflict of Interest Statement

The authors confirm that we have no conflict of interest with the subject matter or materials discussed in the manuscript.

Infants' intentionally communicative vocalisations elicit responses from caregivers and are the best predictors of the transition to language: a longitudinal investigation of infants' vocalisations, gestures, and word production.

Research Highlights

- Infants' vocalisations and gestures are coordinated with gaze above chance at 11 months of age, suggesting that infants are intentionally communicating.
- When infants coordinate gaze to their caregiver's face while vocalising or gesturing, caregivers are more likely to respond.
- A multi-model inference procedure established which infant vocal and gestural behaviours best predict language outcomes and whether gaze-coordination and caregiver responses increase predictive value.
- Infants' gaze-coordinated vocalisations that were met with a timely and contingent caregiver response were the best predictor of expressive language development up to 2 years.

Abstract

What aspects of infants' prelinguistic communication are most valuable for learning to speak, and why? We test whether early vocalisations and gestures drive the transition to word use because, in addition to indicating motoric readiness, they 1) are early instances of intentional communication and 2) elicit verbal responses from caregivers. In study 1, 11-month-olds (*N* = 134) were observed to coordinate vocalisations and gestures with gaze to their caregiver's face at above chance rates, indicating that they are plausibly intentionally communicative. Study 2 tested whether those infant communicative acts that were gaze-coordinated best predicted later expressive vocabulary. We report a novel procedure for predicting vocabulary via multi-model inference over a comprehensive set of infant behaviours produced at 11- and 12-months (*n* = 58). This makes it possible to establish the relative predictive value of different behaviours that are hierarchically organised by level of granularity. Gaze-coordinated vocalisations were the most valuable predictors of expressive vocabulary size up to 24 months. Study 3 established that caregivers were more likely to respond to gaze-coordinated behaviours. Moreover, the dyadic combination of infant gaze-coordinated vocalisation and caregiver response was by far the best predictor of later vocabulary size. We conclude that practice with prelinguistic intentional communication facilitates the leap to symbol use. Learning is optimised when caregivers respond to intentional vocalisations with appropriate language.

Key words: Lexicon; social communication; parenting; infancy; learning.

Within two years of birth, infants begin to use words to direct others' attention and share information. Precisely which cognitive and social mechanisms allow them to make this transition to language is not yet understood. One well-documented prerequisite is the development of specific motor skills required for speech, such as the ability to produce syllables (Vihman, Macken, Miller, Simmons, & Miller, 1985). More recently however, attention has been focused on two additional factors: a) infants' developing intentional control over communication (e.g., Tomasello, 2008), and b) caregiver responses to infants' prelinguistic communicative acts (e.g., Bornstein, Tamis-LeMonda, & Haynes, 1999; Rollins, 2003; Wu & Gros-Louis, 2014). Evidence suggests that both of these factors contribute to the transition to speech. However a major challenge in understanding the unique contribution that each factor makes and how they interact to drive development is that, to date, studies have looked at the diverse behaviours involved in separate studies, often using different methodologies. Recent work investigating early intentional communication has focused almost exclusively on infants' gestural communication (and primarily index-finger pointing), while work on caregiver responsiveness has focused primarily on responses to infants' vocal behaviours. In this paper, we make a first move towards a unified account of the emergence of conventional communication, using novel analytic techniques to consider a comprehensive set of early infant behaviours in a new longitudinal dataset.

The development of intentional communication

From as early as 5 months, infants expect their vocalisations to influence their caregiver's behaviour (Goldstein, Schwade, & Bornstein, 2009; see also Wu & Gros-Louis 2017). By the time infants are 12 months old, adults interpret their infant's vocalisations as requests or expressions of discontent (Esteve-Gibert & Prieto, 2013; Oller et al., 2013; Papaeliou, Minadakis, & Cavouras, 2002; Papaeliou & Trevarthen, 2006) and adults redirect their attention following infants' gestures (Cameron-Faulkner, Theakston, Lieven, & Tomasello, 2015; Carpenter, Nagell, & Tomasello, 1998). However, it is not clear when infants start to *intend* for these vocalisations and gestures to affect others' attention. A key theoretical challenge is therefore how to determine whether an infant's behaviour is intentionally communicative in order to then test whether intentional communication specifically predicts the transition to speech. One approach to this challenge is to appeal to the following criteria for intentional action set out by Bruner (1973, 1975):

'Intention, viewed behaviourally, has several measurable features: anticipation of the outcome of an act, selection among appropriate means for achievement of an end state, sustained direction of behaviour during deployment of means, a stop order defined by an end state, and finally some form of substitution rule whereby alternative means can be deployed for correction of deviation or to fit idiosyncratic conditions' (Bruner, 1973, p. 2).

These Brunerian criteria were initially applied in studies of prelinguistic infants (Bates, Camaioni, & Volterra, 1975). Recently, however they have been applied less frequently with human infants (although see Golinkoff, 1986; Liszkowski, Carpenter, & Tomasello, 2007), and more frequently in the study of intentional communication in non-human primates (Cartmill & Byrne, 2010; Leavens, Russell, & Hopkins, 2005; Pika, Liebal, & Tomasello, 2003). This may be because of the difficulty of observing naturalistic frustration episodes as infants rarely have to persist or elaborate (since caregivers are highly responsive in free play, e.g., Baumwell, Tamis-LeMonda, & Bornstein, 1997) or because of the difficulty, and potential circularity, of inferring what the end state/goal of an infant's behaviour was. However, one indicator that infants anticipate an outcome and are selecting appropriate means to communicate is their use of gaze-checking (i.e., looking to their caregiver's eyes) while vocalising or gesturing. This gaze-coordination has been used as a marker of communicative intention with both human infants, and non-human animals (e.g., Bates et al., 1975; Franco & Butterworth, 1996; Gros-Louis & Wu, 2012; Harding & Golinkoff, 1979; Schel, Townsend, Machanda, Zuberbühler, & Slocombe, 2013; Tomasello et al., 1997). While not a *necessary* condition for intentional communication (Olson & Masur, 2013), it is arguably one of the best markers available.

It is worth noting that, while many developmental psychologists have tended to assume that gaze-coordination indicates that infants intend to affect the attentional state of the caregiver through their communicative acts (either to direct attention to themselves or to initiate joint attention to some third entity), some primatologists have been more cautious, differentiating sub-types of intentional communication (e.g., Townsend et al., 2017). They have assumed that gaze-coordination is evidence only of what Dennett (1983) classified as first-order intentionality (where zero-order intentionality would cover involuntary/reflexive acts, first-order would cover communicative acts produced to affect another's behaviour without reference to their mental states, and second order would cover acts that are produced to affect the mental state of another). These levels likely reflect points on a continuum that infants ascend over the first year of life and it is no simple matter to distinguish between them purely on behavioural evidence. For the purposes of the current studies, we assume that at a minimum, gaze-coordination can be taken as evidence that infants have begun to gain intentional control over communication for the purpose of interacting with others, and at a maximum they have begun to understand that this works by attention-directing specifically. Either way, gaze-coordinated vocalisations and gestures can be taken as more socio-cognitively advanced acts than those produced without gaze-coordination (which may be entirely unintentional). The question we test here is whether the frequency with which infants engage in intentional (gaze-coordinated) prelingustic communication is predictive of their transition to linguistic communication. On social-pragmatic accounts of word learning (e.g., Tomasello, 2003, 2008) such intentional prelinguistic communication is a prerequisite for the use of intersubjectively shared symbols. Through prelinguistic communication, infants learn to share information about the world (e.g., referring, requesting and commenting). Having mastered this, infants arrive at a jumping-off point for word use because the next step is for conventional symbols to replace these early prelinguistic acts as tools for communication.

The role of caregiver responses

Caregiver responses to infants' prelinguistic behaviours are thought to scaffold the transition to language (Bruner, 1976). It has been demonstrated that caregivers' responses to infants' vocalisations and gestures predict infants' later language (e.g., McGillion et al., 2013; Olson & Masur, 2015). However the role of intentional communication in eliciting caregiver responses, and how this in turn predicts later language, has not yet been investigated. Responses to intentional communication may be especially valuable as they provide helpful linguistic information precisely at a moment when infants are motivated to communicate about whatever it is they are attending to. Indeed, it has been hypothesised that infants' use of gestures with vocalisations signal a readiness for verbal input to caregivers, and that timely responses that 'translate' infants' gesture-vocal combinations into adult speech may especially facilitate language learning (Iverson & Goldin-Meadow, 2005). Alternatively, it could be that caregivers' responses to vocalisations and gestures shape these behaviours and promote early conventional language regardless of their infant's communicative intent.

The current studies

The studies in the current paper aim to elucidate the developmental mechanisms that allow the transition to word production. Specifically, we explore a comprehensive range of infant vocalisations and gestures in the same dataset and test whether intentional communication (operationalized as gaze-coordinated gestures/vocalisations) from 11 months is especially predictive of word learning and whether caregiver responses to intentional communication further boost learning. To this end, our investigation consists of three stages. We first explore at the group level whether we can attribute communicative intentionality to any of the vocal and gestural behaviours infants produce at 11 months (study 1). We then test whether individual differences in infants' intentionally communicative vocalisations or gestures are predictive of the transition to language (study 2), and finally we test whether any predictive links hold because caregivers respond to early communicative attempts with relevant speech (study 3).

Study 1

Previous studies have claimed that prelinguistic vocalisations and gestures are intentionally produced because they are gaze-coordinated (Gros-Louis & Wu, 2012; Harding & Golinkoff, 1979; Maljaars, Noens, Jansen, Scholte, & van Berckelaer-Onnes, 2011; Wu & Gros-Louis, 2014). However, these studies do not control for the possibility that the co-occurrence of these behaviours happens by chance. To our knowledge, no such controls have been used in studies of infant gestures. In one of the rare prior studies to provide such controls, D'Odorico & Cassibba (1995; see also D'Odorico, Cassibba, & Salerni, 1997) found tentative evidence that vocalisations are intentionally communicative prior to the end of the first year of life. In an experimental paradigm, 10-month-olds (N = 8) looked to their caregivers' faces prior to, or during vocalising more than would have been expected by chance, whereas 4-, 6- and 8-month-olds did not. There are limits on what can be concluded from this study for our purposes since the sample size was small (and findings possibly depended on two precocious infants), the granularity for coding temporal overlaps between gaze and vocalisations was not fine-grained (due to the available technology), there was no distinction made between Consonant-Vowel (CV) and non-CV vocalisations, and gestures were not analysed. Nonetheless, it suggests that vocalisations are gaze-coordinated at above chance levels and might be intentionally communicative before the end of the first year of life. The question for this study is therefore which of *all* the gestural and vocal behaviours available to infants at 11 months are gaze-coordinated at above chance levels at the group level?

In the current study, we considered infant behaviour at home with the primary caregiver to give an ecologically valid picture of early communication. All major infant gesture types were coded including index-finger pointing, open-hand pointing, showing, giving and conventional gestures. Likewise, all non-vegetative vocalisations were coded and we distinguished between speech-like, CV vocalisations and non-CV vocalisations. We consider gesture-vocal combinations as a separate, mutually exclusive category from gestures and vocalisations produced alone. This is firstly because such combinations are arguably qualitatively different, in that vocalisations produced in combination with gestures have different acoustic properties (Murillo & Capilla, 2016). Secondly, doing so better allows us to tease apart whether the vocal or gestural component, or the unique combination of both represents an early attempt at intentional communication.

Methods

Participants

Video recordings of 134 11-month-old infants (70 female infants, 64 male; mean age = 334 days, *SD* = 4 days) were coded. Recordings were drawn from a larger sample (*N* = 140) that had been collected as baseline measures for a longitudinal randomised controlled trial (McGillion, Pine, Herbert, & Matthews, 2017). All caregivers gave informed consent for their videos to be used for further research. Two families from this larger pool were excluded because there was a third individual present, and a further 4 dyads were excluded for being in shot for less than 7 minutes. Ethical approval was granted by the Psychology Ethics sub-committee at the University of Sheffield. The data is not publicly available due to ethical restrictions.

Procedure

Participants were filmed in their home from two camera angles in an unstructured play session with their primary caregiver lasting 10-15 minutes (McGillion, Pine, et al., 2017). Only the infant and caregiver were present for the duration of the video (the researcher who set up the cameras having left the room). Coding of the naturalistic videos was undertaken in ELAN (Sloetjes & Wittenburg, 2008). All videos were coded by the first author.

Coding

From the videoed session, 10 minutes were selected for coding. This period began from the moment the researcher left the room until 10 minutes later (excluding time off-shot), or until the experimenter returned (if this was prior to 10 minutes being reached). For 9 participants, observation time was below 10 minutes (but above 7 minutes), so prorated frequencies of behaviours were used throughout the analyses.

We coded infant behaviours as either a *vocalisation*, a *gesture* or a *gesture-vocal combination* (for a detailed coding scheme, please see Appendix A). Vocalisations were sub-categorised as *CV vocalisations* (i.e., Consonant Vowel vocalisations) or *non-CV vocalisations*. Gestures were sub-categorised as either *index-finger pointing*, *open-hand pointing*, *giving*, *showing* or *conventional gestures*. Gesture-vocal combinations (where a vocalisation and gesture overlapped in time), were sub-categorized as either *involving a CV vocalisation* or *only non-CV vocalisations* and as involving one of the five gesture types. This gave us 10 types of gesture-vocal combination. The full set of behaviours and their relationship to one another can be seen in figure 1.

Any gesture, vocalisation or gesture-vocal combination was considered to be *gaze-coordinated* if the infant looked to the caregiver within one second of producing the behaviour (see also Desrochers, Morissette, & Ricard, 1995; Matthews, Behne, Lieven, & Tomasello, 2012; Murillo & Belinchón, 2012). Periods when infants were off-shot and when we were unable to determine if they were gazing to a caregiver's face or producing a gesture were marked and excluded from analysis (see Appendix A).

For the analysis relating to whether vocalisations, gestures or combinations were gaze-coordinated above chance rates at the group level, first we calculated the *expected* (chance) rate of gaze-coordination for each type of behaviour. To take the example of vocalisations, we calculated the time each infant spent vocalising during the observation period and the time they spent gazing to the caregiver and then multiplied these to obtain the expected rate of co-occurrence (see also Bakeman & Gottman, 1986, pp. 131–132). A slight modification to this procedure was necessary due to the fact that we counted gaze to the caregiver's face as co-occurring with a vocalisation if it happened within a one-second window of the vocalisation (i.e., from one second before the onset of the vocalisation to one second after the offset of the vocalisation). Thus the time spent vocalising was taken to be the time spent vocalising plus one second before and one second after each vocalisation. When vocalisations were given in quick succession however, these one-second windows sometimes overlapped with another vocalisation, or the one-second window of another vocalisation. It was therefore important not to double-count this overlapping time. In these cases, we counted the intervening time between the behaviours only once. We then calculated the *observed* rate of gaze-coordination for each type of behaviour. To do this, we first identified the time spent engaging in the target behaviour (e.g., vocalisations), and then added the one-second time window to either end of the behaviour (applying the same procedural modification described above to avoid double-counting). We then extracted the duration of gazes to the caregiver's face that occurred within these windows where the target behaviour (e.g., vocalisations) happened.

Reliabilities

Ten percent of videos (randomly selected) were double-coded by a trained research assistant. Reliabilities were calculated on all coding (i.e., identifying and classifying behaviours, and determining rates of gaze-coordination) and revealed excellent rates of agreement in all cases (all κ >.75, r > .80; see Appendix B).

Results

All 11-month-olds gazed to their caregiver's face and produced non-CV vocalisations. Ninety-seven percent (n = 130) produced at least one CV vocalisation. Fewer infants produced gestures (either alone or in a gesture-vocal combination), with 67% (n = 90) producing one or more gestures. Most commonly, infants produced give gestures (produced by 36% of infants, n = 48), but a number also produced show gestures (22%, n = 30), index-finger pointing (21%, n = 28), open-hand pointing (20%, n = 27) and conventional gestures (19%, n = 25). Forty percent of infants (n = 53) produced gesture-vocal combinations (see Appendix C, table 4 for full descriptive statistics).

Figure 2 shows the expected and observed co-occurrence of gaze to the caregivers' face with infant vocalisations, gestures and combinations, with paired t-tests and Bayes Factors for each comparison reported in Table 1. Each Bayes Factor is the ratio of the probability of the hypothesis (that the observed durations are greater than the expected durations) and the null hypothesis (that the durations are the same). According to Kass & Raftery (1995), a Bayes Factor of 1-3 is 'not worth more than a bare mention', 3-20 indicates positive support for the hypothesis over the null, 20-150 is strong support, and >150 is very strong support. Bayes Factors of less than 1 indicate support for the null hypothesis (with values <1/3 indicating positive evidence and so on). We report Bayes Factors in place of p-values because they provide us with a measure of strength of evidence rather than a reject/not reject judgement, allowing us to a) to compare across behaviours, and b) take information from these first tests forward as priors to inform our tests of the subordinate behaviours. We obtain Bayes Factors using the "BayesFactor" package for R (Morey & Rouder, 2015), and for these initial analyses we use a default "medium" prior on the effect size. These analyses revealed that gestures and combinations co-occurred with gaze above the level predicted by chance, with Bayes Factors indicating very strong support for this finding. Vocalisations co-occurred with gaze above the level of chance, but Bayes Factors indicated that such evidence provides only anecdotal support for this finding over the null hypothesis (that they occur at chance levels).

Figure 3 shows the expected and observed co-occurrence of gaze to the caregiver's face with infants' individual vocalisation and gesture subtypes (produced alone and as part of combinations), with comparisons reported in Table 2. As the behaviours considered in these tests are subtypes of those considered in the tests reported in Table 1, the prior on the effect size for each behaviour (specifically the Cauchy scale parameter) was set to reflect the effect sizes (d) observed for its superordinate behaviour in the analyses reported in Table 1; for vocalisations, d = 0.204, for gestures, d = 0.419 and for combinations, d = 0.373.

Closer inspection of the vocalisation sub-types revealed that while non-CV vocalisations co-occurred with gaze above chance levels (Bayes Factors indicated positive support for their being an above-chance association), CV vocalisations did not (Bayes Factors indicated support for the null hypothesis, i.e., chance co-occurrence). Closer inspection of the gesture sub-types revealed strong evidence that showing co-occurred with gaze above chance levels and positive evidence that giving co-occurred with gaze above chance levels. Conventional gestures only anecdotally co-occurred with gaze above chance levels. Neither index-finger pointing nor open-hand pointing co-occurred with gaze above chance levels, with Bayes Factors indicating support for the null hypothesis. Regarding combinations, those involving both types of vocalisations co-occurred with gaze above chance levels, and when separated by gesture type, the picture was broadly the same as when gestures were considered alone (with stronger support for the hypothesis in the case of combinations involving conventional gestures than conventional gestures produced alone, and weaker support in the case of combinations involving giving or showing than these gestures produced alone).

Discussion

This study established that, for many vocal and gestural behaviours, 11-month-old infants coordinate gaze to their caregiver's face above rates that would be expected by chance. This is consistent with the claim that infants are attempting to intentionally communicate. It is important to note, however, that in this analysis two theoretically important behaviours – CV vocalisations and index-finger pointing - were not coordinated above chance at the group level. This is not evidence that these behaviours are never produced with communicative intent, but it might suggest that they are not always produced in this way. For example, while CV vocalisations were very frequently gaze-coordinated, there are many non-gaze-coordinated instances, and these might be characterised as non-communicative 'vocal play' (Bates & Dick, 2002; Oller, 2000). Likewise, while pointing was often gaze-coordinated (43% of points were gaze-coordinated), this was not at above chance rates. Infants at this age point in the absence of others (Bates et al., 1975, p. 217; Carpendale & Carpendale, 2010; Delgado, Gómez, & Sarriá, 2011), suggesting that pointing is not always necessarily communicative. It is possible that infants' interspersion of intentionally communicative and non-communicative CV vocalisations and pointing yielded chance levels of coordination overall.

It should be noted that pointing is perhaps not comparable to other gestures considered here in terms of the ease with which infants can coordinate gaze to caregivers. Giving and showing gestures (both coordinated with gaze at above chance levels) are adult-directed, and the physical configuration of showing in particular (holding objects up to caregiver's face) likely facilitates infants' looking to their caregivers' face more readily than with pointing to other entities. Likewise, giving and showing gestures involve objects within the infant's grasp, while pointing (especially in relation to more distal stimuli) is thought to be more cognitively complex, perhaps accounting for greater difficulty in gaze-coordination at this age (see also Boundy, Cameron-Faulkner, & Theakston, 2019; Carpenter, Nagell, & Tomasello, 1998).

In study 2 we look at whether the frequency with which children produce gaze-coordinated instances of behaviours is more predictive of their language development than the frequency with which they produce those behaviours regardless of gaze-coordination. In other words, we explore whether specifically those instances that we assume to be attempts to intentionally communicate (e.g., rather than babble or undirected gestures) are predictive of language outcomes.

Study 2

Infant's prelinguistic vocalisations and gestures predict later language abilities (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Igualada, Bosch, & Prieto, 2015; Laakso, Poikkeus, Katajamaki, & Lyytinen, 1999; McCune & Vihman, 2001; McGillion, Herbert, et al., 2017). The question addressed in this study is whether this is because they represent instances of prelinguistic intentional communication. In the case of vocalisation-vocabulary links, CV vocalisations are a motoric prerequisite of speech, and thus might predict later language because they indicate motoric readiness for speech, not because they represent early practice with intentional communication. In contrast, there is an assumption that gesture-vocabulary links, in particular pointing-vocabulary links, exist because such gestures are early attempts to intentionally communicate (Tomasello, 2003, 2008). However, it remains an untested empirical question as to whether this is the case. The premise of this study is that if prelinguistic communicative behaviours are predictive of word use because they are early instances of intentional communication then we should expect that measures of gaze-coordinated behaviours specifically should be the best predictors of expressive vocabulary.

One limitation of previous investigations into which prelinguistic behaviours predict first words is that often just one predictor (e.g., pointing) is considered. A pitfall of this approach is potentially hidden correlations with other unmeasured behaviours. This makes drawing conclusions about the relative predictive power of specific behaviours problematic. We therefore need to consider a more complete set of infant behaviours and then test the predictive value of each behaviour alone or in combination with others. Taking this approach results in a large set of possible models and high potential for collinearity, giving rise to significant model uncertainty, i.e., it is possible that there are many ways in which these predictors might explain the data and that these accounts might be difficult to distinguish from one another. We confront these issues using a multi-model inference procedure (Burnham & Anderson, 2002) rather than a traditional single-model approach. The specific approach we take is bootstrap smoothing. This involves fitting the space of plausible models to a large set of simulated datasets generated by resampling from our data. The explanatory value of a given predictor is then taken to be the *proportion of simulated datasets for which that predictor was included in the best fitting model*. The slope estimate for that predictor is calculated by averaging over the best models for all simulated datasets.

In this study, we first assess whether the frequency of infants' vocalisations, gestures and gesture-vocal combinations produced regardless of gaze-coordination at 11 and 12 months predict children's later expressive vocabulary at 15, 18 and 24 months using this modelling approach. Thus we initially take the necessary step of providing a more complete picture of what behaviours (regardless of gaze-coordination) predict later language than is currently available in the literature. Crucially, we then investigate whether gaze-coordinated instances of these behaviours are better predictors of vocabulary. Thus, we investigate whether specifically intentionally communicative instances of each of these behaviours are better predictors of later language.

Method

Participants

For this study, we restricted the sample from study 1 to the 70 dyads who participated in the control condition of the original randomised controlled trial (McGillion, Pine, et al., 2017) since those in the experimental condition had received a parenting intervention (after the 11 month videos were recorded) aimed at promoting language development (making it difficult to study growth over time without taking potential effects of the intervention into account). We included 58 infants (33 female, 25 male) for whom we had naturalistic observations at both 11 and 12 months and a measure of expressive vocabulary at 15, 18 and/or 24 months (see Appendix D for detailed breakdown).

Materials

Expressive vocabulary was assessed using the Lincoln Communicative Development Inventory (LCDI) Infant Form at 15 and 18 months and the Toddler Form at 24 months (see McGillion, Pine, et al., 2017 for full details of data collection procedure; Meints & Woodward, 2011).

Procedure

Infant behaviours at 12 months were coded following the method described in study 1. Half of the videos at 12 months were coded by the first author, with the remaining half coded by a trained research assistant. Reliabilities were calculated for all behaviours coded at 12 months following the method described in study 1. Excellent rates of agreement were reached in all cases (all $\kappa > .75$, $r > .90$; see Appendix B). Data from the 11- and 12-month time points were collapsed to maximise variance in the frequency of observed gestures during infancy (since gestures were produced relatively infrequently at 11 months alone).

Analysis

To address the question of which prelinguistic behaviours predicted language outcomes, we built mixed effects Poisson regression models with expressive vocabulary as the outcome variable (measured at 15, 18 and 24 months) and participant as a random effect on the intercept. Age was included as a predictor in all models and all predictor variables were mean centred and scaled to units of their standard deviation to aid interpretation. All combinations of behaviours (within constraints noted below) were compared for fit. Note that as behaviours were coded at different levels of granularity (e.g., the behaviour *vocalisation* had two sub-types: CV and non-CV), models could only include predictors that were not a subset of the data points of another (see figure 1). For example, we did not construct models with *frequency of vocalisations* and *frequency of CV vocalisations* as predictors in the same model as one predictor represents a subset of the data points of the other. To accommodate the fact that some of the models under consideration were large given the sample size (requiring the use of a second-order information criterion, Sugiura, 1978), and overdispersed, we use QAICc (Burnham & Anderson, 2002 section 2.5) to estimate the fit of each model, with the dispersion parameter taken from the largest model in our candidate set (a model including all lowest level predictors, i.e., the most fine-grained behaviour subtypes - those furthest right in figure 1).

As noted above, for many datasets there is no single correct model, particularly when the hypothesis space is large, and that model uncertainty needs to be taken into consideration (Buckland, Burnham, & Augustin, 1997). In order to accommodate this, rather than selecting a single best model, we performed bootstrap smoothing (a kind of model averaging; Burnham and Anderson, 2002) to determine the value of each predictor and to give a more robust estimate of its effect. This procedure involved using sampling with replacement from our original dataset to produce 10,000 new datasets. We ranked all models fitted to the original dataset by QAICc (lowest values to highest – lower values indicating a better fit). Models that were within 2 QAICc units of the best fitting model were considered candidate models to be tested against the resampled datasets. We selected the best model for each dataset from this set of candidate models. Having selected the models that gave the best fit to each new dataset, we then a) used the proportion of the datasets for which each predictor was included in the best fitting model as an estimate of its predictive value (its *inclusion probability*) and b) used the coefficients from the 10,000 models to calculate means and confidence intervals for each predictor. The inclusion probability for each predictor can be taken as a measure of relative predictive value. It is important to note that since the estimates are based on model averaging, it cannot be assumed that predictors would have the same estimates if they were all included in a single model together.

Results

First, we attempt to replicate and extend previous findings concerning which infant vocal and gestural behaviours predict productive vocabulary size. We consider a full range of behaviours and ask which predict later language. The inclusion probabilities (i.e., the measure of predictive value) for all of these vocal and gestural predictors can be seen in figure 4. For ease of visualisation, all predictors that had an inclusion probability of >0.1 (10%) are shown in bold, and can be considered the most valuable predictors of later expressive vocabulary. Figure 5 then plots the effect sizes and direction of effects for these frequently included predictors (together with 90% confidence intervals). As all predictors were scaled, the X-axis of figure 5 represents effect sizes in terms of the change in number of words we would predict a child to be able to produce (at 19 months, the mean age for this model) given one standard deviation increase in the given behaviour at 11 and 12 months. This change in number of words can be both positive and negative. For example, a child whose index-finger pointing frequency around their first birthday is one standard deviation higher than the mean, is predicted to produce 20 extra words at 19 months. By contrast, a child who produces non-CV vocalisations whilst open-hand pointing at a frequency one standard deviation higher than the mean, is predicted to produce over 20 fewer words than the mean at 19 months.

The next critical step was to test whether prelinguistic behaviours are predictive because they are used in an intentionally communicative manner. In order to do this, we reran the model fitting and bootstrap smoothing process, this time including in the model space both the overall frequency of vocalisations, gestures and combinations and their frequency of occurrence specifically when coordinated with gaze. The inclusion probability for each predictor can be seen in figure 6. The *gaze-coordinated inclusion probability* is given without parentheses and the *regardless-of-gaze inclusion probability* is given in parentheses. Critically, because the inclusion probabilities here are derived from a model space that includes both gaze-coordinated and regardless-of-gaze versions of all predictors, we can take the rates of inclusion for the two versions of each behaviour as an indicator of their relative predictive value and thereby answer the question as to whether gaze-coordinated behaviours are better predictors. All predictors that had an inclusion probability of >0.1 are shown in bold, and their means and confidence intervals shown in figure 7.

Gaze-coordinated vocalisations have the highest inclusion probability, being included in over 70% of models. A child who produced these vocalisations at a frequency of one standard deviation above the mean at 11 & 12 months, is predicted to produce 20 words more than the average child at 19 months (Figure 7). Figure 6 shows that vocalisations are the only category of prelinguistic behaviour where the gaze-coordinated frequency of the behaviour is a substantially better predictor than the behaviour considered regardless of gaze-coordination.

Finally, because combinations involving open-hand pointing with non-CV vocalisations are almost always gaze-coordinated when they appear, and this interchangeability (all except 3 children have identical counts regardless of gaze-coordination) is the cause of both the gaze-coordinated and regardless-of-gaze predictors having relatively high inclusion probabilities, we include only the slightly preferred gaze-coordinated predictor in the plot and in further discussion.

Discussion

In study 2 we first looked at the predictive value of the full range of vocal and gestural behaviours regardless of whether they were gaze-coordinated to provide a more complete picture of what behaviours predict later language than provided in the literature. An array of vocalisations, gestures and specific gesture-vocal combinations produced at 11 and 12 months predicted later expressive vocabulary. Secondly, we addressed the question of whether these behaviours predicted language development because they were early instances of intentional communication by expanding the space of possible predictors to include gaze-coordinated instances of these behaviours. This changed the picture in important ways, discussed below, which suggest that it is of crucial importance to consider whether behaviours are intentionally communicative.

The most notable change was seen with vocalisations. In the initial regardless-of-gaze analysis (figure 5), CV and non-CV vocalisations had the two highest inclusion probabilities, with the former being a positive predictor and the latter being negative. However when gaze-coordinated versions of behaviours were added to the model space, both regardless-of-gaze vocalisation subtypes had a much lower inclusion probability. Instead the *gaze-coordinated* version of the single combined vocalisation predictor appeared in the best model 74% of the time as a positive predictor. When non-CV vocalisations are considered regardless-of-gaze, they are a negative predictor of later vocabulary, yet when only gaze-coordinated (intentional) instances are considered, they are a positive predictor. Indeed gaze-coordinated non-CV vocalisations are indistinguishable from gaze-coordinated CV vocalisations in their relationship to later vocabulary as they are both positively related to vocabulary size and give the best fit when represented by a single composite predictor (i.e., gaze-coordinated vocalisations). We therefore provide evidence that the strongest predictor of infants' later language is the frequency with which they produce gaze-coordinated (and thus, we infer, intentionally communicative) vocalisations at 11 & 12 months.

Show gestures were a valuable positive predictor of language development. Since this behaviour was almost always gaze-coordinated, it is impossible to test whether specifically gaze-coordinated instances of the gesture were predictive. While it is likely that this gesture is produced with communicative intent (Boundy et al., 2019), this is hard to unpack using our data. However, we have demonstrated the link between showing and later language that has been hypothesised, but empirically tested only once, on a small sample (Bates et al., 1979). The physical configuration of showing (holding objects up to caregiver's face) allows infants' to attend to both an object of interest and the attention of their caregiver to that object, which plausibly scaffolds the transition to later triadic communication.

In contrast, open-hand pointing (both produced alone and combined with non-CV vocalisations) was a reliable *negative* predictor of language. This provides convergent evidence with recent studies that suggest that open-handed pointing is a marker for risk of delay (Luke, Grimminger, Rohlfing, Liszkowski, & Ritterfeld, 2016). Furthermore, as there was no evidence that gaze-coordination affected the negative value of this predictor, we conclude that this may be a marker of a motoric delay rather than a social-cognitive one.

Finally, while we found that index-finger pointing positively predicted later expressive vocabulary (Desrochers et al., 1995), we found a) no evidence that it was more predictive when only gaze-coordinated instances were considered, and b) that it was not a substantive predictor when the model space was expanded to include gaze-coordinated versions of all behaviours. This provides convergent evidence that index-finger pointing is not a crucial predictor in the transition to first words when other infant behaviours are also considered and when all behaviours are measured under naturalistic conditions (McGillion, Herbert, et al., 2017).

To summarise, we assumed that if prelinguistic vocal and gestural behaviours are predictive of word use because they are early instances of intentional communication then measures of gaze-coordinated behaviours specifically should be the best predictors of expressive vocabulary. This was unambiguously demonstrated to be the case for infant vocalisations, but not for their gestures. The remaining question is whether these positive predictors relate to later language purely because they indicate infants' readiness for intentional communication to become conventional communication and/or because apparently intentional acts are particularly effective in eliciting a response from caregivers.

Study 3

Gaze-coordinated prelinguistic behaviours could be more likely to provoke a linguistic response from caregivers who then provide relevant lexical material at precisely the moment when infants are most able to learn from it (see also Iverson and Goldin-Meadow 2005 for a similar discussion concerning gesture-vocal combinations). Our questions in this final study are 1) whether caregivers are indeed more likely to respond to gaze-coordinated prelinguistic behaviours, and 2) whether when they do respond, such episodes are the better predictors of language outcomes than the infant behaviours alone (while controlling for overall rates of caregiver speech). Answering these questions will offer insight into whether any of the predictive value of gaze-coordination can be attributed to the fact that caregivers are more likely to respond to gaze-coordinated behaviours.

Of particular interest for word learning are caregiver responses that are both *temporally* and *semantically* contingent on infants' vocalisations and gestures (i.e., caregivers say something in quick temporal succession of an infant behaviour that relates to the infant's focus of attention). Previous studies have established that the amount that caregivers respond in a semantically contingent manner to infant behaviour predicts later expressive vocabulary (McGillion et al., 2013; Olson & Masur, 2015). However, prior studies have not considered responses to a comprehensive range of vocalisations and gestures or taken into account infant intentions.

As previously, we treated gesture-vocal combinations as a separate category. Caregivers may respond differently to combinations (compared to gestures/vocalisations alone) since they may more reliably infer what infants are trying to communicate about when cues from both modalities are available (Balog & Brentari, 2008; Fasolo & D'Odorico, 2012; Grünloh & Liszkowski, 2015; Rowe & Goldin-Meadow, 2009).

Method

Participants

The same 58 infants from study 2 were included in this study, along with their primary caregivers. All caregivers were female, spoke English to their children and were from socio-economically diverse backgrounds: 66% had a university degree, and 12% lived in areas considered to be within the most deprived 10% of England, as defined by the Index of Multiple Deprivation 2015 (ONS, 2015).

Coding

All caregivers' infant-directed speech had been transcribed and coded for semantic contingency on infant focus of attention as part of the longitudinal study from which the dataset originated (McGillion, Pine, et al., 2017). An utterance was coded as contingent if its semantic content was related to the attentional state of the infant in the five seconds prior to the onset of the utterance.

Measures

We extracted all instances of semantically contingent infant-directed speech occurring after an infant began a vocalisation, gesture or gesture-vocal combination, and within 1 second of that infant behaviour ending. This captured all speech that was both temporally and semantically contingent on an infants' vocalisations and gestures.

Analysis

To test whether gaze-coordinated behaviours were proportionally more likely to be responded to (in a temporally and semantically contingent way) than those produced without gaze-coordination, and whether this differed by behaviour type, we fitted a multi-level logistic regression model, considering each individual behaviour as a data point. For each behaviour, the outcome variable was whether it was met with a response (1 = response, 0 = no response). Infant was included as a random effect on the intercept and on all slopes.

Our analysis of the effect of responsiveness on later language took the same approach as the test of the predictive value of gaze-coordination in study 2. We added to the model space counts of the behaviours (and the gaze-coordinated behaviours) that included only instances that were responded to by caregivers. However, the combinatorial explosion that would arise from considering all combinations of all behaviours (both regardless-of-gaze and gaze-coordinated) in responded-to and regardless-of-response form made the exhaustive approach taken in Study 2 infeasible with the computational resources available. We therefore reduced the problem by taking only the models that were considered in the bootstrap smoothing for the final analysis of study 2 (note that these models contained all predictors included in Figure 6, just not in an exhaustive set of combinations of predictors).

We then derived all alternate models that arise from allowing each behaviour/subset of behaviours to be restricted to its/their responded-to only frequency. For example, taking a model containing two predictors - gaze-coordinated vocalisations and index-finger pointing (regardless of gaze) - we derived three alternate models; 1) a model in which both predictors only included instances of the behaviour that were responded to, 2) a model in which only gaze-coordinated vocalisations that were responded to were included, but all instances of pointing were included, and 3) a model in which all instances of gaze-coordinated vocalisations were included, but only those index-finger points that were responded to were included. We then took the subset of these additional models that were within 2 QAICc units of the best fitting model for our data and added these to the original models considered for study 2. Our enlarged model set for bootstrap smoothing thus included all credible combinations of responded-to and regardless-of-response counts for the predictors in each model considered in study 2. The following analysis thus examines the relative predictive value of responded-to forms, with the minor caveat that, due to the restricted model search space, we are looking at the effect of responsiveness on only those behaviours that had plausible predictive value independent of caregiver response.

Reliabilities

Reliabilities for semantic contingency coding of caregiver infant-directed speech were calculated as part of the original cohort study. Eleven percent of videos (randomly selected) were double-coded by a trained research assistant, with excellent rates of agreement, $\kappa = .87$ (McGillion, Pine, et al., 2017).

Results

Table 5 in Appendix E provides descriptive statistics for infant behaviours that were met with a caregiver response at 11 and 12 months combined (note caregiver responses reported here are both temporally and semantically contingent). Adding gaze-coordination (1= gaze-coordinated, 0 = not) to a null logistic regression model predicting whether a behaviour was responded to significantly improved fit (χ^2 (1) = 16.33, $p < .001$). Adding behaviour type (vocalisation, gesture or combination) further improved fit (χ^2 (2) = 86.27, $p < .001$) but adding an interaction term did not. As can be seen from Table 3 (where vocalisations regardless of gaze-coordination are the baseline case) a significantly higher proportion of gestures and combinations were met with a response than vocalisations, and further a significantly higher proportion of behaviours that were gaze-coordinated were met with a response than those that were not gaze-coordinated (Table 3). Thus intentionally communicative vocalisations, gestures and combinations were more successful in eliciting contingent responses from caregivers.

We next wanted to explore whether the frequency of responded-to behaviours (either gaze-coordinated or regardless-of-gaze) is particularly valuable in predicting later expressive vocabulary. It is important to note that the frequency of semantically contingent caregiver utterances (contingent talk) is a valuable predictor of vocabulary development, regardless of whether they are in response to a child's behaviour, i.e., regardless of temporal contingency (see McGillion, Pine, et al., 2017; Rollins, 2003). As a control, we thus introduced additional versions of each model in which the total frequency of caregivers contingent talk utterances (M = 173.5, SD = 66.43, Med = 172.50, range 35-307) was added, as well as a model in which this was the only predictor.

Critically, because the inclusion probabilities reported below are derived from a model space that includes both responded-to and regardless-of-response versions of all predictors, we can take the rates of inclusion for each behaviour as an indicator of their relative predictive value and thereby answer the question as to whether responded-to behaviours are better predictors. Unlike in study 2 where a large number of predictors had non-zero inclusion probabilities, here there is much reduced model uncertainty and only five predictors appear in any models at all: 1) gaze-coordinated vocalisations met with a caregiver response (inclusion probability of .521); 2) vocalisations (regardless of gaze) met with a caregiver response (.111), 3) caregiver contingent speech (.296), 4) gaze-coordinated non-CV vocalisations met with a caregiver response (.062) and 5) non-CV vocalisations (regardless of gaze) met with a caregiver response (.010). The effects of the three predictors with an inclusion probability greater than 0.1 are shown in figure 8.

Discussion

The first analysis in this study indicated that infants' prelinguistic behaviours are more successful in eliciting contingent responses from caregivers when they are gaze-coordinated. The second analysis demonstrated that, when specifically responded-to behaviours are added to the candidate model space, then the most valuable predictor of expressive vocabulary is the frequency with which a child produced gaze-coordinated vocalisations that were responded to. In a bootstrap procedure, responded-to gaze-coordinated vocalisations were included in the best model 52% of the time (with responded-to vocalisations regardless-of-gaze included 11% of the time). Caregiver contingent talk (speech that was semantically contingent, but not necessarily given as a response to the child's vocalisations, gestures or combinations) was included in the best model 29% of the time. All other variables were included less than 10% of the time. We conclude from these findings both that gaze-coordination is a valuable tool in eliciting caregiver contingent responses, and that caregiver responses further increase the predictive value of infant communicative behaviours.

It is worth clarifying that while some behaviours had predictive value in study 2 but not in study 3, this disappearance is not evidence that they have no relationship with vocabulary development. The unique contribution of this paper is in considering all behaviours in a single analysis and quantifying their relative predictive value. What we can infer from this analysis is that responded-to gaze-coordinated vocalisations have the greatest predictive value with regard to later language. Other predictors, e.g., gestures, have less value in the task of prediction but the earlier observed relationships remain of theoretical importance, as discussed in study 2.

General Discussion

The studies presented here provide a first move towards a unified account of the transition to word production based on a consideration of the full range of infants' prelinguistic vocalisations and gestures. We asked whether intentional communication from 11 months is especially predictive of word learning and whether caregiver responses to intentional communication further promote learning.

The first of these questions was addressed in studies 1 and 2. In study 1 we demonstrated that 11-month-olds coordinate many prelinguistic behaviours (both vocal and gestural) with gaze to their caregiver's face at above chance rates. This is consistent with the hypothesis that, as a group, 11-month-olds intend their actions to be communicative. We also, however, noted that some much-discussed behaviours (CV vocalisations and pointing) did not occur with gaze above chance at the group level, and noted the possibility that some instances of these behaviours might be intentionally communicative while others may serve a different function. In study 2, we demonstrated that individual differences in rates of production of gaze-coordinated vocalisations were valuable positive predictors of later expressive vocabulary. This is consistent with the hypothesis that instances of prelinguistic intentional vocal communication are especially predictive of later language because infants who can produce them are ready to make the leap to symbol use.

Together these results suggest that, while not all vocalisations are produced with communicative intent, those that plausibly are intentionally communicative play a role in driving later language. Previous work on the predictive role of babble has focused on CV vocalisations and established babble-language links (D'Odorico, Salerni, Cassibba, & Jacob, 1999; McCune & Vihman, 2001; McGillion, Herbert, et al., 2017; Menyuk, Liebergott, & Shultz, 1986; Stoel-Gammon, 1992). The current work suggests that the predictive value of vocalisations in general may not solely derive from being motoric prerequisites for speech but also from being an attempt to communicate intentionally. This is a critical developmental step.

Gestures were coordinated with gaze at above chance rates but were less valuable predictors of early word production. Showing gestures were the best positive gestural predictors of later language and seem to be produced intentionally, while rates of open-hand pointing were negative predictors, perhaps indicating motoric delay (as discussed in study 2). While the current studies underline the importance of vocalisations at the end of the first year of life in predicting the transition to language, it is possible that gestures become more important predictors later in development. Our measurements were taken around the age of pointing onset, where a majority of infants are unlikely to produce a high frequency of these gestures (Butterworth & Morissette, 1996; Carpenter, Nagell, & Tomasello, 1998; Desrochers et al., 1995; Leung & Rheingold, 1981). There is evidence to suggest that infants' gestures produced during the second year of life predict later language outcomes (Rowe & Goldin-Meadow, 2009; Rowe, Özçalışkan, & Goldin-Meadow, 2008) as do their gesture-vocal combinations (Goldin-Meadow & Butcher, 2003). Indeed, a meta-analysis of the relation between pointing and language development found that pointing became a stronger predictor of language outcomes with age (Colonnesi, Stams, Koster, & Noom, 2010). Moreover, caregiver response to infant gestures is a potential mechanism by which gestures facilitate language learning (Olson & Masur, 2015), and it is plausible that combining gestures and vocalisations gives caregivers additional information to provide timely, relevant input (Balog & Brentari, 2008; Fasolo & D'Odorico, 2012; Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007). Future work could use the methods employed in this paper to simultaneously evaluate the contribution of infant vocalisations and gestures, and caregiver speech produced later in development, to determine whether there is a change with time in the relative importance of these factors in predicting the transition to language.

In study 3, we demonstrated that caregivers were more likely to respond with semantically contingent speech to gaze-coordinated behaviours and indeed it was the dyadic combination of an infant's gaze-coordinated vocal behaviours with contingent caregiver responses that best predicted growth in expressive vocabulary in the second year. Our interpretation of the results from study 2 is that gaze-coordinated vocalisations are predictive because they indicate an ability to communicate intentionally; an ability that would bridge to language use. However, the results from study 3 support the claim that the behavioural indicator of intentional communication (i.e., gaze-coordination) is valuable at least in part because it is a powerful tool in eliciting responses. This could perhaps be due to caregivers viewing their infant's behaviour as intentionally communicative and responding informatively. We consider it unlikely that intentional communication predicts vocabulary development *only* because it elicits responses, but cannot conclusively rule this out. Whether infants' attempts to intentionally communicate represent efforts to shape their environment, driving their learning by provoking informative responses from their caregivers is a key question for future study.

It is worth noting that in studies 2 and 3 we focus on *frequencies* of behaviours rather than looking at the *proportion* of cases of a behaviour type that were gaze-coordinated or responded to by a caregiver. While both types of measure are of interest, to calculate proportions, a given type of behaviour has to be produced at least once, and therefore any infant who did not produce a given behaviour (e.g., did not produce an index-finger point) would have to be excluded from an analysis based on proportions (because we could not say what proportion of their index-finger points were gaze-coordinated). Given that many behaviours were produced infrequently, and some by a minority of infants, analyses with proportional predictors would have resulted in a substantially reduced sample size and selected for precociousness, making them less informative in terms of how children develop language in general. In addition, using proportions would not allow evaluation of the relative predictive value of each type of behaviour (as no infant produced all the behaviours considered in our analyses and even considering a narrower range of behaviours would result in substantial reductions in the sample). In short, using proportional measures would limit both the sample size and scope of analyses. Nonetheless, since proportional measures are of interest (e.g., Donnellan, 2017; Wu & Gros-Louis, 2014), we have included them in appendices C and E for descriptive purposes. Later in development, when gestures are produced by more infants and at a higher frequency, it may be possible to use the approach outlined in this paper to assess the relative value of proportional measures in predicting language development.

A novel contribution of this paper is the analytic methods (used in Study 2 and 3) that allowed us to look at all behaviours at once and thus compare their relative predictive value. It would benefit from replication on another cohort that also takes the unified approach outlined here. A second contribution of this paper is in the investigation of early intentional communication, and determining whether gaze-coordinated behaviours were particularly valuable predictors of the transition to later language. Many have argued that producing prelinguistic intentional communication is a theoretically important step towards producing language (e.g., Bates et al., 1979; Tomasello, 2008). This would be the case whether gaze-coordination is taken as an indicator of first-order or second-order intentional communication (as outlined in the introduction). In the case of first-order intentionality, gaze-coordination would indicate that the infant is using prelinguistic means to engage an interlocutor and is looking towards them in anticipation of a behavioural response, thus approximating the way in which words are eventually used. Many argue that gaze-coordination is a marker of second-order intentionality from around 12 months. For example, when infants point to things and check their caregiver's gaze, this is assumed to be an early instance of intentional, triadic communication in the sense that the infant intends to direct their interlocutor's attention to something in the external world (Bates et al., 1975; Matthews et al., 2012). When infants produce prelinguistic acts that are intentional in this second-order sense, we assume that they are at a jumping-off point for word use because all that needs to happen next is for conventional symbols to be used for the purpose of directing attention. In reality, infants may communicate sometimes with first-order intentionality and sometimes with second-order intentionality in a single play session. We assume there is a fluid transition to mastery and while our measure of intentionality collapses these levels in order to distinguish from behaviours that are less likely to be intentionally communicative (i.e., zero-order cases), future research could pick these levels apart.

Finally, the approach taken in this paper gives a general account of how behaviours at the end of the first year of life predict language in the second year. It is possible with a larger sample that our approach could be extended to identify different clusters of caregiver and infant behaviours that together form different communicative profiles. Such profiles may predict developmental trajectories and potentially highlight ways in which the caregiving environment might play a different role for children taking different routes to language.

In sum, infants intentionally communicate at 11 months of age, gazing to their caregiver's face whilst producing certain vocalisations and gestures at above chance rates. The frequency with which infants produce intentionally communicative vocalisations is the best predictor of their later expressive vocabulary, over and above the contribution of their early gestures. Moreover, these vocalisations elicit contingent responses from caregivers, and it was the dyadic combination of infant gaze-coordinated vocalisation and caregiver response that was by far the best predictor of later vocabulary size. We conclude that practice with prelinguistic intentional communication facilitates the leap to symbol use. Learning is optimised when caregivers respond to intentionally communicative vocalisations with appropriate language.

References

Bakeman, R., & Gottman, J. M. (1986). *Observing Interaction: An Introduction to Sequential Analysis*. New York, NY: Cambridge University Press.

Balog, H. L., & Brentari, D. (2008). The relationship between early gestures and intonation. *First Language*, *28*(2), 141–163. https://doi.org/10.1177/0142723708088722

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). Cognition and Communication from Nine to Thirteen Months: Correlational Findings. In E. Bates (Ed.), *The Emergence of Symbols: Cognition and Communication in Infancy* (pp. 69–140). New York; San Francisco; London: Academic Press.

Bates, E., Camaioni, L., & Volterra, V. (1975). The Acquisition of Performatives Prior to Speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology*, *40*(3), 293–310. https://doi.org/10.1002/dev.10034

Baumwell, L., Tamis-LeMonda, C. S., & Bornstein, M. H. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior and Development*, *20*(2), 247–258. https://doi.org/10.1016/S0163-6383(97)90026-6

Bornstein, M. H., Tamis-LeMonda, C. S., & Haynes, O. M. (1999). First Words in the Second Year : Continuity , Stability , and Models of Concurrent and Predictive Correspondence in Vocabulary and Verbal Responsiveness Across Age and Context. *Infant Behavior and Development*, *22*(1), 65–85.

Boundy, L., Cameron-Faulkner, T., & Theakston, A. (2019). Intention or Attention Before Pointing : Do Infants ' Early Holdout Gestures Reflect Evidence of a Declarative Motive ? *Infancy*, *24*(2), 228–248. https://doi.org/10.1111/infa.12267

Bruner, J. S. (1973). Organization of Early Skilled Action. *Child Development*, *44*(1), 1–11.

Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language*, *2*(01), 1–19. https://doi.org/10.1017/S0305000900000866

Bruner, J. S. (1976). From communication to language - A psychological perspective. *Cognition*, *3*(3), 255–287.

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model Selection: An Integral Part of Inference. *Biometrics*. https://doi.org/10.2307/2533961

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. *Ecological Modelling*.

https://doi.org/10.1016/j.ecolmodel.2003.11.004

Butterworth, G., & Morissette, P. (1996). Onset of pointing and the acquisition of language in infancy. *Journal of Reproductive and Infant Psychology*, *14*(3), 219–231. https://doi.org/10.1080/02646839608404519

Cameron-Faulkner, T., Theakston, A., Lieven, E., & Tomasello, M. (2015). The Relationship Between Infant Holdout and Gives, and Pointing. *Infancy*, *20*(5), 576–586. https://doi.org/10.1111/infa.12085

Carpendale, J. I. M., & Carpendale, A. B. (2010). The Development of Pointing: From Personal Directedness to Interpersonal Direction. *Human Development*, *53*(3), 110–126. https://doi.org/10.1159/000315168

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i-vi+1-166.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i–vi, 1-174.

Cartmill, E. A., & Byrne, R. W. (2010). Semantics of primate gestures: intentional meanings of orangutan gestures. *Animal Cognition*, *13*(6), 793–804. https://doi.org/10.1007/s10071-010-0328-7

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

D'Odorico, L., & Cassibba, R. (1995). Cross-Sectional Study of Coordination Between Infants' Gaze and Vocalizations Towards Their Mothers. *Early Development and Parenting*, *4*(1), 11–19.

D'Odorico, L., Cassibba, R., & Salerni, N. (1997). Temporal relationships between gaze and vocal behavior in prelinguistic and linguistic communication. *Journal of Psycholinguistic Research*, *26*(5), 539–556. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9329205

D'Odorico, L., Salerni, N., Cassibba, R., & Jacob, V. (1999). Stability and change of maternal speech to Italian infants from 7 to 21 months of age: a longitudinal study of its influence on early stages of language acquisition. *First Language*, *19*(57), 313–346. https://doi.org/10.1177/014272379901905702

Delgado, B., Gómez, J.-C., & Sarriá, E. (2011). Pointing gestures as a cognitive tool in young children: experimental evidence. *Journal of Experimental Child Psychology*, *110*(3), 299–312. https://doi.org/10.1016/j.jecp.2011.04.010

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The “Panglossian paradigm” defended. *Behavioral and Brain Sciences*, *6*, 343–390. https://doi.org/10.1017/S0140525X00016393

Desrochers, S., Morissette, P., & Ricard, M. (1995). Two perspectives on pointing in infancy. In P. J. Dunham & C. Moore (Eds.), *Joint Attention: Its origins and role in development* (pp. 85–101). Hillsdale, NJ: Lawrence Erlbaum.

Donnellan, E. (2017). *Intentional Communication in Infants and Toddlers (PhD Thesis)*. University of Sheffield.

Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: evidence from Catalan-babbling infants. *Journal of Child Language*, *40*(5), 919–944. https://doi.org/10.1017/S0305000912000359

Fasolo, M., & D’Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture*, *12*(1), 1–15. https://doi.org/10.1075/gest.12.1.01fas

Franco, F., & Butterworth, G. (1996). Pointing and social awareness: declaring and requesting in the second year. *Journal of Child Language*, *23*(02), 307–336. https://doi.org/10.1017/S0305000900008813

Goldin-Meadow, S., & Butcher, C. (2003). Pointing Toward Two-Word Speech in Young Children. In S. Kita (Ed.), *Pointing: Where Language, Culture and Cognition Meet* (pp. 85–108). Lawrence Erlbaum Associates.

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science*, *10*(6), 778–785. https://doi.org/10.1111/j.1467-7687.2007.00636.x

Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child Development*, *80*(3), 636–644. https://doi.org/10.1111/j.1467-8624.2009.01287.x

Golinkoff, R. M. (1986). ‘I beg your pardon?’: the preverbal negotiation of failed messages. *Journal of Child Language*, *13*(3), 455–476. https://doi.org/10.1017/S0305000900006826

Gros-Louis, J., & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions: sensitivity to parents' attention and responses. *Infant Behavior and Development*, *35*(4), 773–778. https://doi.org/10.1016/j.infbeh.2012.07.016

Grünloh, T., & Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *Journal of Child Language*, *42*(06), 1312–1336. https://doi.org/10.1017/S0305000914000816

Harding, C. G., & Golinkoff, R. M. (1979). The origins of intentional vocalizations in prelinguistic infants. *Child Development*, *50*(1), 33–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/446215

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, *39*, 42–52. https://doi.org/10.1016/j.infbeh.2015.02.004

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science*, *16*(5), 367–371.

Kass, R., & Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.1995.10476572

Laakso, M. L., Poikkeus, A. M., Katajamaki, J., & Lyytinen, P. (1999). Early intentional communication as a predictor of language development in young toddlers. *First Language*, *19*(56), 207–231. https://doi.org/10.1177/014272379901905604

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (Pan troglodytes). *Child Development*, *76*(1), 291–306. https://doi.org/10.1111/j.1467-8624.2005.00845.x

Leung, E. H. L., & Rheingold, H. L. (1981). Development of Pointing as a Social Gesture. *Developmental Psychology*, *17*(2), 215–220.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Reference and attitude in infant pointing. *Journal of Child Language*, *34*, 1–20. https://doi.org/10.1017/S0305000906007689

Luke, C., Grimminger, A., Rohlfing, K., Liszkowski, U., & Ritterfeld, U. (2016). In Infants' Hands : Identification of Preverbal Infants at Risk for Primary Language Delay. *Child Development*, *00*(0), 1–9. https://doi.org/10.1111/cdev.12610

Maljaars, J., Noens, I., Jansen, R., Scholte, E., & van Berckelaer-Onnes, I. (2011). Intentional communication in nonverbal and verbal low-functioning children with autism. *Journal* *of Communication Disorders*, *44*(6), 601–614. https://doi.org/10.1016/j.jcomdis.2011.07.004

Matthews, D., Behne, T., Lieven, E., & Tomasello, M. (2012). Origins of the human pointing gesture: a training study. *Developmental Science*, *15*(6), 817–829. https://doi.org/10.1111/j.1467-7687.2012.01181.x

McCune, L., & Vihman, M. M. (2001). Early Phonetic and Lexical Development: A Productivity Approach. *Journal of Speech, Language and Hearing Research*, *44*(3), 670–684. https://doi.org/10.1044/1092-4388(2001/054)

McGillion, M. L., Herbert, J. S., Pine, J. M., Keren-portnoy, T., Vihman, M. M., & Matthews, D. E. (2013). Supporting early vocabulary development: What sort of responsiveness matters?

McGillion, M. L., Herbert, J. S., Pine, J., Vihman, M. M., DePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child Development*, *88*(1), 156–166. https://doi.org/10.1111/cdev.12671

McGillion, M. L., Pine, J. M., Herbert, J., & Matthews, D. (2017). A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse SES backgrounds. *Journal of Child Psychology and Psychiatry*, (June).

Meints, K., & Woodward, A. Electronic vocabulary database Lincoln Lincoln UK-CDI Infants and/or the Lincoln UK-CDI Toddlers. (2011). WWW document: Computer software & manual].

Menyuk, P., Liebergott, J., & Shultz, M. (1986). Predicting Phonological Development. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of Early Speech* (pp. 79–94). New York, NY: Stockton Press.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes Factors for Common Designs. Retrieved from http://cran.r-project.org/package=BayesFactor

Murillo, E., & Belinchón, M. (2012). Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture*, *12*(2012), 16–39. https://doi.org/10.1075/gest.12.1.02mur

Murillo, E., & Capilla, A. (2016). Properties of vocalization- and gesture-combinations in the transition to first words. *Journal of Child Language*, *43*(4), 890–913. https://doi.org/10.1017/S0305000915000343

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6318–6323. https://doi.org/10.1073/pnas.1300337110

Olson, J., & Masur, E. F. (2013). Mothers respond differently to infants’ gestural versus nongestural communicative bids. *First Language*, *33*(4), 372–387. https://doi.org/10.1177/0142723713493346

Olson, J., & Masur, E. F. (2015). Mothers’ labeling responses to infants’ gestures predict vocabulary outcomes. *Journal of Child Language*, *42*(6), 1289–1311. https://doi.org/10.1017/S0305000914000828

Papaeliou, C. F., Minadakis, G., & Cavouras, D. (2002). Acoustic patterns of infant vocalizations expressing emotions and communicative functions. *Journal of Speech, Language, and Hearing Research*, *45*(2), 311–317. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12003513

Papaeliou, C. F., & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing ‘communication’ and ‘apprehension.’ *Journal of Child Language*, *33*, 163–178. https://doi.org/10.1017/S0305000905007300

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (Gorilla gorilla): gestural repertoire, learning, and use. *American Journal of Primatology*, *60*(3), 95–111. https://doi.org/10.1002/ajp.10097

Rollins, P. R. (2003). Caregivers’ contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, *24*(May 2017), 221–234. https://doi.org/10.1017.S0142716403000110

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, *12*(1), 182–187. https://doi.org/10.1111/j.1467-7687.2008.00764.x

Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture’s role in predicting vocabulary development. *First Language*, *28*(2), 182–199. https://doi.org/10.1177/0142723707088310

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PloS One*, *8*(10), e76674. https://doi.org/10.1371/journal.pone.0076674

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*.

Stoel-Gammon, C. (1992). Prelinguistic Vocal Development: Measurement and Predictions. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research, Implications* (pp. 439–456). Timonium, Maryland: York Press.

Sugiura, N. (1978). Further analysis of the data by Akaike's Information Criterion and the finite corrections. *Communications in Statistics - Theory and Methods*. https://doi.org/10.1080/03610927808827599

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The Ontogeny of Chimpanzee Gestural Signals: A Comparison Across Groups and Generations. *Evolution of Communication*, *1*(2), 223–259. https://doi.org/10.1075/eoc.1.2.04tom

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., … Manser, M. B. (2017). Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals. *Biological Reviews*. https://doi.org/10.1111/brv.12289

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From Babbling to Speech: A Re-Assessment of the Continuity Issue. *Language*, *61*(2), 397. https://doi.org/10.2307/414151

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, *34*(1), 72–90. https://doi.org/10.1177/0142723714521925

Wu, Z., & Gros-Louis, J. (2017). The Value of Vocalizing: 10-Month-Olds' Vocal Usage Relates to Language Outcomes at 15 Months. *Infancy*, *22*(1), 23–41. https://doi.org/10.1111/infa.12150

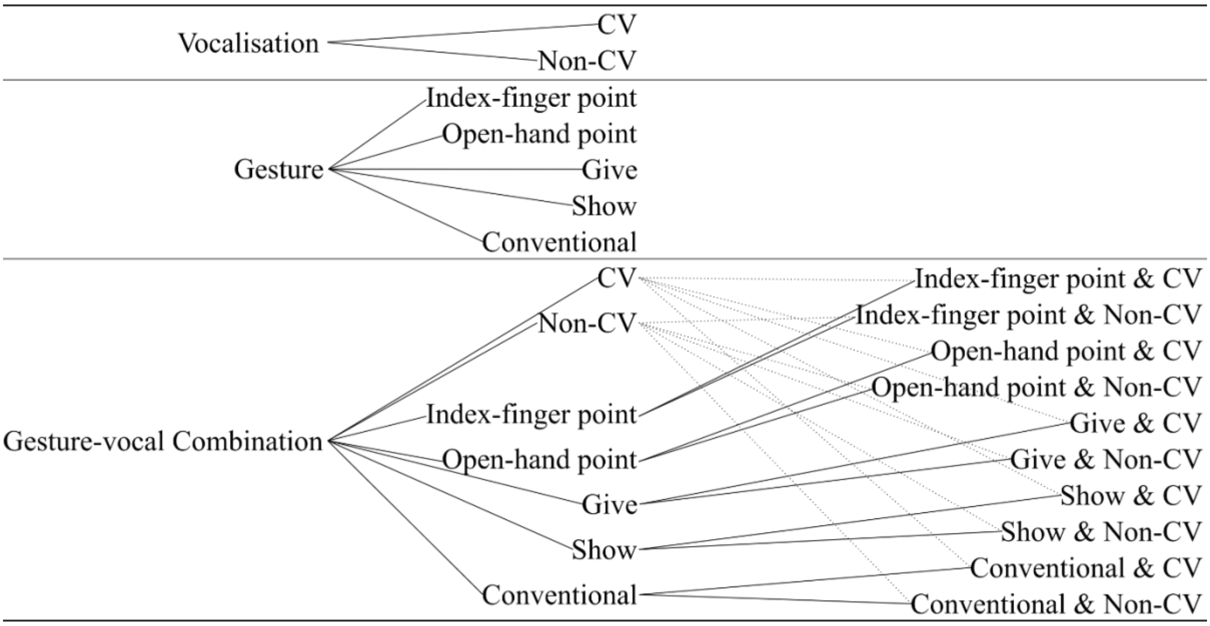

Figure 1: Full set of infant prelinguistic behaviours, and sub-types of these behaviours

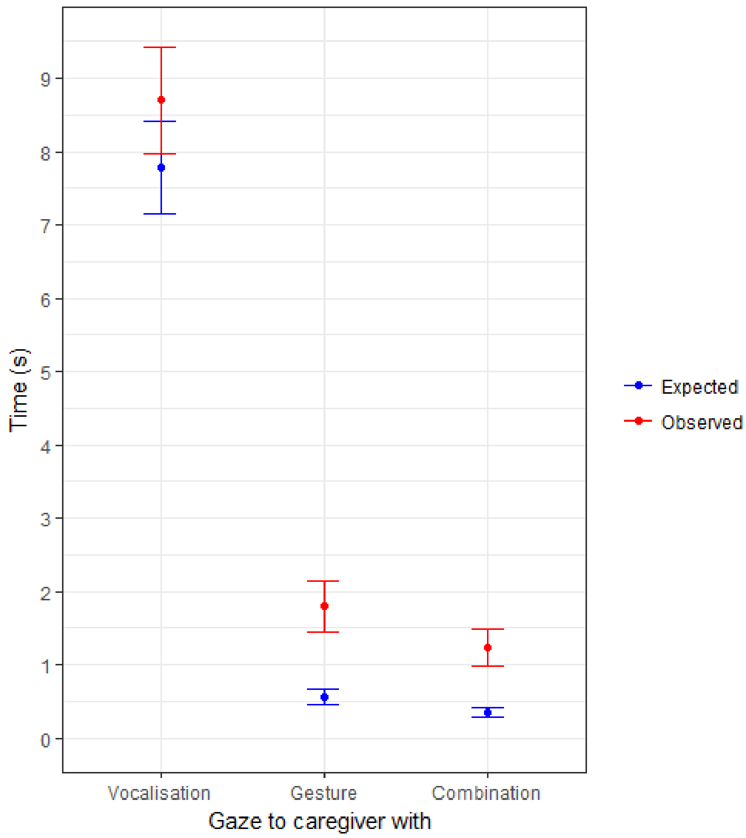

Figure 2: Mean expected and observed co-occurrence of vocalisations, gestures and gesture-vocal combinations with gaze to caregiver's face at 11 months (n = 134). Error bars represent standard error about the mean

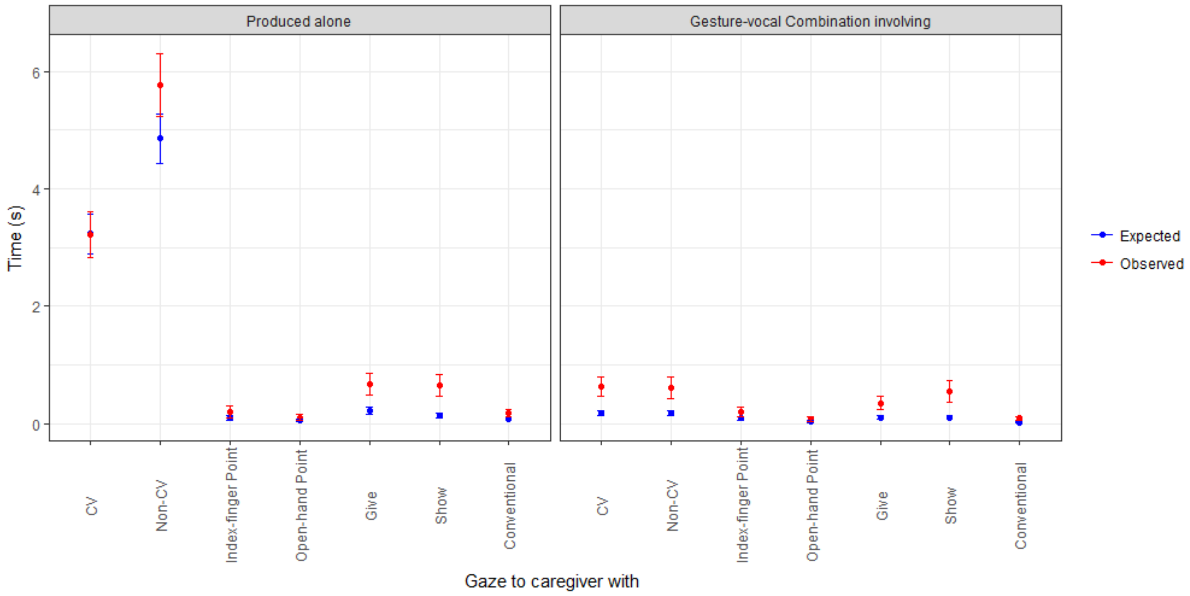

Figure 3: Mean expected and observed co-occurrence of individual vocalisation and gestures types (both produced alone and as part of a gesture-vocal combination) with gaze to caregiver's face at 11 months (n = 134). Error bars represent standard error about the mean

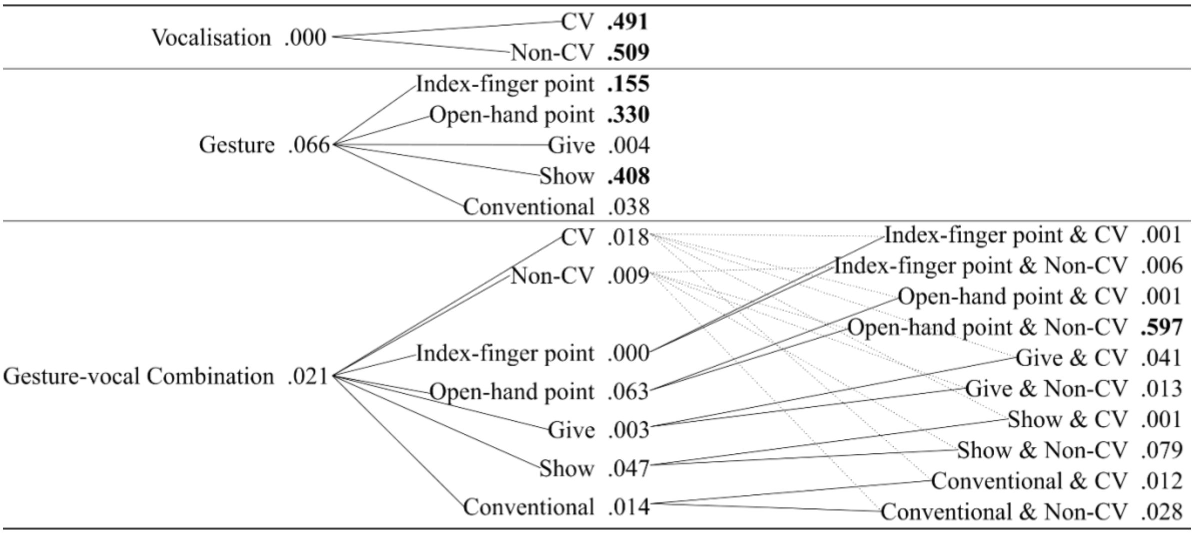

Figure 4. Inclusion probability of predictors relating to the frequency of prelinguistic infant behaviours (regardless of whether these behaviours are gaze-coordinated). For ease of visualisation, predictors with an inclusion probability >0.1 are shown in bold.

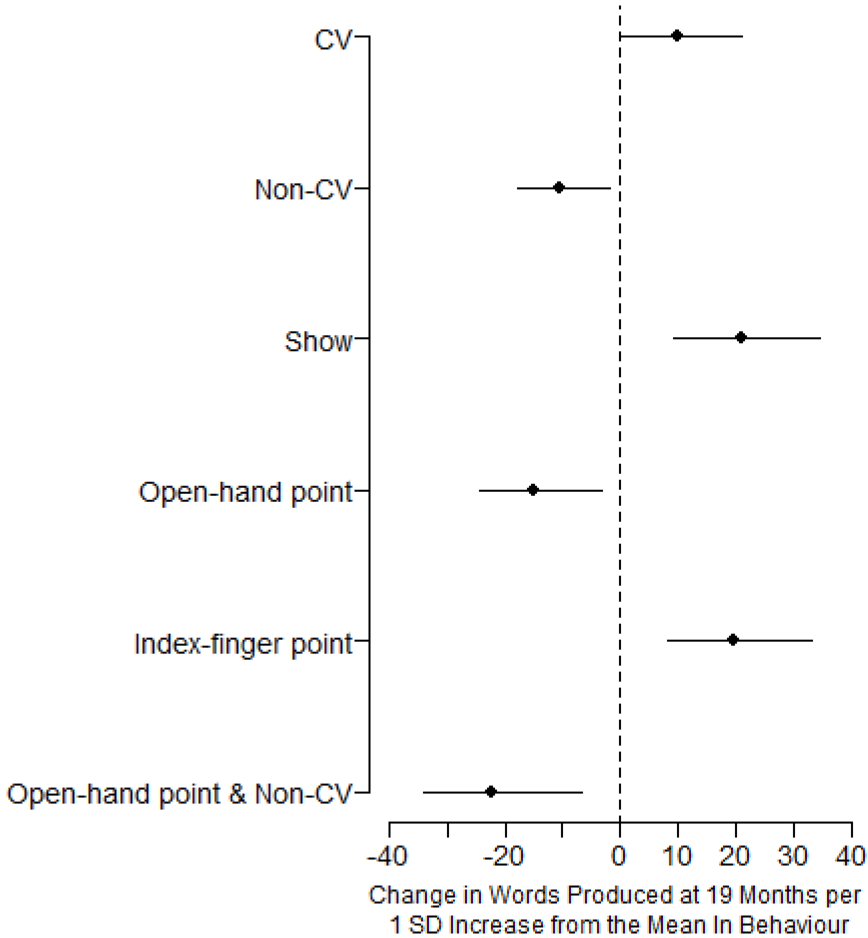

Figure 5. Effect size of predictors with an inclusion probability of >0.1 from figure 4. The X-axis represents the change in the number of words we would expect the child to produce (at 19 months, the mean age for the model) given 1 SD increase in the behaviour at 11 and 12 months. Negative values denote a negative relationship between the predictor behaviour and the outcome (e.g. more non-CV vocalisations are associated with a smaller expressive vocabulary than the mean), whilst positive values denote a positive relationship between predictor behaviour and outcome (e.g. more showing gestures are associated with larger vocabulary).

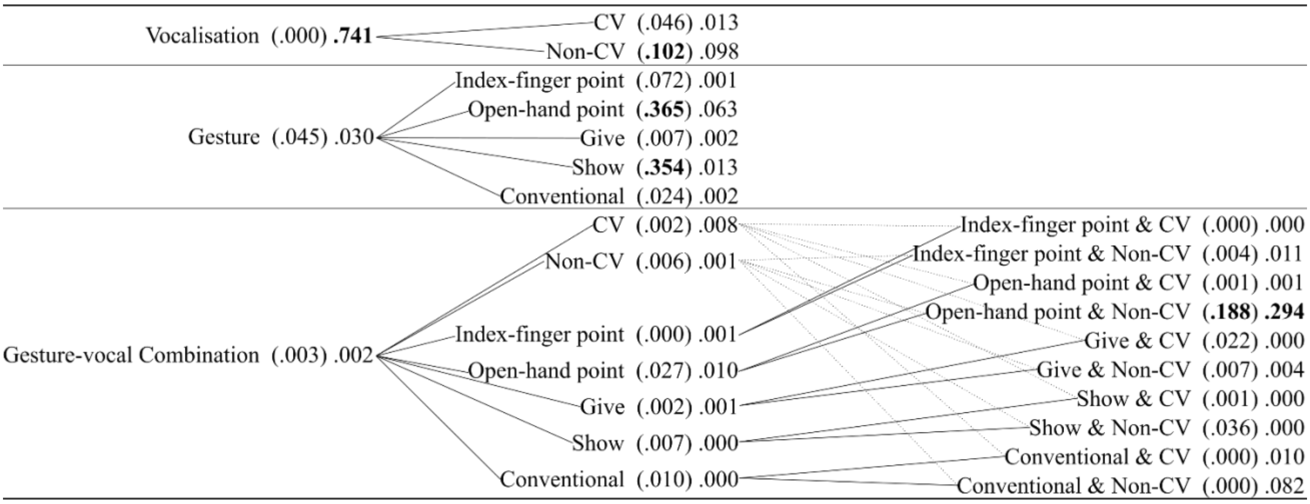

Figure 6. Inclusion probability of predictors relating to the frequency of prelinguistic infant behaviours. Inclusion probability of the predictor consisting of all instances of the behaviour (i.e., regardless of gaze coordination) is denoted in parentheses. Inclusion probability of the version of the predictor consisting only of those instances where the behaviour was gaze-coordinated is denoted without parentheses. For ease of visualisation, predictors with an inclusion probability > 0.1 are shown in bold.

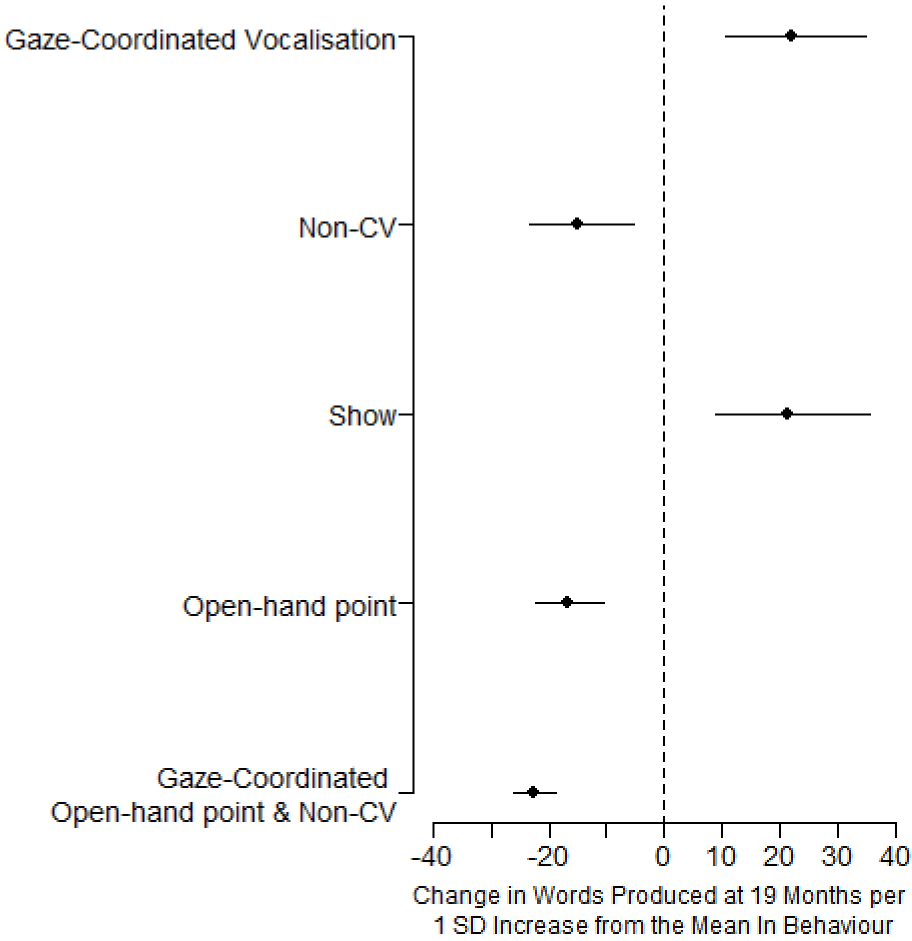

Figure 7. Effect size of predictors with an inclusion probability of >0.1 from figure 6. The X-axis represents the change in the number of words we would expect the child to produce (at 19 months, the mean age for the model) given 1 SD increase in the behaviour at 11 and 12 months.

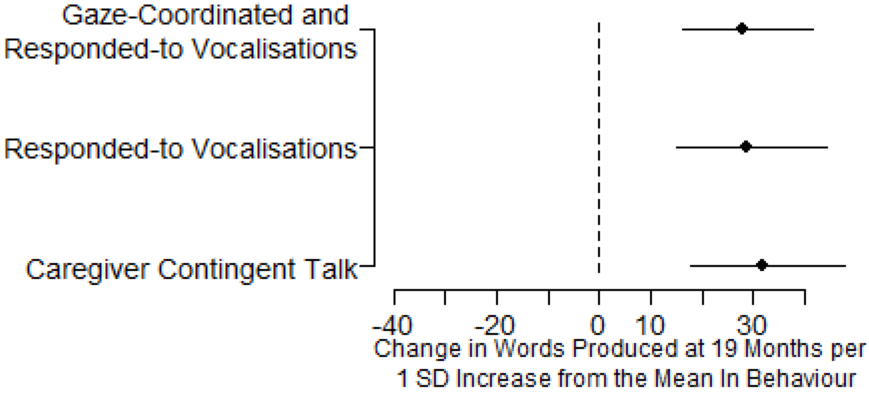

Figure 8. Effect size of predictors with an inclusion probability of >0.1. The X-axis represents the change in the number of words we would expect the child to produce (at 19 months, the mean age for the model) given 1 SD increase in the behaviour at 11 and 12 months.

Table 1

Summary of T-tests and Bayes Factor Analyses Comparing Mean Expected and Observed Co-occurrence of Vocalisations, Gestures and Gesture-vocal Combinations with Gaze to Caregiver's Face at 11 months (n = 134)

	$\bar{x}$	*df*	*t*	*BF*
Vocalisations (without gesture)	0.92	133	2.36	1.39
Gestures (without vocalisations)	1.23	133	4.85	4000.87
Gesture-vocal combinations	0.90	133	4.32	504.40

Table 2

Summary of T-tests and Bayes Factor Analyses Comparing Mean Expected and Observed Co-occurrence of Subtypes of Vocalisations, Gestures and Gesture-vocal Combinations with Gaze to Caregiver's Face at 11 months (n = 134)

	$\bar{x}$	*df*	*t*	*BF*
Vocalisations (without gesture)				
CV	-0.01	133	-0.03	0.30
Non-CV vocalisations	0.91	133	2.94	9.71
Gestures (without vocalisations)				
Index-finger pointing	0.10	133	1.60	0.52
Open-hand pointing	0.51	133	1.36	0.37
Giving	0.46	133	3.20	16.63
Showing	0.52	133	3.44	32.88
Conventional Gestures	0.11	133	2.27	1.71
Gesture-vocal combinations				
By vocalisation				
CV	0.46	133	3.22	18.40
Non-CV vocalisations	0.44	133	3.17	15.82
By gesture				
Index-finger pointing	0.10	133	1.48	0.48
Open-hand pointing	0.03	133	1.14	0.32
Giving	0.25	133	2.59	3.68
Showing	0.45	133	2.95	8.79
Conventional Gestures	0.07	133	2.42	2.48

Table 3

Summary of Fixed Effects from a Logistic Regression Model Fitting Vocalisations (*n* = 5129), *Gestures* (*n* = 264) *and Combinations (n* = 164) *that were Met with a Caregiver Response (1= Response, 0 = No response) by Gaze Coordination and Behaviour Type at 11 & 12 months (n = 58)*

	B	*SE*	*z*	*p*
Intercept (*No Gaze Coordination: Vocalisation*)	-1.50	0.10	-15.66	< .001
Gaze Coordination	0.50	0.10	5.11	< .001
Gesture	2.06	0.17	12.02	< .001
Combination	2.00	0.22	9.28	< .001

LLRI = .02, C = .73, Dxy = .46.

Appendix A

Coding Scheme

Gaze to caregiver's face. All instances where the infant looked to the caregiver's face were coded. These were marked from the frame that was judged to be the beginning of the look, to the last frame where the infant was judged to be looking at their caregiver's face.

Vocalisations. All infant vocalisations were coded except crying vocalisations, vegetative noises, and fussing noises (D'Odorico & Cassibba, 1995; Murillo & Belinchón, 2012; Wu & Gros-Louis, 2014). For each vocalisation, the beginning of the vocalisation was marked at the frame where the vocalisation began, and the end was marked at the last frame where the vocalisation was still audible. Vocalisations were considered separate when separated by 200ms of silence, in line with the literature suggesting that a short pause, often (but not necessarily) including a breath, delineates distinct vocalisations (Murillo & Belinchón, 2012; Vihman et al., 1985).

CV vocalisations consisted of at least one syllable that itself contained at least one consonant (C) and vowel (V) (see also D'Odorico et al., 1999; Grünloh & Liszkowski, 2015). In line with McCune and Vihman (2001), we code only supraglottal consonants, excluding glides and glottals. All vocalisations that did not contain a CV syllable were coded as non-CV vocalisations.

Gestures. We coded 5 types of infant gesture. While not an exhaustive set of infant gestures, any remaining types were so rare as to not warrant coding. For all these gestures, the beginning of the gesture was marked at the frame where arm reached maximum extension, and the end is marked at the frame where retraction of the arm began. To create continuity with the vocalisation coding scheme, if the arm was extended within 200ms of the previous arm retraction, this is counted as the same gesture.

Index-finger points and open-hand points were coded when an infant extended either hand (or both) while looking at an object or event of interest. The arm(s) had to be extended, the hand(s) had to be empty, and the child was not leaning forward and did not touch what was being pointed at (Matthews et al., 2012; McGillion, Herbert, et al., 2017). For index-finger points, the index finger(s) was clearly and visibly separate from the other fingers, which were partially or entirely curled back, and the index finger extended in the direction of the object or event being looked at. For open-hand points, a majority of fingers were extended in the direction of the object or event being looked at.

Giving and showing were coded when the infant held out an object with either (or both) arms extended towards the caregiver while holding the object. For a show, the object was held up towards the caregiver's face, while for a give the object was extended in the direction of the caregiver's hands, or extended in a way so as to deliver the object into the vicinity of the caregiver (Cameron-Faulkner et al., 2015; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998).

Due to low frequencies, a number of remaining gestures were coded under one category of conventional gestures. These included 1) arm up where the infant raised both arms in order to initiate being picked up, 2) wave where the infant waved with palm vertical (or close to vertical) and moving side to side, 3) all gone where the infant shrugged with palm of hand facing up, similar to adults asking, 'where?', and 4) baby sign were also coded.

Gesture-vocal combinations. When all or part of a vocalisation and gesture overlapped in time, this was considered a gesture-vocal combination (see also Igualada et al., 2015). For all these gesture-vocal combinations, the beginning of the combination was marked at the frame where the first element (either vocal or gestural) of the combination began (as coded above), and the end was marked at the frame the last element of the combination ended (as coded above).

Combinations could either involve a CV vocalisation or involve a non-CV vocalisation. In cases where they involved both CV and non-CV vocalisations, they were counted as involving CV vocalisations. Combinations could also involve any of the five gesture types. No instances of combinations involving two different gesture types was observed. This gave us 10 types of gesture-vocal combination (2 vocalisation types x 5 gesture types).

Off-shot measures. We did not code data from periods where 1) the infant was completely out of shot, 2) it was not possible to tell if the infant was looking to their caregiver's face, (more detail below) and, 3) the infants arms were not visible, making it impossible to ascertain whether they had produced a gesture. Only behaviours with full temporal windows (i.e., where the 1 second window around a behaviour did not overlap with one of these off-shot periods) were included in analyses.

Regarding (2), this could be when the infant's eyes were not in shot, the position of caregiver's face was not known, or the infant was looking in the direction of caregiver's face, but there was partial occlusion between caregiver and infant that made it impossible to tell if the infant was looking to the caregiver's face (i.e., unclear whether they had a direct line of sight). To exclude these periods it had to be possible that the infant could have looked to their caregiver’s face, but it was not possible to conclusively determine if they had. For example, if infant’s eyes were not in shot (i.e., they were looking straight down at the floor, with their caregiver behind them), it was clear that the infant was not capable of gazing to the caregiver’s face, so data was still coded from this period.

Appendix B

Reliabilities

In order to calculate reliabilities on the 11 month data (used in study 1) a trained research assistant blind to the aims of the study coded gaze to caregiver's face, vocalisations and gestures for 10% of participants (*n* = 14). In study 2, we collapsed the 11 month data with data from 12 months, so here we also present reliabilities for 12 month data. In order to calculate reliabilities on the 12 month data, we used a 10% (of the full sample, *n* = 13) overlap in coding between the first author and the same research assistant.

Agreement on the frequency of infant behaviours was high at both 11 months (for gaze to caregiver's face, $r = .95$; for vocalisations, $r = .99$; for gestures, $r = .82$; for combinations, $r = .93$) and 12 months (for gaze to caregiver's face, $r = .95$; for vocalisations, $r = .98$; for gestures, $r = .97$; for combinations, $r = .95$).

Additionally, we tested whether the frequency of vocalisations, gestures and gesture-vocal combinations with gaze coordination was reliable, and again, agreement was high at both 11 months (for vocalisations, $r = .95$; for gestures, $r = .89$; for combinations, $r = .94$) and 12 months (for vocalisations, $r = .96$; for gestures, $r = .96$; for combinations, $r = .94$).

For agreed vocalisations, gestures and combinations, Cohen's kappa was calculated for gaze coordination (was the behaviour coordinated with gaze or not), and indicated high levels of agreement at both ages. At 11 months, Cohen's kappa was high for vocalisations, $\kappa = .82$, $p < .001$ (agreement on coding of 96%); for gestures, $\kappa = .86$, $p <.001$ (93%); and for combinations, $\kappa = .77$, $p = .013$ (89%). At 12 months, Cohen's kappa was high for vocalisations, $\kappa = .85$, $p < .001$ (96%); for gestures, $\kappa =.97$, $p < .001$ (98%); and for combinations, $\kappa = .92$, $p < .001$ (96%).

In terms of gesture type coding, we intended to calculate kappas on gesture type (whether gestures were classified as index-finger pointing, open-hand pointing, giving, showing or conventional gestures) on agreed gestures, however there was 100% agreement at 11 months. At 12 months, Cohen's kappa for gestures was, $\kappa =0.85$, $p <.001$ (agreement on coding of 90%), indicating excellent agreement.

Finally, for vocalisation type coding (whether they were classified as CV or non-CV), a separate phonologically trained researcher (the third author) independently classified vocalisations for 10% of the sample at both ages. Cohen's kappa for vocalisations at 11 months indicated excellent agreement, $\kappa = .80$, $p < .001$ (agreement on coding of 91%), as did Cohen's kappa at 12 months, $\kappa = .81$, (agreement on coding of 91%).

Appendix C

Infant behaviours at 11 months

Table 4

Mean Frequency of Infant Behaviours, and Frequency and Proportion of Behaviours Produced With Gaze Coordination at 11 months (n = 134)

					Gaze-Coordinated				
	Frequency Produced				*Frequency*				*Prop*
Behaviour	*M*	*SD*	*Med*	*Range*	*M*	*SD*	*Med*	*Range*	*M*
Gaze to Caregiver's Face	22.21	12.11	19	1-53					
Vocalisations (without gesture)	47.50	28.82	41.5	4-172	8.48	7.53	7	0-36	.18
CV	18.45	17.94	14	0-108	3.16	4.32	2	0-23	.16
Non-CV	29.05	17.49	25.5	2-82	5.33	4.85	4	0-22	.19
Gestures (without vocalisation)	1.64	2.26	1	0-11	0.99	1.73	0	0-9	.56
Index-finger point	0.22	0.74	0	0-7	0.10	0.51	0	0-5	.43
Open-hand point	0.20	0.58	0	0-4	0.07	0.34	0	0-3	.29
Give	0.68	1.41	0	0-8	0.41	1.07	0	0-7	.61
Show	0.29	0.77	0	0-4	0.28	0.76	0	0-4	.93
Conventional gesture	0.25	0.80	0	0-6	0.13	0.45	0	0-3	.52

Gesture-vocal combinations	1.10	2.13	0	0-12	0.62	1.20	0	0-6	.59
By vocalisation type									
CV	0.63	1.45	0	0-9	0.32	0.78	0	0-4	.54
Non-CV	0.47	1.15	0	0-8	0.30	0.86	0	0-5	.57
By gesture type									
Index-finger point	0.31	1.32	0	0-12	0.11	0.50	0	0-4	.37
Open-hand point	0.19	0.96	0	0-10	0.07	0.55	0	0-6	.24
Give	0.34	0.99	0	0-8	0.21	0.64	0	0-4	.61
Show	0.16	0.58	0	0-3	0.16	0.57	0	0-3	.92
Conventional gesture	0.10	0.36	0	0-2	0.07	0.25	0	0-1	.75
By gesture and vocalisation type									
Index-finger point & CV	0.24	0.99	0	0-9	0.10	0.45	0	0-3	.44
Index-finger point & Non-CV	0.07	0.39	0	0-3	0.01	0.09	0	0-1	.07
Open-hand point & CV	0.10	0.45	0	0-4	0.03	0.21	0	0-2	.22
Open-hand point & Non-CV	0.09	0.71	0	0-8	0.04	0.44	0	0-5	.33
Give & CV	0.14	0.51	0	0-4	0.07	0.25	0	0-1	.60
Give & Non-CV	0.20	0.70	0	0-4	0.14	0.56	0	0-3	.63

Show & CV	0.10	0.44	0	0-3	0.09	0.43	0	0-3	.88
Show & Non-CV	0.07	0.33	0	0-2	0.07	0.33	0	0-2	1.00
Conventional gesture & CV	0.05	0.25	0	0-2	0.03	0.17	0	0-1	.58
Conventional gesture & Non-CV	0.04	0.24	0	0-2	0.04	0.19	0	0-1	.90

Appendix D

Study 2 Participants Supplementary Data

We included infants for whom we had naturalistic observations at both 11 and 12 months and a measure of expressive vocabulary at 15, 18 and/or 24 months We had naturalistic observations for 58 caregiver-infant dyads (33 female infants, 25 male) at both ages (11 months mean age = 334 days, SD = 4 days; 12 months mean age = 365 days, SD = 4 days) who had expressive vocabulary outcomes at 15, 18 or 24 months. We had expressive vocabulary outcomes for 53 caregiver-infant dyads (30 female infants, 23 male) at 15 months (mean age = 456 days, SD = 17 days); 40 dyads (20 female, 20 male) at 18 months (mean age = 572 days, SD = 10 days); and 49 dyads (28 female, 21 male) at 24 months (mean age = 773 days, SD = 40 days).

Appendix E

Infant behaviours at 11 and 12 months

Table 5

Mean and Median Frequency of Infant Behaviours, Frequency and Proportion of Behaviours Produced With Gaze Coordination and Frequency and Proportion of Behaviours Met with a Caregiver Response at 11 & 12 months (n = 58)

			Gaze-Coordinated			*Responded to (Regardless of Gaze Coordination)*			*Responded to and Gaze-Coordinated*		
	Frequency Produced		*Frequency*		*Prop*	*Frequency*		*Prop*	*Frequency*		*Prop*
Behaviour	*M (SD)*	*Med (Range)*	*M (SD)*	*Med (Range)*	*M (SD)*	*M (SD)*	*Med (Range)*	*M (SD)*	*M (SD)*	*Med (Range)*	*M (SD)*
Gaze to Caregiver's Face	44.61 (22.25)	40 (5-99)									
Vocalisations (without gesture)	88.99 (39.99)	83.5 (17-197)	15.55 (10.62)	13.5 (0-51)	.18 (.11)	19.43 (14.15)	16 (0-66)	.21 (.11)	4.55 (4.02)	4 (0-19)	.05 (.04)
CV	35.17 (25.58)	29.5 (2-125)	5.77 (5.85)	4 (0-30)	.17 (.14)	8.24 (9.75)	4 (0-51)	.22 (.15)	1.69 (2.39)	1 (0-13)	.05 (.06)
Non-CV	53.81 (26.51)	52 (12-145)	9.77 (6.76)	9 (0-26)	.19 (.12)	11.20 (7.7)	11 (0-44)	.21 (.11)	2.86 (2.61)	2.5 (0-9)	.05 (.05)
Gestures (without vocalisation)	4.60 (5.16)	2 (0-22)	2.25 (2.74)	1.5 (0-12)	.51 (.33)	3.22 (3.87)	2 (0-16)	.71 (.29)	1.61 (2.03)	1 (0-8)	.37 (.30)
Index-finger point	0.81 (1.83)	0 (0-9)	0.22 (0.77)	0 (0-5)	.28 (.40)	0.53 (1.44)	0 (0-7)	.55 (.4)	0.14 (0.58)	0 (0-4)	.12 (.20)

Open-hand point	0.28 (0.76)	0 (0-4)	0.12 (0.48)	0 (0-3)	.39 (.49)	0.14 (0.5)	0 (0-3)	.47 (.51)	0.07 (0.39)	0 (0-3)	.17 (.35)
Give	2.53 (3.27)	1 (0-13)	1.2 (1.95)	0 (0-8)	.45 (.38)	1.93 (2.76)	1 (0-12)	.77 (.28)	0.94 (1.57)	0 (0-7)	.37 (.34)
Show	0.48 (0.94)	0 (0-4)	0.43 (0.88)	0 (0-4)	.89 (.27)	0.34 (0.71)	0 (0-3)	.75 (.39)	0.29 (0.65)	0 (0-3)	.64 (.42)
Conventional gesture	0.50 (1.16)	0 (0-7)	0.28 (0.62)	0 (0-3)	.62 (.42)	0.28 (0.77)	0 (0-5)	.57 (.42)	0.17 (0.42)	0 (0-2)	.43 (.45)
Gesture-vocal combinations	2.84 (3.51)	1.5 (0-14)	1.76 (2.49)	1 (0-12)	.57 (.34)	1.96 (2.46)	1 (0-9)	.71 (.28)	1.21 (1.75)	0.5 (0-8)	.41 (.31)
By vocalisation type											
CV	1.51 (2.34)	1 (0-12)	0.94 (1.38)	0 (0-6)	.66 (.37)	1.09 (1.66)	0 (0-7)	.77 (.29)	0.70 (1.08)	0 (0-4)	.50 (.40)
Non-CV	1.34 (2.25)	0 (0-13)	0.83 (1.83)	0 (0-11)	.51 (.44)	0.87 (1.54)	0 (0-7)	.59 (.40)	0.52 (1.17)	0 (0-6)	.30 (.36)
By gesture type											
Index-finger point	0.71 (1.98)	0 (0-12)	0.33 (0.98)	0 (0-6)	.43 (.45)	0.45 (1.37)	0 (0-7)	.65 (.43)	0.14 (0.61)	0 (0-4)	.13 (.28)
Open-hand point	0.48 (1.47)	0 (0-10)	0.26 (0.93)	0 (0-6)	.42 (.38)	0.28 (1.01)	0 (0-7)	.50 (.40)	0.14 (0.48)	0 (0-3)	.24 (.24)

Give	1.10 (2.07)	0 (0-12)	0.68 (1.6)	0 (0-10)	.59 (.38)	0.82 (1.55)	0 (0-7)	.78 (.32)	0.54 (1.18)	0 (0-6)	.52 (.40)
Show	0.40 (0.79)	0 (0-3)	0.40 (0.79)	0 (0-3)	1.00 (.00)	0.31 (0.68)	0 (0-3)	.77 (.39)	0.31 (0.68)	0 (0-3)	.77 (.39)
Conventional gesture	0.16 (0.41)	0 (0-2)	0.10 (0.31)	0 (0-1)	.69 (.46)	0.10 (0.31)	0 (0-1)	.75 (.46)	0.09 (0.28)	0 (0-1)	.63 (.52)
By gesture and vocalisation type											
Index-finger point & CV	0.53 (1.80)	0 (0-12)	0.26 (0.91)	0 (0-6)	.47 (.42)	0.33 (1.15)	0 (0-7)	.61 (.44)	0.12 (0.59)	0 (0-4)	.15 (.32)
Index-finger point & Non-CV	0.17 (0.53)	0 (0-3)	0.07 (0.32)	0 (0-2)	.43 (.53)	0.12 (0.46)	0 (0-3)	.64 (.48)	0.02 (0.13)	0 (0-1)	.07 (.19)
Open-hand point & CV	0.26 (0.61)	0 (0-2)	0.12 (0.38)	0 (0-2)	.40 (.39)	0.14 (0.44)	0 (0-2)	.45 (.44)	0.07 (0.26)	0 (0-1)	.20 (.26)
Open-hand point & Non-CV	0.22 (1.09)	0 (0-8)	0.14 (0.69)	0 (0-5)	.73 (.44)	0.14 (0.80)	0 (0-6)	.45 (.45)	0.07 (0.41)	0 (0-3)	.28 (.44)
Give & CV	0.45 (0.85)	0 (0-4)	0.32 (0.63)	0 (0-3)	.74 (.40)	0.40 (0.80)	0 (0-4)	.91 (.26)	0.28 (0.59)	0 (0-3)	.68 (.43)
Give & Non-CV	0.65 (1.66)	0 (0-11)	0.36 (1.28)	0 (0-9)	.45 (.42)	0.42 (1.08)	0 (0-6)	.60 (.42)	0.26 (0.83)	0 (0-5)	.33 (.40)
Show & CV	0.17 (0.53)	0 (0-3)	0.17 (0.53)	0 (0-3)	1.00 (.00)	0.16 (0.52)	0 (0-3)	.86 (.38)	0.16 (0.52)	0 (0-3)	.86 (.38)

Show & Non-CV	0.22 (0.53)	0 (0-2)	0.22 (0.53)	0 (0-2)	1.00 (.00)	0.16 (0.41)	0 (0-2)	.75 (.42)	0.16 (0.41)	0 (0-2)	.75 (.42)
Conventional gesture & CV	0.09 (0.28)	0 (0-1)	0.07 (0.26)	0 (0-1)	.80 (.45)	0.07 (0.26)	0 (0-1)	.80 (.45)	0.07 (0.26)	0 (0-1)	.80 (.45)
Conventional gesture & Non-CV	0.07 (0.26)	0 (0-1)	0.03 (0.18)	0 (0-1)	.50 (.58)	0.03 (0.18)	0 (0-1)	.50 (.58)	0.02 (0.13)	0 (0-1)	.25 (.50)