Infant Behavior and Development 71 (2023) 101828

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

Infants' behaviours elicit different verbal, nonverbal, and multimodal responses from caregivers during early play

Anika van der Klis *, Frans Adriaans, René Kager

Institute for Language Sciences, Utrecht University, the Netherlands

ARTICLE INFO

Keywords:
Caregiver-child interactions
Responsiveness
Multimodal language
Free play
Infancy
YOUth Cohort Study

ABSTRACT

Caregivers use a range of verbal and nonverbal behaviours when responding to their infants. Previous studies have typically focused on the role of the caregiver in providing verbal responses, while communication is inherently multimodal (involving audio and visual information) and bidirectional (exchange of information between infant and caregiver). In this paper, we present a comprehensive study of caregivers' verbal, nonverbal, and multimodal responses to 10-month-old infants' vocalisations and gestures during free play. A new coding scheme was used to annotate 2036 infant vocalisations and gestures of which 87.1 % received a caregiver response. Most caregiver responses were verbal, but 39.7 % of all responses were multimodal. We also examined whether different infant behaviours elicited different responses from caregivers. Infant bimodal (i.e., vocal-gestural combination) behaviours elicited high rates of verbal responses and high rates of multimodal responses, while infant gestures elicited high rates of nonverbal responses. We also found that the types of verbal and nonverbal responses differed as a function of infant behaviour. The results indicate that infants influence the rates and types of responses they receive from caregivers. When examining caregiver-child interactions, analysing caregivers' verbal responses alone undermines the multimodal richness and bidirectionality of early communication.

1. Introduction

During early play sessions, infants may babble while pointing at a doll. Their caregiver may pick up the doll and ask: "Do you want this doll?". The infant starts reaching for the doll, extending both arms while opening and closing their fingers. Their caregiver smiles in understanding and hands over the doll. Such interactions between infants and caregivers help the infant to identify the label "doll" for the object they were interested in. More generally, such interactions teach the infant to communicate effectively by producing sounds and gestures. Previous studies have shown that caregiver responses to infant vocalisations and gestures differ in terms of frequency and contents (e.g., Ger et al., 2018; McGillion et al., 2013; Olson & Masur, 2013; Tamis-LeMonda & Bornstein, 2002; Wu & Gros-Louis, 2014). However, while communication is inherently multimodal (involving audio and visual information) and bidirectional (exchange of information between infant and caregiver), previous studies have mostly focused on verbal responses provided by the caregiver. It is therefore not known to what extent caregivers produce nonverbal or multimodal responses, and to what extent such caregiver responses are elicited by different infant behaviours. Understanding the full extent of early infant-caregiver interactions is crucial for understanding infants' early socio-cognitive development. The current study has taken an important step towards this goal by investigating caregivers' verbal, nonverbal, and multimodal responses to 10-month-old infants' vocalisations and gestures during

* Correspondence to: Institute for Language Sciences, Utrecht University, Trans 10, 3512 JK, Utrecht, the Netherlands.
E-mail address: a.vanderklis@uu.nl (A. van der Klis).

https://doi.org/10.1016/j.infbeh.2023.101828
Received 14 September 2022; Received in revised form 3 February 2023; Accepted 16 February 2023
Available online 23 February 2023

free play.

1.1. Infants learn to communicate

Infants learn to communicate by producing vocalisations and gestures. A study by Donnellan et al. (2019) found that 11-month-olds varied greatly in their vocalisations during a play session at home, with some infants producing two or three vocalisations, and other infants producing more than a hundred vocalisations during the session. This shows that early vocalisations are characterised by large variability. The developmental trajectories of gestures are also cross-linguistically characterised by both variability and stability (Fenson et al., 1994; Frank et al., 2021). The earliest deictic gestures typically involve giving and showing, later followed by index-finger pointing and requesting, although the ages of onset can vary drastically across children (Frank et al., 2021). These individual differences in the productions of vocalisations and deictic gestures, specifically index-finger pointing, have been positively associated with children's vocabulary outcomes (e.g., Brooks & Meltzoff, 2008; Choi et al., 2021; Colonnesi et al., 2010; Rowe et al., 2022). It remains unclear, however, why individual differences are related to children's later language skills. Some studies have suggested that infant behaviours elicit specific contingent responses from caregivers that facilitate language development (Ger et al., 2018; Olson & Masur, 2015), for example, providing a label for an object that the infant was pointing at. Crucially, infants must first produce vocalisations and gestures to create opportunities for their caregivers to respond. The information flow is in both directions.

1.2. Communication is bidirectional

Young infants already expect their caregivers to respond to their prelinguistic vocalisations and gestures. At 5 months of age, infants have learned that vocalisations elicit caregiver responses (Goldstein et al., 2009). By 10–12 months, infants use deictic gestures with the motive to share attention and interest with others (Boundy et al., 2019; Liszkowski et al., 2004). The onset of declarative communication therefore takes place before infants learn how to speak. In the process, infants learn which behaviours are effective at eliciting which types of caregiver responses. In turn, caregivers should be sensitive to differences in infants' communicative behaviours and respond contingently and appropriately. This bidirectional view of communication has recently shifted the focus from studying individual behaviours to examining a shared system in which infants and caregivers both shape the interaction (Chen et al., 2021; Renzi et al., 2017). When examining individual differences in caregivers' language input to their infant, we cannot attribute all variation to the caregivers themselves because infants also influence the input they receive from their caregivers.

In particular, studies have shown that caregiver responses vary as a function of child behaviour (e.g., Choi et al., 2021; Gros-Louis et al., 2006; Olson & Masur, 2013). Olson and Masur (2013) showed that mothers provide more object labels to gestural than non-gestural bids. This shows that infants who produce more gestures tend to elicit more labelling responses from their caregivers. The type of vocalisation or gesture also influences the response. More specifically, infant index-finger pointing gestures have been found to elicit more labelling responses than reaching gestures (Kishimoto et al., 2007; Wu & Gros-Louis, 2015). In a recent study, Choi et al. (2021) showed that caregivers respond more often to their 10-month-old infants' showing + giving gestures than to their pointing gestures. They only examined these types of gestures and did not distinguish between different types of responses. In addition, mothers were found to use more verbal than nonverbal responses to infant vocalisations, and they responded with more imitations to consonant-vowel (CV) sequences compared to vowel-like sounds (Gros-Louis et al., 2006). Recently, Yurkovic et al. (2021) showed that multimodal behaviours (looks combined with touch) by infants aged 12–48 months elicited higher caregiver response rates than unimodal behaviours and elicited more multimodal (looks combined with touch) responses from caregivers. These studies provide some initial evidence that infant behaviours tend to elicit caregiver responses in the same modality. Symmetry between modalities could suggest high synchrony between children and their caregivers (see Leclère et al., 2014), but more research is needed to establish from which age and to what extent caregiver-child dyads match modalities in infant behaviour and caregiver response sequences. The studies so far suggest that certain infant behaviours tend to elicit higher response rates and different response types, but they only examined a few types of infant behaviours and caregiver responses. We currently miss a detailed characterisation of caregivers' verbal, nonverbal, and multimodal behaviours in response to infants' vocalisations, gestures, and bimodal behaviours.

1.3. Relevance of verbal and nonverbal responses

Caregivers individually differ in their verbal responsiveness which has been found to positively relate to children's vocabulary development (e.g., Donnellan et al., 2019; McGillion et al., 2013; Olson & Masur, 2015; Wu & Gros-Louis, 2014). Variation in caregiver responsiveness is rooted in a variety of factors. One factor is socio-economic status (SES). Mothers of a higher SES have been found to produce more speech and verbally respond more often than mothers from lower SES backgrounds (e.g., Hart & Risley, 1995; McGillion et al., 2017; Vanormelingen & Gillis, 2016). It has been suggested that this is the main reason why children from lower SES backgrounds tend to have smaller vocabularies (Huttenlocher et al., 2010), although it is also possible that some infants produce fewer behaviours that elicit verbal responses, giving fewer opportunities for their caregivers to provide contingent language input. A contingent verbal response, such as labelling an object that the infant is pointing at, creates a temporal and semantic contingency that allows the infant to match the phonological form of a word with its meaning. Studies examining individual differences in caregiver responsiveness have therefore largely focused on differences in caregivers' verbal responses.

However, there is evidence that nonverbal responses play a facilitative role as well. Caregivers' responsiveness measured both verbally *and* nonverbally positively relates to their infants' socio-cognitive skills, including language development (see Bornstein & Tamis-LeMonda, 1989). Studies have found that specific caregivers' nonverbal behaviours, such as handing over a toy, pointing, or

smiling, predict vocabulary outcomes and social skills (Pearson et al., 2011; Ruddy & Bornstein, 1982). In addition, nonverbal behaviours regularly co-occur with speech. Children appear to rely on visual information when speech is novel (e.g., a label for an unfamiliar object) or unclear (e.g., in the case of referential ambiguity). Studies have found that children use gaze direction, body orientation, and index-finger pointing as cues to learn the reference of novel words from both humans and robots (Baldwin et al., 1996; Grassmann & Tomasello, 2010; Kory Westlund et al., 2017; Verhagen et al., 2019). Recently, Chen et al. (2021) showed that caregivers touched objects more often while naming them when the object was unfamiliar to the child. Overall, approximately 40 % of all caregivers' utterances are accompanied by at least one visual cue (Ger et al., 2018; Vigliocco et al., 2019). These studies suggest that caregivers tend to use many different types of nonverbal cues when providing children with novel speech. Studies have not yet addressed variation across caregivers in nonverbal or multimodal responsiveness, except for one study which found that mothers are more likely to respond verbally, while fathers are equally likely to produce verbal or nonverbal responses (Flippin and Watson, 2011). Existing studies have not yet identified which types of nonverbal and multimodal responses occur during early caregiver-child interactions, nor whether infants also affect their caregivers' nonverbal and multimodal responsiveness.

1.4. Current study

Previous studies have documented caregivers' verbal responses in much detail, while different strands of research have shown that nonverbal and multimodal behaviours occur frequently in caregivers' communication with infants. The first aim of this study therefore was to examine which infant vocalisations, gestures, and bimodal behaviours and which caregiver verbal, nonverbal, and multimodal responses occur during early play. For this analysis, we developed a new coding scheme that includes various types of caregivers' nonverbal behaviours, such as gestures, facial expressions, and other non-gestural bodily behaviours, including body orientation. By applying this coding scheme to a large sample of caregiver-child dyads, our study obtained new insights into the richness and variability of early interactions. In addition, while most research focused on the role of the caregiver in providing contingent responses, some studies suggest that infants play a role in eliciting specific caregiver response rates and types. The second aim of this study was to assess whether infants' vocalisations, gestures, and bimodal behaviours elicited different verbal, nonverbal, and multimodal responses. We examined this through statistical analyses of co-occurring infant behaviours and caregiver responses. These analyses informed us to what extent infants affect their caregivers' responsiveness during early play – thereby shaping their own language experience.

2. Method

2.1. Participants

The data for this study are derived from YOUth, an ongoing longitudinal cohort study that is part of Utrecht University and University Medical Center Utrecht (see Onland-Moret et al., 2020). YOUth has repeated measurements at regular intervals ("waves"). The current study uses measurements obtained at the age of 9–11 months. From the original sample, we excluded 5 dyads due to technical issues during the recordings resulting in unclear or distorted audio/video, and 1 dyad was excluded because the child was vocalising non-stop throughout the entire video. The final sample consisted of 117 infants (66 females) around 9–11 months of age ($M = 10.5$ months, $SD = 0.9$) and their caregivers (92 mothers; 25 fathers). These dyads were selected because they spoke Dutch at home and completed the caregiver-child interaction task. The YOUth cohort study is carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki), and all caregivers have signed informed consent. The study was approved by the medical ethical committee of the University Medical Center Utrecht (application number 14–7–221). Children received a picture book after participating.

We collected information regarding the caregivers' languages spoken at home and their education level via questionnaires. All caregivers included in this study speak only Dutch at home. Most caregivers in our sample were willing to share information regarding their education level (94.9 %). We rated their highest level of education on a 5-point scale. Most caregivers (81.1 %) completed university or college education and another 15.3 % completed senior secondary vocational education. Only a small percentage of caregivers (3.6 %) did not continue their education after secondary school. We also collected each partner's educational level when applicable. We found that caregivers' educational levels were moderately correlated with their partners' ($\rho = .33$, $p < .01$). Thus, most caregivers in this sample are highly educated.

2.2. Procedure

During the caregiver-child interaction task, the infant and their caregiver were asked to sit next to each other within touching distance on a playing rug in a sparsely furnished room. Before the task, the research assistant placed various toys on the rug. The positioning of all items was the same for each participant. The toys were directly in front of both the caregiver and the infant on the rug. The research assistant read a specific set of instructions to ensure each participant received the same instructions. Participants were filmed from four camera angles placed around the rug. Three Dome cameras could be controlled (i.e., moved and zoomed in and out) by the research assistant. One camera captured both the infant and their caregiver from the side, one camera focused on the infant's face, and the other camera focused on the caregiver's face. There was also one fixed camera providing an overview of the entire scene. To capture sound, a fixed, standing (Sennheiser ME64/K6P condenser) microphone is positioned next to the rug. After reading out the set of instructions, the research assistant would take place behind a screen so the caregiver and child could not see them while being filmed. Other caregivers or siblings were not allowed in the lab to minimise distractions.

The caregiver and their infant were filmed for a total of fifteen minutes, subdivided into five different tasks. The session started with three minutes of free play with a standard set of toys and ended with three minutes of free play before the toys were cleaned up. The toys were a baby doll with a milk bottle, a green toy car, a Bumba pop-up toy, and a sun-shaped rattle, in addition to a shape sorter and a picture book that were only available in the last three minutes of free play. The caregivers were asked beforehand to carry the infant back to the rug in case they crawled away. Furthermore, after every three-minute episode, the research assistant gave clear instructions for the next task, e.g.: "It is now time for free play.". After reading the instructions, the research assistant started a stopwatch to film for three minutes. For the present study, we analysed the first and last sessions of free play, analysing six minutes of free play per dyad in total.

2.3. Coding scheme

In this section, we present a new coding scheme for annotating infants' vocalisations, gestures, and bimodal behaviours and caregivers' responses. This is the first coding scheme to include various types of verbal, nonverbal, and multimodal responses.

2.3.1. Coding infant behaviours

All videos were coded in ELAN 6.0 (Sloetjes & Wittenburg, 2008). Based on previous studies, an infant vocalisation was any sound produced by the infant except vegetative sounds (e.g., hiccoughs or coughs) or distress sounds (e.g., crying or fussing). A vocalisation was coded as CV vocalisation when at least one syllable contained a consonant-vowel sequence ("baba", "ma" etc.), excluding glides ("ja") and glottals ("ha") (following e.g., Donnellan et al., 2019). All other types of vocalisations were coded as non-CV vocalisation. The categories of infant gestures were (1) index-finger pointing, (2) whole-hand pointing, (3) other rudimentary forms of pointing (e.g., using a fist), (4) showing, (5) giving, (6) reaching, (7) requesting, and (8) other conventional gestures, such as waving or nodding. The selection of these gestures was partially based on previous studies examining infant gestures and caregiver responses (e.g., Donnellan et al., 2019; McGillion et al., 2013; Olson & Masur, 2015; Wu & Gros-Louis, 2014) and the gestures included in the widely used checklist to measure infants' vocabularies: The MacArthur-Bates Communicative Development Inventory (CDI) - Words and Gestures (Fenson et al., 2007). In ELAN, the end of each gesture was marked at the frame where the retraction of the arm began. Infant gestures and vocalisations were coded independently as they could occur simultaneously. This automatically revealed which infant behaviours were bimodal (i.e., gesture-vocal combinations). There had to be at least a partial overlap between the vocalisation and gesture. Otherwise, they were annotated as two separate infant behaviours. For the detailed coding scheme and definitions for each behaviour, see Appendix A.

2.3.2. Coding caregiver responses

After the offset of the infant gesture or vocalisation, a period of exactly two seconds was measured for the caregiver response (following e.g., McGillion et al., 2013; Wu & Gros-Louis, 2014). The onset of the caregiver response had to occur either during the infant's behaviour or within this two-second time frame. If not, the response was not considered temporally contingent and not analysed in this study. First, we annotated a binary measure indicating whether there was a caregiver response or not. This could be any type of response in the coding scheme (i.e., any verbal or nonverbal behaviour). Then in detail, we annotated which types of caregiver responses occurred. This could be any verbal, gestural, facial, and/or bodily response. These four categories were not mutually exclusive: more than one type of behaviour could occur at the same time. *Verbal* responses were coded as either (1) semantically contingent (i.e., a follow-in response), (2) onomatopoeias or sound effects, (3) infant imitations, or (4) any other type of verbal response that was not semantically related (i.e., non-contingent), such as an affirmation. This selection was based on previous studies (e.g., Donnellan et al., 2019; Motamedi et al., 2021; Tamis-LeMonda & Bornstein, 2002). If the verbal response was not an onomatopoeia or sound effect or an infant imitation, the verbal contents were coded as either contingent or non-contingent. A response was semantically contingent if its semantic content was related to the attentional state of the infant (following Donnellan et al., 2019; McGillion et al., 2017). We assumed the object or activity was the infant's focus of attention when the infant was vocalising while either holding the object or playing with the object, performing the activity, looking at the object, and/or gesturing towards the object. Otherwise, the verbal response was coded as non-contingent.

For nonverbal responses, we included different types of gestures, facial expressions, and other (non-gestural and non-facial) bodily behaviours. *Gestural* responses included various types of manual and non-manual gestures. They were further subdivided into (1) pointing, (2) passing, (3) showing (i.e., without manipulating the object), (4) accepting, (5) a representational gesture, 6) object manipulations (see Murgiano et al., 2021), or (7) any other conventional gesture (e.g., nodding or waving). We also included representational gestures (i.e., showing the size, shape, or how an object works without the object in hand) and object manipulations (i.e., when the caregiver interacts physically with an object to play with it or communicate about it) based on the ECOLANG project (Vigliocco et al., in prep).

For *facial* responses, we distinguished between (1) smiling (including laughter), (2) surprise, and (3) other facial expressions different from neutral. We included smiling because it is typical of infant-directed speech (see Benders, 2013). We included surprise because we may expect this to occur frequently during early play, as "mock" surprise in the context of, for example, playing peek-a-boo. The latter (other) category was added to ensure that the coding process was exhaustive. A facial expression only counted as a response if the caregiver was not already showing the expression before the start of the infant's behaviour. This was to ensure that the annotated facial expressions were truly responses to infants' behaviours. Finally, all *bodily* responses were subdivided into (1) leaning closer to the infant, (2) turning to the infant, (3) turning to the toy, and (4) any affective behaviour (e.g., hugging or touching the infant). Body orientations were included in the coding scheme because they may serve as referential cues when hearing novel speech

(e.g., Kory Westlund et al., 2017). For the full coding scheme including definitions, see Appendix A.

2.4. Training, improving, and reliabilities

We had a three-step process to complete data annotation: training the research assistant, improving the coding scheme after a pilot, and checking inter-rater reliabilities. The first author wrote coding instructions and the initial version of the coding scheme. The first author and research assistant annotated the same three randomly selected videos. They verbally went over all annotations to discuss any differences and uncertainties. After that, the assistant and first author both separately annotated the videos again including an additional seven randomly selected videos. To assess inter-annotator reliabilities, we report chance-corrected modified Cohen's kappa (κ) using the built-in calculator in ELAN 6.0 which is based on the EasyDIAg toolbox (Holle & Rein, 2015). The modified kappa considers both the categorisation of behaviours and the temporal overlap of annotations (i.e., segmentation). While this is a good measure of reliability, kappa values are affected by the large number of coding categories and the infrequent occurrence of some codes. When the marginal distributions are not uniform, the maximum value of kappa cannot reach 1.0 (von Eye & von Eye, 2008). Therefore, we also report maximum kappa, which aids interpretation of the reported kappa values, and raw agreement, representing the number of agreements on cases divided by the total number of cases, which is a more intuitive measure.

For the first ten videos of the pilot, we found high agreement on categorising different behaviours ($\kappa = .87$; $\kappa_{max} = .96$; raw = .91). When also including unmatched annotations, the overall agreement dropped to a level that is below satisfaction ($\kappa = .35$; $\kappa_{max} = 91$; raw = .46). This suggests that there was high agreement on the behaviours that were annotated by both annotators, but there were still many false positives or false negatives (i.e., situations in which behaviours were only annotated by one of the two annotators). The first author manually examined any other deviations in annotations and used this to redefine definitions or coding criteria (e.g., we changed the criteria for separating annotations of infant vocalisations and gestures, and more clearly defined the offset of an infant gesture which is important for starting the two-second response window, and better clarified some definitions). After improving the coding scheme, both annotators updated the pilot set accordingly. We achieved a satisfactory reliability score of $\kappa = .75$ ($\kappa_{max} = .95$; raw = .83), including unmatched annotations.

The last step involved an additional blind inter-annotator reliability check at the end. A random selection of seven videos was again double-coded by the first author. Overall, chance-corrected Cohen's kappa shows agreement of $\kappa = .81$ ($\kappa_{max} = .94$; raw = .87), including unmatched annotations, which is excellent. We also looked at agreement for each coding category. For infant behaviours, there was high agreement on the classification of infant vocalisations ($\kappa = .70$; $\kappa_{max} = .85$; raw = .97) and infant gestures ($\kappa = .79$; $\kappa_{max} = .86$; raw = .98). We found that the frequencies of the total number of observations per dyad were strongly correlated between the two annotators ($r_s = .95$, $p < .01$). This suggests that overall, identification and classification of infant vocalisations and gestures was strongly reliable. We also examined agreement on the categorisation of caregivers' verbal and nonverbal responses. For the binary variable indicating whether there was a response of any type, we find excellent agreement ($\kappa = .95$, $\kappa_{max} = .95$; raw = .99). We also examined agreement on the different categories of caregivers' verbal responses ($\kappa = .97$; $\kappa_{\max} = .98$; raw = .98), gestural responses ($\kappa = .81$; $\kappa_{max} = 94$; raw = .90), facial responses ($\kappa = .74$; $\kappa_{max} = .87$; raw = .97), and bodily responses ($\kappa = .66$; $\kappa_{max} = .66$; raw = .99) were all excellent. We find no more than three bodily responses in this set which is reflected in the lower maximum kappa. We reflect more on this in the discussion.

2.5. Statistical analyses

All analyses were carried out in *R* version 4.2.0 (R Core Team, 2022). First, we aimed to examine which infants' vocalisations, gestures, and bimodal behaviours and which caregivers' verbal, nonverbal, and multimodal response types occurred in this large, naturalistic dataset. In the first sections, we present descriptive statistics of all annotated behaviours to address this first aim. We calculated the total number of productions for each behaviour, the production range, and the proportion of participants who produced the behaviour at least once. Lastly, we examined which infant vocalisations and gestures and which caregiver verbal and nonverbal responses were often combined to form an infant bimodal behaviour or a caregiver multimodal response.

The second aim of this study was to examine whether different infant behaviours elicit different caregiver response rates and types. To examine this, we first fitted sets of logistic mixed-effects models using three binary outcomes indicating the presence or absence of a verbal, nonverbal, or multimodal response. All models were fitted with a random intercept for subjects using the lme4 package version 1.1–31 (Bates et al., 2015) since we have multiple observations per dyad. In the first set of models, we used infant behaviour (vocalisation, gesture, or bimodal) as the predictor variable. Then, we also wanted to examine in more detail whether different gestures or different vocalisations also elicited different response rates. In the next set of models, we used infant gesture (index-finger pointing, whole-hand pointing, showing, giving, reaching, requesting, or other) or infant vocalisation (CV or non-CV) as the predictor variables. We used dummy coding with the category containing the largest number of observations (infant behaviours: vocalisations; infant gestures: reaching; infant vocalisations: non-CV) as reference levels. Lastly, we present Chi-square test statistics to examine whether there was a relationship between infants' vocalisations and gestures and caregivers' verbal and gestural responses. We used the non-parametric Fisher's exact test to examine this for facial and bodily response types due to the low frequencies in these categories. The results indicate whether different infant behaviours elicited different response rates and types.

Table 1
Total frequencies of infant behaviours including production range and percentage of infants who produced the behaviour.

Infant behaviour	Frequency	Range	Percentage
Vocalisations			
CV vocalisation	55	0–13	19.7
Non-CV vocalisation	1837	1–54	100.0
Total	1892		
Gestures			
Index-finger pointing	25	0–4	12.0
Whole-hand pointing	4	0–1	3.4
Showing	18	0–5	9.4
Giving	56	0–7	18.8
Reaching	78	0–6	30.8
Requesting	9	0–2	6.0
Other	17	0–7	4.3
Total	207		

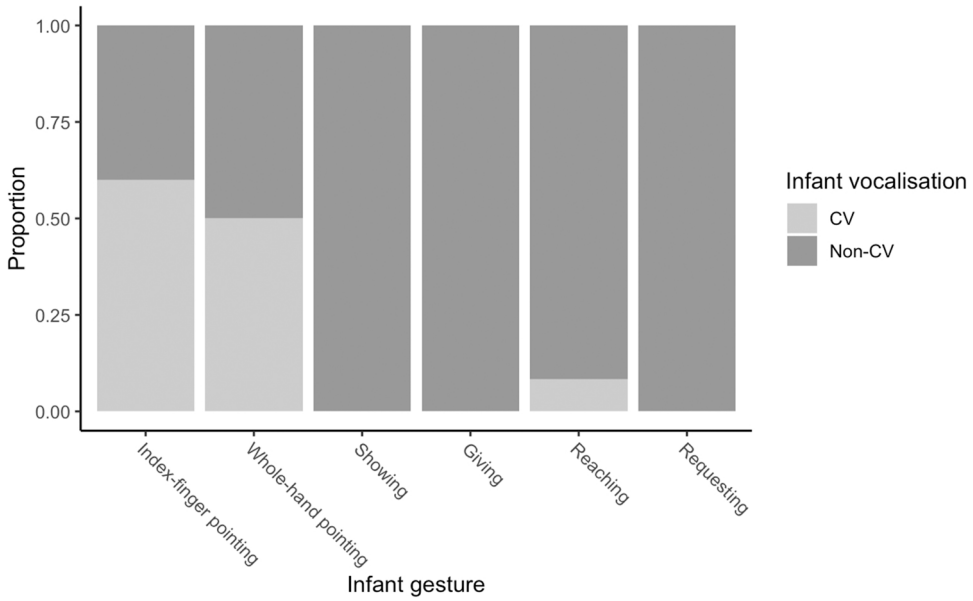

Fig. 1. Proportions of infants' bimodal gestures combined with CV and non-CV vocalisations.

3. Results

3.1. Infant behaviours

In total, we annotated 2036 infant behaviours of which 1892 were infant vocalisations. All infants included in this study produced at least one non-CV vocalisation during the session. Of all vocalisations, only 55 were classified as CV, consisting of at least one syllable that did not only involve a glide or glottal. Approximately twenty percent of the infants in this study produced at least one CV vocalisation. Most of them produced only a few, although one infant produced thirteen instances during the session.

We annotated 207 infant gestures. Slightly more than half (53 %) of the infants in our study produced at least one unimodal gesture. The most frequent infant gesture is reaching. We also annotated many instances of giving and index-finger pointing. Despite the young age, index-finger pointing was much more common than whole-hand pointing. The infants in this study did not spontaneously produce many showing or requesting gestures. The total frequencies of all infant behaviours are shown in Table 1.

We also examined whether infants produced bimodal behaviours (i.e., vocal-gestural combinations). In total, only 63 infant behaviours were bimodal. At least one bimodal behaviour was produced by a quarter (25.6 %) of the infants in this study. The gestures that were most often combined with a vocalisation were requesting (55.6 % of instances were bimodal), whole-hand pointing (50 %), and index-finger pointing (40 %). Nine (14.3 %) of the vocalisations in bimodal behaviours were CV vocalisations. In the full data set, only 2.8 % of all vocalisations were classified as CV vocalisation. Therefore, infants tended to produce more CV vocalisations in bimodal behaviours compared to unimodal vocalisations. Figure 1 shows the proportions of each infant gesture combined with CV and non-CV vocalisations. This shows that pointing gestures were more often combined with CV vocalisations compared to other gestures.

Table 2
Total frequencies of caregivers' verbal responses including production range and percentage of caregivers who produced the response.

Caregiver response	Frequency	Range	Percentage
Contingent	670	0–27	93.2
Non-contingent	726	0–23	94.0
Infant imitation	127	0–11	47.0
Onomatopoeia or sound effect	43	0–5	24.8
Total	1566		

Table 3
Total frequencies of caregivers' nonverbal responses including production range and percentage of caregivers who produced the response.

Caregiver response	Frequency	Range	Percentage
Gestural responses			
Pointing	61	0–5	29.1
Passing	74	0–6	36.8
Showing	101	0–22	36.8
Accepting	60	0–7	20.5
Representational	2	0–1	1.7
Object manipulation	388	0–13	83.8
Other	11	0–3	7.7
Total	697		
Facial responses			
Smile	168	0–10	65.0
Surprise	23	0–2	15.4
Total	191		
Bodily responses			
Leaning closer	26	0–3	17.9
Turning to infant	12	0–3	6.8
Turning to toy	6	0–2	4.3
Affective	36	0–3	20.5
Total	80		

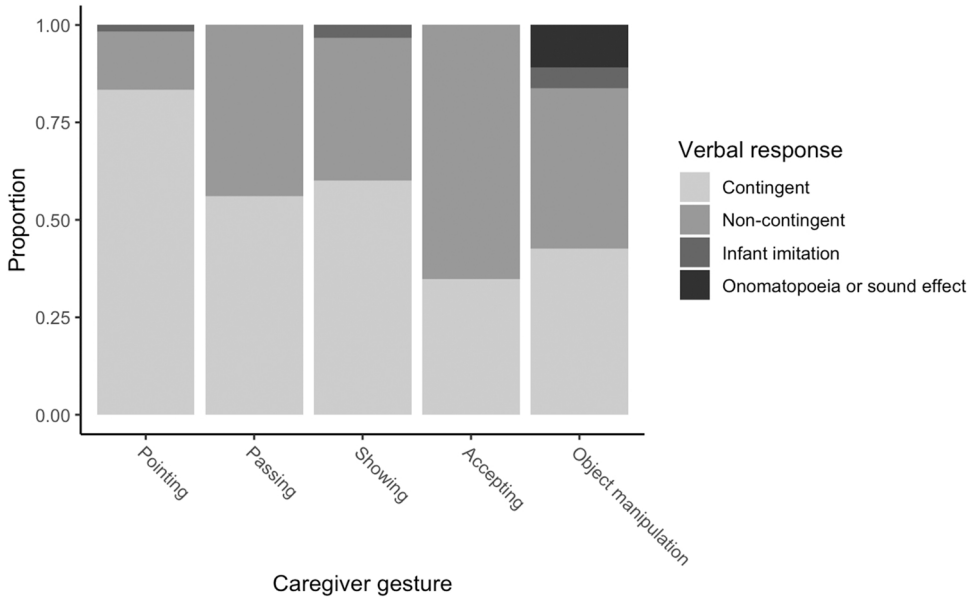

Fig. 2. Proportions of caregivers' multimodal gestures combined with different verbal response types.

Table 4
Percentages of infant behaviours that received a caregiver response of any type.

Vocalisations		Gestures	
	Percentage		Percentage
CV	87.3	Index-finger pointing	96.0
Non-CV	86.4	Whole-hand pointing	75.0
		Showing	94.4
		Giving	98.2
		Reaching	91.0
		Requesting	100.0
		Others	100.0

Table 5
Percentages of each infant behaviour receiving a caregiver verbal, nonverbal, and multimodal response.

Infant behaviour	Verbal responses	Nonverbal responses	Multimodal responses
Vocalisation	76.2	42.2	32.0
Gesture	79.2	72.2	55.6
Bimodal	92.1	60.3	60.3

3.2. Caregiver responses

In total, we annotated 2036 infant behaviours of which 87.1 % received a caregiver response of any type. A caregiver response could fall into multiple categories in the case of a multimodal response. Overall, caregivers produced more verbal than nonverbal responses. Table 2 shows verbal response frequencies, ranges, and percentages of caregivers who produced the response at least once. Most verbal responses are classified under contingent (e.g., talking about a toy that the infant is showing) or non-contingent (e.g., an affirmation) verbal responses, but we also annotated imitations of infant vocalisations and some onomatopoeias or sound effects (e.g., "broom broom").

We also annotated nonverbal responses shown in Table 3. We found many occurrences of manual gestures, such as object manipulations (e.g., riding the toy car around), and deictic gestures (e.g., showing or pointing). Deictic gestures were produced by approximately a third of the participants in this study. There were few representational gestures (e.g., demonstrating how the pop-up toy works without pressing the buttons) and only eleven gestures were classified as other, for example, waving or nodding. These categories were infrequent and therefore not included in further analyses. To a lesser extent, caregivers used their faces or bodies to respond to infants. Caregivers also frequently began to smile during or after the offset of the infant behaviour. Smiling as a response occurred in more than half of all caregivers in this study. There were not many occurrences of surprise, and we did not annotate any other facial expressions. The other non-facial and non-gestural bodily behaviours, such as a change of body orientation, did not occur frequently as a response to infants' vocalisations and gestures.

Of all 1774 caregiver responses, 39.7 % were multimodal. This indicates that a verbal response was accompanied by a gestural, facial, and/or bodily response – at least partially overlapping in time. Most caregivers (94 %) produced at least one multimodal response. There were only 209 unimodal nonverbal responses. Caregivers' gestural responses were most often multimodal out of all response categories (80.9 %). Almost all pointing gestures produced by caregivers (98.4 %) were multimodal. In contrast, less than half (45 %) of verbal responses were multimodal. Only onomatopoeias or sound effects show a high degree of multimodality (79.1 %). Semantically contingent (i.e., follow-in) verbal responses are more often produced multimodally (50.6 %) than non-contingent verbal responses (41.9 %). Infant imitations were less often multimodal (21.3 %). Lastly, other facial and bodily responses also occurred in high proportions of multimodal responses (58.6 % and 78.8 % respectively). Caregivers tend to combine nonverbal responses - mainly gestures and bodily behaviours – with verbal responses.

We further examined which caregivers' verbal and gestural responses often co-occurred. Fig. 2 shows that, although frequently occurring in a multimodal response, onomatopoeias or sound effects are only combined with object manipulations. Caregivers' pointing gestures are more often combined with a semantically contingent response. Caregivers' showing and passing gestures are also more likely to be combined with a semantically contingent response. Caregivers rarely use infant vocal imitations in a multimodal response.

3.3. Predicting response rates

We first compared response rates by calculating the proportions of responded-to behaviours for each infant vocalisation and gesture. Table 4 shows that infant gestures had higher overall response rates than infant vocalisations. All gestures examined in this study were highly frequently responded to. Next, we examined whether different categories of infant behaviours (vocalisations, gestures, and bimodal behaviours) elicited different proportions of verbal, nonverbal, and multimodal responses. Table 5 shows the percentages of each response category to each infant behaviour. The results show that infant vocalisations received more caregiver verbal responses than nonverbal or multimodal responses. Infant gestures elicited many verbal and nonverbal responses. Lastly, infant

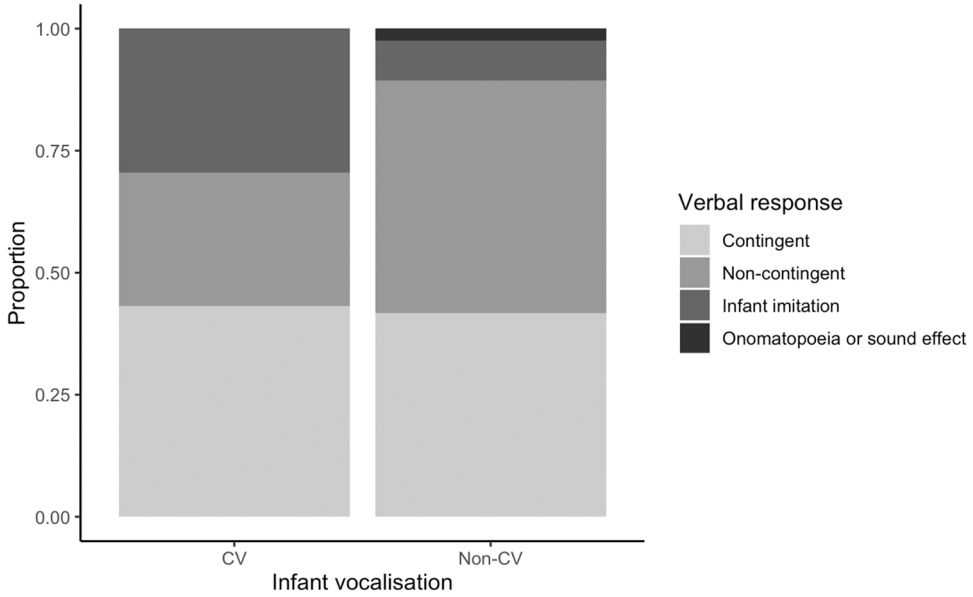

Fig. 3. Infants' vocalisations eliciting different proportions of verbal response types.

bimodal behaviours elicited the highest percentages of verbal and multimodal responses.

Three mixed-effects logistic regression models were fitted to test whether certain behaviours were statistically more likely to elicit (1) verbal responses, (2) nonverbal responses, and (3) multimodal responses. The first model indicates that infant bimodal behaviours are 6.42 times (95 % CI [2.39, 17.23]) more likely to receive a verbal response than infant vocalisations. Infant vocalisations and infant gestures do not significantly differ in eliciting verbal response rates. The second model shows that infant bimodal behaviours are 1.97 times (95 % CI [1.13, 3.43]) more likely to receive a nonverbal response, and infant gestures are 3.97 times (95 % CI [2.63, 5.99]) more likely to receive a nonverbal response compared to infant vocalisations. The last model reveals that infant bimodal behaviours are 3.48 (95 % CI [2.00, 6.06]) times more likely to elicit a multimodal response, and infant gestures are 2.96 (95 % CI [2.02, 4.34]) times more likely to receive a multimodal response than infant vocalisations.

We also examined whether specific types of vocalisations or gestures are more likely to elicit specific caregiver response types. None of the models including infant vocalisation (CV or non-CV) as a predictor reached significance, indicating that the two types of vocalisations do not differ in eliciting verbal, nonverbal, or multimodal response rates. The models including infant gesture type as the predictor variable indicate that the best predictor of a caregiver verbal response was infant index-finger pointing. Infant index-finger pointing was 14.82 times (95 % CI [1.25, 175.08]) more likely to elicit a caregiver verbal response compared to infant reaching. The other gestures did not significantly differ from infant reaching in eliciting verbal response rates. In contrast, infant giving gestures elicited higher nonverbal and multimodal response rates. Infant giving was 13.63 times (95 % CI [3.10, 59.85]) more likely to receive a nonverbal response, and 5.11 times (95 % CI [1.92, 13.65]) more likely to elicit a multimodal response compared to infant reaching. The other infant gestures did not significantly differ from reaching.

In sum, infant bimodal behaviours elicited more verbal and multimodal responses, while infant gestures elicited more nonverbal responses. Overall, infant vocalisations are less likely to elicit caregiver responses, but they received more verbal than nonverbal or multimodal responses. When examining the data in more detail, we found that verbal responses are more often elicited by infant index-finger pointing compared to other infant gestures. Lastly, we found that infant bimodal behaviours elicited more caregiver multimodal responses compared to unimodal vocalisations or gestures, although infant gestures elicited more multimodal responses than infant vocalisations.

3.4. Predicting caregiver response types

We also aimed to examine whether different infant vocalisations and gestures elicited different caregiver verbal, gestural, facial, and bodily response types. In the case of a significant difference, we also present figures showing the proportions of each response category in response to different infant vocalisations and gestures to visually examine the differences.

Although we showed earlier that different infant vocalisations did not elicit different response rates, we do find that infant vocalisations elicited different verbal response types ($X^2 = 27.35$, $p < .001$). Fig. 3 shows that caregivers verbally imitated CV vocalisations more often than non-CV vocalisations. Infant vocalisations did not elicit different gestural responses ($X^2 = 1.53$, $p = 0.82$), facial responses ($p = 0.06$), or bodily responses ($p = 1.0$). More data could be necessary for the latter two categories to detect small effects, but these initial results suggest that vocalisations did not elicit different nonverbal responses.

Next, we examined whether different infant gestures elicited different verbal and nonverbal response types. First, different infant

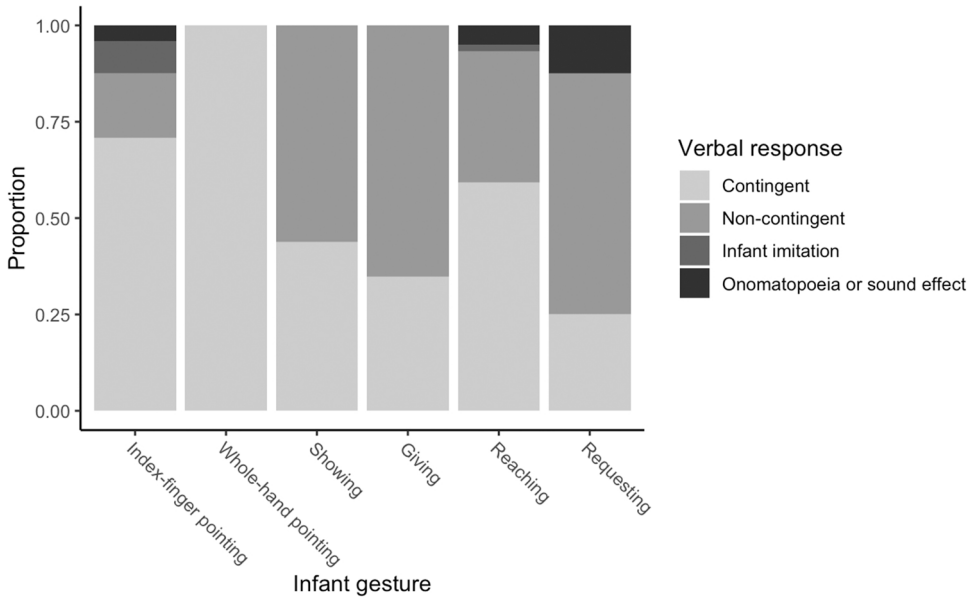

Fig. 4. Infants' gestures eliciting different proportions of verbal response types.

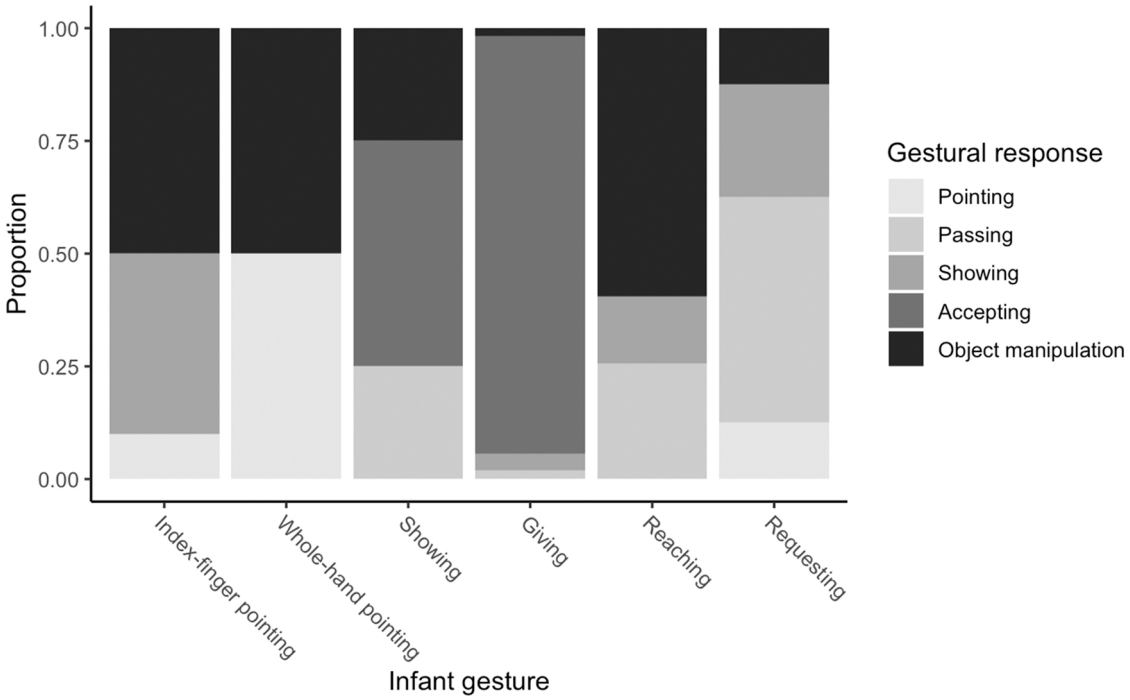

Fig. 5. Infants' gestures eliciting different proportions of gestural response types.

gestures elicited different verbal responses ($X^2 = 35.20$, $p < .01$). Fig. 4 shows that index-finger pointing, whole-hand pointing, and reaching elicited more contingent verbal responses (e.g., naming the object that the infant gesture was directed to), while all other gestures elicited more non-contingent verbal response (e.g., saying "oh nice" after the infant showed a toy).

Infant gestures also elicited different gestural responses ($X^2 = 157.07$, $p < .001$). Fig. 5 shows that infant giving was usually responded to by caregiver accepting. This is the most predictable caregiver gesture in response to any infant gesture. Infant requesting tended to elicit caregiver passing, and to a lesser extent caregiver pointing and caregiver showing. We see that infant pointing and requesting were the only types of gestures that elicited caregiver pointing in response. Infant showing elicited mostly caregiver accepting or object manipulations.

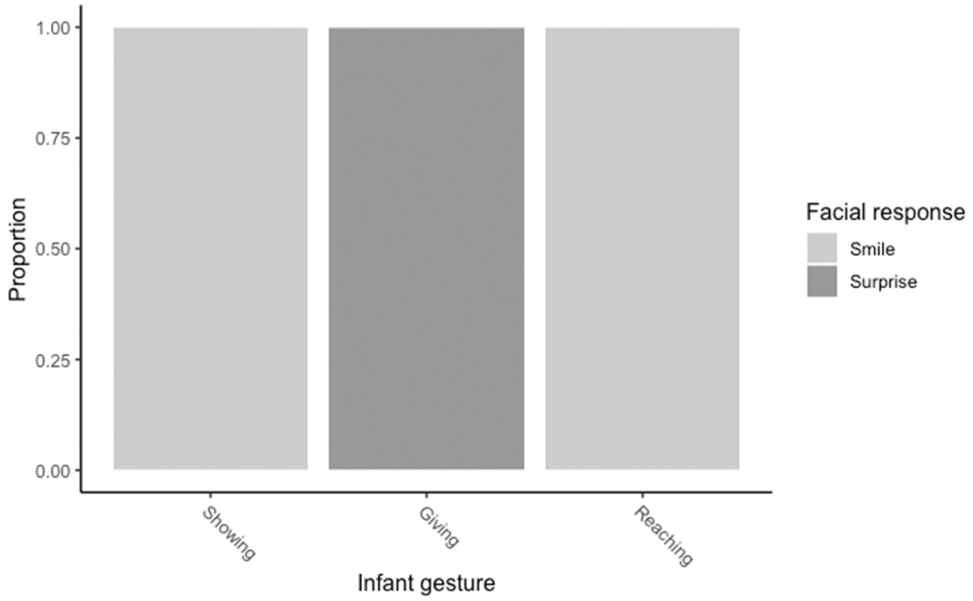

Fig. 6. Infants' gestures eliciting different proportions of facial response types.

Lastly, we found that infant gestures elicited different facial expressions ($p < .01$) but not different bodily responses ($p = 0.31$). Again, more data may be needed in the latter category to detect any small effects. Nevertheless, the Fisher's exact test does indicate a difference in eliciting facial expressions. Fig. 6 shows that infant showing and reaching only elicited smiles, while infant giving only elicited surprise. This is the first initial evidence that caregivers use different facial expressions to respond to different infant gestures, further suggesting that infants play an important role in the type of verbal and nonverbal language they receive.

4. Discussion

In the present study, we developed a new coding scheme to annotate caregivers' verbal, nonverbal (i.e., gestural, facial, and bodily), and multimodal responses to infants' vocalisations and gestures. The first aim was to determine which infant vocalisations, gestures, and bimodal behaviours and which caregiver verbal, nonverbal, and multimodal responses occur in a large, naturalistic data set. The second aim was to assess whether different infant behaviours elicited specific caregiver response rates and types.

4.1. Variability in infant behaviours

Most infant behaviours were non-CV vocalisations, and they were produced at least once by all infants in our sample. This is to be expected, as this behaviour emerges early in development. We do find large individual variability across infants in the number of productions during the annotated six minutes. Many infants only produced up to ten non-CV vocalisations, while others produced over fifty. To a lesser extent, infants produced CV vocalisations. We only annotated 55 CV vocalisations in total, produced by only one-fifth of the infants in our sample. This is considerably less than the number of infants who produced CV vocalisations in the study by Donnellan et al. (2019), who found that 97 % of 11-month-olds produced at least one CV vocalisation. The difference could be explained by the ages of the infants. The infants in their study were on average one month older, which is an important difference in the development of early vocalisations. Another explanation could be the duration of the interactions. In the study by Donnellan et al. (2019), caregiver-infant dyads were observed for 10–15 min which is much longer compared to the six minutes in the present study, giving them more time to produce less frequent behaviours.

The most frequent infant gestures in the current study were reaching, giving, and pointing. We found that reaching is the most frequent gesture overall, and it is produced by the largest group of infants in our study. Although less frequent than reaching, giving shows the largest individual variability across the infants in the current study. Although most infants did not produce these gestures even once, we find some infants who produced them six or seven times in the annotated session. It may be surprising that we found more occurrences of index-finger pointing than showing, although the latter shows more individual variability. The present study was not designed to elicit pointing gestures. Yet, it was the third most frequent gesture among the infants in the current study. In the study by Donnellan et al. (2019), index-finger pointing was also produced by 21 % of infants in their study, the largest group after giving and showing. Once infants have acquired index-finger pointing, they may rapidly start to produce many of them. Index-finger pointing has a clear deictic function which may be useful for prelinguistic infants to refine their early communicative bids.

We also examined which infant gestures were frequently combined with infant vocalisations, resulting in bimodal behaviours. Although the 10-month-olds in the current study did not produce many bimodal behaviours, the results show that infant requesting

and pointing gestures have the largest proportions of bimodal productions. Approximately half of the total productions were combined with infant vocalisations. Pointing and requesting are among the later emerging infant gestures (Frank et al., 2021). The results of this study suggest that these later emerging gestures are more often combined with vocalisations. Infants only start producing bimodal behaviours later in development. Possibly, when infants are ready to produce pointing and requesting gestures, they are also ready to produce bimodal behaviours.

4.2. Caregivers' verbal, nonverbal, and multimodal responses

More than three-quarters of infant behaviours received a verbal response, making verbal responses the most common type of caregiver response. Caregivers tended to produce many contingent and non-contingent verbal responses. We also annotated some imitations of infant vocalisations, as well as onomatopoeias or sound effects. Regarding nonverbal responses, we annotated many gestural responses. We did not specifically elicit gestures during the caregiver-child interaction task, so these results suggest that caregivers tend to use many gestures naturally during early interactions with their infants. We annotated fewer facial and bodily responses. This could be due to the criteria that we set for annotating a facial response: A facial expression was only annotated if the caregiver was not already showing the facial expression prior to the infant behaviour. We did find many smiles, compared to other facial responses, which is characteristic of infant-directed speech (Benders, 2013). We did not find many bodily responses which negatively affected the coding reliability of these responses. Nevertheless, there was high agreement between the two annotators on the few bodily responses that did occur. Therefore, it is likely that caregivers' vocalisations and gestures are more often used to respond to infants, while other non-gestural bodily behaviours naturally occur during interactions but may not be used specifically to respond to infants' vocalisations and gestures.

We also examined which verbal and nonverbal responses are often combined into multimodal responses. Of all verbal responses, approximately 40 % were multimodal. Less than 10 % of caregiver responses were nonverbal without any verbal component. This result agrees with the findings by Ger et al. (2018) who reported that approximately 40 % of caregiver responses to infant index-finger pointing gestures were multimodal. We extended this finding to caregiver behaviours in response to all infant vocalisations and gestures. In a recent study, Chen et al. (2021) showed that caregivers often touched objects while naming them. Caregivers did this more often when the object was unfamiliar to the child, suggesting that they may do this to reduce referential ambiguity in learning contexts. Although we did not compare familiar to unfamiliar toys, our study extends this finding by showing that all caregivers' deictic gestures and object manipulations often co-occur with verbal responses. Caregivers' pointing gestures were most often multimodal. They were also more often combined with a semantically contingent verbal response compared to other gestures. This could be due to the deictic function of pointing. The contents of a semantically contingent response are often about one of the toys in the room. Caregivers frequently produced, for example, a labelling response while simultaneously pointing at the labelled object. In addition, caregivers' showing gestures were, although to a lesser extent, also more often combined with contingent verbal responses. A showing gesture also has a more deictic function than passing or accepting a toy. We also found that only object manipulations are combined with onomatopoeias or sound effects. Caregivers tended to produce onomatopoeias or sound effects while playing with a toy, such as moving the toy car while saying "broom broom". This non-arbitrary connection could help the infant to connect the linguistic sound to the referent, aiding vocabulary development. Indeed, onomatopoeias are generally highly present in infant-directed speech (Motamedi et al., 2021). These visual cues, in the form of deictic gestures or object manipulations, are characteristic of infant-directed speech and may aid the infant's interpretation of unknown words (e.g., Baldwin et al., 1996; Gogate et al., 2000). Previous studies have predominantly examined caregivers' verbal responses, but the results of the current study suggest that it is important to also analyse gestural responses to gain a more complete picture of caregivers' responses.

4.3. Infant behaviours elicit specific responses

Our study found that different infant behaviours elicited different verbal, nonverbal, and multimodal responses. Different vocalisations by the infant elicited different types of caregiver verbal responses. Caregivers imitated infants' CV vocalisations significantly more often than non-CV vocalisations, as previously found by Gros-Louis et al. (2006). Vocalisations did not elicit different nonverbal responses. Infant gestures, however, elicited different verbal, gestural, and facial responses. For example, infant index-finger pointing gestures elicited more semantically contingent verbal responses compared to the other infant gestures. Earlier research found that mothers provide more object labels to gestures compared to non-gestures (Olson & Masur, 2013) and that mothers are specifically more likely to provide object labels to infant index-finger pointing compared to infant reaching or infant vocalisations (Kishimoto et al., 2007; Wu & Gros-Louis, 2015). The results of the current study show that infant index-finger pointing gestures are often bimodal, and bimodal behaviours elicited the highest proportion of verbal responses. It could be possible that index-finger pointing has clear communicative intent, particularly when coordinated with a vocalisation, thereby eliciting more fine-tuned responses from caregivers. We also see many predictable patterns in the type of caregiver gestures that are elicited by infants. For example, infant giving and showing gestures usually elicited caregiver accepting, while infant reaching and requesting elicited higher rates of caregiver passing. Lastly, it appears that when caregivers accept toys from their infants, they tend to show a "mock surprised" facial expression. They did not show this expression in response to any other infant gesture. This suggests that even caregivers' facial expressions are dependent on

infant behaviours. These results show the richness of multimodal communication between infants and their caregivers with highly predictable patterns within and across modalities.

Predictable infant behaviour and caregiver response patterns may highlight synchrony between infants and their caregivers. Synchrony can involve matching behaviours, such as smiling simultaneously. Previous studies showed that caregivers use more verbal than nonverbal responses to infant vocalisations, and infants' multimodal behaviours (looks combined with touch) elicited more caregivers' multimodal responses (looks combined with touch) (Gros-Louis et al., 2006; Yurkovic et al., 2021). This points to synchrony through matching modalities. That is, a co-occurrence of modality could suggest high synchrony between children and their caregivers (see Leclère et al., 2014). Yet, these studies only examined infant vocalisations, looking behaviour, and touch. Our results add to these previous studies by showing that infant gestures were the best predictors of a nonverbal response, while infant bimodal behaviours elicited more multimodal responses compared to other behaviours. We did not find that infant vocalisations are the best predictors of a verbal response, because infant vocalisations elicited fewer caregiver responses in general. Nevertheless, of all caregiver responses that were elicited by infant vocalisations, the majority were indeed verbal. We thus extend previous findings regarding matched modalities to a wider set of behaviours, including infant vocalisations and gestures, as well as more caregiver verbal and nonverbal responses. It appears that caregivers tend to respond using the same modality as the infant behaviour.

The bidirectional approach taken in the current study poses new questions regarding the large variability in caregiver responsiveness. Previous studies have shown that caregivers individually differ in their verbal responsiveness which positively relates to children's socio-cognitive outcomes, including language (e.g., McGillion et al., 2013; Olson & Masur, 2015; Wu & Gros-Louis, 2014). Caregivers differ in the number of verbal responses they tend to produce. For example, mothers are more likely to respond verbally than fathers (Flippin and Watson, 2011), and mothers of higher SES verbally respond more often than mothers of lower SES (e.g., Hart & Risley, 1995; McGillion et al., 2017; Vanormelingen & Gillis, 2016). Our results add to these previous studies by showing that responsiveness does not depend solely on caregiver characteristics. Infants influence the rates and types of responses they elicit from their caregivers by producing different types of vocalisations and gestures. More specifically, infants who produce many gestures and bimodal behaviours tend to elicit more caregiver responses than infants who predominantly produce vocalisations. This shifts the attention from individual behaviours to the dyad (Renzi et al., 2017; Chen et al., 2021). The question remains to what extent caregivers' responses in turn reinforce infants' behaviours. We should not study infant behaviours or caregiver responses in isolation to understand variability during early caregiver-child interactions, but rather we should examine the bidirectional effects that infants have on their caregivers and vice versa.

4.4. Limitations and future directions

The present study characterised infants' behaviours and caregivers' responses during free play. Caregivers likely use different types of cues in different learning environments. For example, some cues, such as body orientation, may become more important when objects are not in close proximity to the child and the caregiver. We may also expect infants to change their behaviours in different contexts. During book reading, infant index-finger pointing may be used more often than infant showing and passing, while the latter were most frequent during free play. Subsequently, caregivers' responsiveness will likely also be affected by the change in infant behaviours. Future studies can examine whether the predictable patterns between infant behaviours and caregiver responses remain stable, or whether we find differences in both infant behaviours, as well as caregiver response rates and types, across different learning environments.

The present study did not annotate nonverbal behaviours based on their contingency. In a recent study, Ger et al. (2018) have shown the important role played by contingent versus non-contingent responses that were measured both verbally and nonverbally. Future studies can examine the extent to which both verbal and nonverbal responses are semantically contingent on infants' behaviours. Previous studies have highlighted the important role played by intersensory redundancy in, for example, word learning (Gogate & Bahrick, 1998). Redundancy implies that there is some overlap in meaning between the verbal and nonverbal behaviour. Hence, in the case of redundancy, if the verbal response is contingent, the nonverbal behaviour should also be contingent. The question remains what proportion of multimodal responses typically contains redundant information.

We lastly want to acknowledge that the development of the coding scheme was highly informed by previous studies and empirical observations in the present study of western, educated, industrialised, rich, and democratic (WEIRD) caregivers. We recommend caution when using this coding scheme to annotate interactions of non-WEIRD dyads since the coding scheme reflects many cultural phenomena that may not be universal across cultures. By including onomatopoeias and nonverbal responses when examining caregivers' responsiveness, we already encompass more cultural diversity since caregivers from diverse cultures may produce different types of vocalisations, for example, Japanese mothers tend to produce far more onomatopoeias (Fernald & Morikawa, 1993), and Chinese mothers produce more pointing gestures compared to American mothers (So & Lim, 2012). However, it is always good to keep in mind that certain behaviours may occur in other cultures that are not included in the current coding scheme, but which do contribute importantly to caregiver responsiveness. It is important to validate a measuring instrument for specific populations.

4.5. Conclusions

This study provides an overview of caregivers' verbal, nonverbal, and multimodal responses to their 10-month-old infants' communicative behaviours in a large, naturalistic data set. During free play, caregivers most often produced verbal responses, but approximately 40 % of those were multimodal. Caregivers often coordinated speech with manual and deictic gestures, and to a lesser extent with facial expressions and other bodily behaviours. Multimodal responses could be useful in learning contexts as they provide

children with useful cues to disambiguate novel or unclear speech. We also examined whether different infant behaviours elicited different caregiver verbal, nonverbal, and multimodal responses. Infant bimodal (i.e., vocal-gestural combination) behaviours elicited high rates of verbal and multimodal responses, while unimodal gestures elicited high rates of nonverbal responses. We also found that different infant vocalisations elicited different verbal responses, while different infant gestures elicited different verbal, gestural, and facial responses. The results indicate that infants show large variability in the frequency and types of vocalisations and gestures they produce, which in turn affect when and how their caregivers respond. When examining caregiver-child interactions, analysing caregivers' verbal responses alone undermines the multimodal richness and bidirectionality of early communication.

CRediT authorship contribution statement

Anika van der Klis: Conceptualization, Methodology, Validation, Formal analysis, Visualization, Writing - original draft. **Frans Adriaans:** Conceptualization, Methodology, Supervision, Writing - review & editing. **René Kager:** Supervision, Conceptualization, Methodology, Funding acquisition, Writing - review & editing.

Declarations of interest

None.

Data availability

YOUth is a longitudinal study that aims to produce and safely store FAIR and high-quality data. The data can be accessed for both use and verification purposes upon request (see https://www.uu.nl/en/research/youth-cohort-study/data-access). All other materials, detailed coding instructions, and *R* scripts are available online: https://osf.io/nvm54/.

Acknowledgements

We would like to thank Joyce van Zwet for her help with data annotation. We are also grateful to all families who participate in the YOUth study. YOUth is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003). A complete listing of the study investigators and study management can be found at https://www.uu.nl/en/research/youth-cohort-study/about-us/who-is involved. YOUth investigators and management designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the YOUth study investigators or YOUth management.

Appendix A. - Coding scheme

See Appendix Tables A1, A2, A3, A4, A5, A6.

Infant vocalisations and gestures are annotated independently. Whenever the infant combines or overlaps a vocalisation and a gesture, annotate both behaviours and start measuring two seconds after the offset of the last behaviour for the caregiver's response. If the infant produces the same vocalisation or gesture again within these two seconds, we only annotate a caregiver response once, unless the caregiver gives two separate responses. If the infant produces a vocalisation and a gesture which overlap in time, we measure two seconds after the onset of the last behaviour (i.e., we annotate only one "no response" when the infant produces two pointing gestures within two seconds or when a vocalisation and gesture overlaps in time and their caregiver did not respond to either of them). This was done to allow caregivers enough time to respond to the behaviours, while some infants tend to repeat vocalisations continually.

Infant vocalisations

An infant vocalisation is any sound produced by the infant except vegetative sounds (e.g., hiccoughs or coughs) or distress sounds (e.g., crying or fussing). The two types of infant vocalisations are defined in Table A1.

Infant gestures

We coded eight types of infant gestures, following previous studies (e.g., McGillion et al., 2013; Wu & Gros-Louis, 2014; Olson & Masur, 2015; Donnellan et al., 2019) and items defined in the CDIs (Fenson et al., 2007). We initially distinguished between three types of pointing gestures: index-finger pointing, whole-hand pointing, and any other precursor of pointing. We did not annotate any occurrences of the latter category, so this was excluded from further analyses. The beginning of each gesture should be marked at the frame where the arm reached maximum extension, and the end should be marked at the frame where retraction of the arm begins. If the arm is extended within 200 ms of the previous arm retraction, and the infant produces the same gesture again, this counts as a single occurrence (following Donnellan et al., 2019). All gestures are defined in Table A2.

Table A1
Coding scheme for infant vocalisations.

Infant vocalisation	Definition
Consonant-Vowel (CV)	At least one syllable contains a consonant-vowel sequence ("baba", "ma" etc.), excluding glides ("ja") and glottals ("ha") (e.g., Donnellan et al., 2019).
Non-CV	All other types of vocalisations except vegetative sounds (e.g., hiccoughs or coughs) or distress sounds (e.g., crying or fussing).

Table A2
Coding scheme for infant gestures.

Infant gesture	Definition
Index-finger pointing	Infant extends their index-finger in the direction of the object or event while the other fingers are partially or entirely curled back while looking at an object or event. The arms must be extended, with empty hands, and the child should not lean forward or touch the object (following McGillion et al., 2017).
Whole-hand pointing	Infant extends a majority of fingers in the direction of the object or event (Donnellan et al., 2019).
Other pointing	Infant produces another precursor of pointing: e.g., fist extension or thump extension in the direction of the object or event.
Showing	Infant holds out an object with extended arm(s) towards the caregiver's face (adapted from Masur, 1982).
Giving	Infant holds out an object with either (or both) arms extended towards the caregiver's hands or in a way as to deliver the object to the caregiver.
Reaching	Infant extends either (or both) hands to get to an object out of reach. In this case, the infant may lean forward. Excluding movements that were the first phase of grasping an object already within reach (Masur, 1982). If the infant starts moving closer to the object to eventually grasp it, still annotate reaching before the infant started moving.
Requesting	Infant extends either (or both) hands to get an object out of reach without leaning towards it. The infant may open and close their hand.
Other	Infant produces any remaining conventional gestures, such as the infant raising their arms to initiate being picked up, waving, shrugging, nodding "yes" or "no", or blowing a kiss.

Table A3
Coding scheme for verbal responses.

Verbal response	Definition
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 (Donnellan et al., 2019). The utterance refers to an object that the child is holding, looking at, or has referenced by a gesture, or if the utterance is related to the activity in which the child is engaged (following McGillion et al., 2017).
Non-contingent	Other types of non-contingent responses, for example, affirmations ("good job!"), routines ("peek-a-boo"), directive acts ("Now get the doll", if the doll was not the infant's current focus of attention), and other questions not specifically about the current focus of attention ("What do you want next?").
Onomatopoeias or sound effects	Onomatopoeias ("knor knor") or other sound effects such as noises made with the mouth that represent a sound (e.g., snorting like a pig or making the noise of drinking something) (Vigliocco et al., in prep).
Infant imitation	Child vocal imitations.

Table A4
Coding scheme for gestural responses.

Gestural response	Definition
Pointing	Caregiver points towards an object or event.
Passing	Caregiver gives a toy to the infant.
Showing	Caregiver holds out an object with either (or both) hands to show the object to the infant without manipulating it.
Accepting	Caregiver accepts the toy that the infant is giving by grabbing it out of the infant's hand(s).
Representational	Caregiver shows the size, shape or how an object works without the object in hand (e.g., pretending to drink from a bottle or holding up hands to demonstrate the size) (Vigliocco et al., in prep).
Object manipulation	Caregiver depicts how to use an object or imitate how object moves or act (e.g., letting the baby doll drink from the bottle or showing how to use the pop-up toy) physically with the object (Vigliocco et al., in prep).
Other	Any remaining conventional gestures, such as picking the infant up, waving, shrugging, nodding "yes" or "no", or blowing a kiss.

Table A5
Coding scheme for facial responses.

Facial response	Definition
Smiling	Caregiver shows a happy expression, typically with the corners of the mouth turned up.
Surprise	Caregiver shows a surprised expression, typically with eyebrows raised and jaw dropped down.
Other	Caregiver shows any other facial expression different from neutral.

Table A6
Coding scheme for bodily responses.

Bodily response	Definition
Leaning closer	Caregiver comes closer in proximity to the infant.
Turning to infant	Caregiver turns their head or full body towards the infant when the caregiver was facing elsewhere.
Turning to toy	Caregiver turns their head or full body towards the toy when the caregiver was facing elsewhere.
Affective language	Caregiver shows positive affect towards the infant by cuddling, patting, kissing, or caressing the infant's cheek.
Other	Caregiver shows another clear non-gestural bodily reaction.

Caregiver responses

Responses should be temporally contingent and can occur during the infant behaviour or within two seconds after the offset of the infant behaviour (McGillion et al., 2013). A caregiver response occurs when the caregiver produces any type of verbal and/or nonverbal behaviour within this time frame. Each category is coded independently. All verbal responses are defined in Table A3. For nonverbal responses, we distinguished between gestural, facial, and bodily responses. These can be found in Table A4, A5 and A6, respectively.

References

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, 67*(6), 3135–3153. https://doi.org/10.2307/1131771

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior & Development, 36*(4), 847–862. https://doi.org/10.1016/j.infbeh.2013.09.001

Bornstein, M. H., & Tamis-LeMonda, C. S. (1989). Maternal responsiveness and cognitive development in children. *Maternal responsiveness: Characteristics and consequences* (pp. 49–61). Jossey-Bass.

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

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language, 35*(1), 207–220. https://doi.org/10.1017/s030500090700829x

Chen, C., Houston, D. M., & Yu, C. (2021). Parent–child joint behaviors in novel object play create high-quality data for word learning. *Child Development, 92*(5), 1889–1905. https://doi.org/10.1111/cdev.13620

Choi, B., Wei, R., & Rowe, M. L. (2021). Show, give, and point gestures across infancy differentially predict language development. *Developmental Psychology, 57*(6), 851–862. https://doi.org/10.1037/dev0001195

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

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2019). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science, 23*(1), 1–21. https://doi.org/10.1111/desc.12843

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *The MacArthur Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Paul H. Brookes Publishing Co., Inc.,.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5). https://doi.org/10.2307/1166093

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development, 64*(3), 637–656. https://doi.org/10.2307/1131208

Flippin, M., & Watson, L. R. (2011). Relationships between the responsiveness of fathers and mothers and the object play skills of children with autism spectrum disorders. *Journal of Early Intervention, 33*(3), 220–234. https://doi.org/10.1177/1053815111427445

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project.* MIT Press,. ⟨https://langcog.github.io/wordbank-book/⟩.

Ger, E., Altınok, N., Liszkowski, U., & Küntay, A. C. (2018). Development of infant pointing from 10 to 12 months: The role of relevant caregiver responsiveness. *Infancy, 23*(5), 708–729. https://doi.org/10.1111/infa.12239

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology, 69*(2), 133–149. https://doi.org/10.1006/jecp.1998.2438

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development, 71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197

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

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science, 13*(1), 252–263. https://doi.org/10.1111/j.1467-7687.2009.00871.x

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30*(6), 509–516. https://doi.org/10.1177/0165025406071914

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing,.

Holle, H., & Rein, R. (2015). EasyDIAg: A tool for easy determination of interrater agreement. *Behavior Research Methods, 47*(3), 837–847. https://doi.org/10.3758/s13428-014-0506-7

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Kishimoto, T., Shizawa, Y., Yasuda, J., Hinobayashi, T., & Minami, T. (2007). Do pointing gestures by infants provoke comments from adults. *Infant Behavior and Development, 30*(4), 562–567. https://doi.org/10.1016/j.infbeh.2007.04.001

Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., & Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction, 13*, 1–9. https://doi.org/10.1016/j.ijcci.2017.04.001

Leclère, C., Viaux, S., Avril, M., Achard, C., Chetouani, M., Missonnier, S., & Cohen, D. (2014). Why synchrony matters during mother-child interactions: A systematic review. *PLOS One, 9*(12), Article e113571. https://doi.org/10.1371/journal.pone.0113571

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science, 7*(3), 297–307. https://doi.org/10.1111/j.1467-7687.2004.00349.x

Masur, E. F. (1982). Mothers' responses to infants' object-related gestures: Influences on lexical development. *Journal of Child Language, 9*(1), 23–30. https://doi.org/10.1017/S0305000900003585

McGillion, M. L., Pine, J. M., Herbert, J. S., & Matthews, D. (2017). A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *Journal of Child Psychology and Psychiatry, 58*(10), 1122–1131. https://doi.org/10.1111/jcpp.12725

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. *IEEE Transactions on Autonomous Mental Development, 5*(3), 240–248. https://doi.org/10.1109/TAMD.2013.2275949

Motamedi, Y., Murgiano, M., Perniss, P., Wonnacott, E., Marshall, C., Goldin-Meadow, S., & Vigliocco, G. (2021). Linking language to sensory experience: Onomatopoeia in early language development. *Developmental Science, 24*(3), Article e13066. https://doi.org/10.1111/desc.13066

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating language in the real-world: The role of multimodal iconicity and indexicality. *Journal of Cognition, 4*(1), 38. https://doi.org/10.5334/joc.113

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

Onland-Moret, N. C., Buizer-Voskamp, J. E., Albers, M. E. W. A., Brouwer, R. M., Buimer, E. E. L., Hessels, R. S., de Heus, R., Huijding, J., Junge, C. M. M., Mandl, R. C. W., Pas, P., Vink, M., van der Wal, J. J. M., Hulshoff Pol, H. E., & Kemner, C. (2020). The Youth study: Rationale, design, and study procedures. *Developmental Cognitive Neuroscience, 46*, Article 100868. https://doi.org/10.1016/j.dcn.2020.100868

Pearson, R. M., Heron, J., Melotti, R., Joinson, C., Stein, A., Ramchandani, P. G., & Evans, J. (2011). The association between observed non-verbal maternal responses at 12 months and later infant development at 18 months and IQ at 4 years: A longitudinal study. *Infant Behavior & Development, 34*(4), 525–533. https://doi.org/10.1016/j.infbeh.2011.07.003

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing,. ⟨https://www.R-project.org/⟩.

Renzi, D. T., Romberg, A. R., Bolger, D. J., & Newman, R. S. (2017). Two minds are better than one: Cooperative communication as a new framework for understanding infant language learning. *Translational Issues in Psychological Science, 3*, 19–33. https://doi.org/10.1037/tps0000088

Rowe, M. L., Wei, R., & Salo, V. C. (2022). Early gesture predicts later language development. *Gesture in language: Development across the lifespan* (pp. 93–111). American Psychological Association,. https://doi.org/10.1037/0000269-004

Ruddy, M. G., & Bornstein, M. H. (1982). Cognitive correlates of infant attention and maternal stimulation over the first year of life. *Child Development, 53*(1), 183–188. https://doi.org/10.2307/1129651

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category: ELAN and ISO DCR. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: LREC 2008,.

So, W. C., & Lim, J. Y. (2012). "What is this?" Gesture as a potential cue to identify referents in discourse. *Applied Psycholinguistics, 33*(2), 329–342. https://doi.org/10.1017/S0142716411000373

Tamis-LeMonda, C. S., & Bornstein, M. H. (2002). Maternal responsiveness and early language acquisition. In R. V. Kail, & H. W. Reese (Eds.), *Advances in Child Development and Behavior* (Vol. 29, pp. 89–127). JAI.

von Eye, A., & von Eye, M. (2008). On the marginal dependency of Cohen's k. *European Psychologist, 13*(4), 305–315. https://doi.org/10.1027/1016-9040.13.4.305

Vanormelingen, L., & Gillis, S. (2016). The influence of socio-economic status on mothers' volubility and responsiveness in a monolingual Dutch-speaking sample. *First Language, 36*(2), 140–156. https://doi.org/10.1177/0142723716639502

Verhagen, J., van den Berghe, R., Oudgenoeg-Paz, O., Küntay, A., & Leseman, P. (2019). Children's reliance on the non-verbal cues of a robot versus a human. *PLOS One, 14*(12), Article e0217833. https://doi.org/10.1371/journal.pone.0217833

Vigliocco, G., Gu, Y., Donnellan, E., Grzyb, B., Brekelmans, G., Murgiano, M., Motamedi, Y., Brieke, R., & Perniss, P. (in preparation). The Ecological Language (ECOLANG) corpus of multimodal dyadic communication.

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C.R., Milan Maillo, I., & Perniss, P., (2019). Onomatopoeias, gestures, actions and words in the input to children: How do caregivers use multimodal cues in their communication to children? In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*.

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. (2015). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *Journal of Child Language, 42*(3), 538–561. https://doi.org/10.1017/S0305000914000221

Yurkovic, J. R., Kennedy, D. P., & Yu, C. (2021). Multimodal behaviors from children elicit parent responses in real-time social interaction. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*(43). ⟨https://escholarship.org/uc/item/2qw0d216⟩.