Abstract: Deixis - a fundamental part of communication involves combinations of speech, gesture and eye-gaze, yet little is known about the temporal dynamics of this coordination. We analysed eye-gaze, pointing gestures and verbal productions in 514 deictic episodes during triadic, semi-naturalistic, book-reading sessions performed by Italian children (1;08-2;07) and their caregivers. Results show three new findings. First, deictic communication is overwhelmingly preceded and accompanied by shared attention (of consistent duration) on an object, and only sometimes by disjoint attention. Second, children are synchronously multimodal (conveying information via speech, pointing gesture and eye-gaze) in their deictic communications. Third, the form of deictic communication used is not related to the complexity of the linguistic structures of the sample. Deictic communication is remarkably consistent in children ranging from approximately 1;08 to 2;07 years of age.

Introduction

Deixis is one of the most common strategies we use to manipulate an interlocutor's focus of attention during interaction (Diessel, 2006; Kita, 2003; Stukenbrock, 2015). Deixis projects a reference, usually from a speaker (*origo*, Bühler, 1934) to an intended referent, with interpretation related to the extra-linguistic, or meta-linguistic, context of production. A speaker can convey deictic reference through different modalities: speech (demonstratives, *this/that/here/there*), eye-gaze contact, and gestures (Iverson & Capirci, 1994; Liszkowski, Brown, Callaghan, Takada, & De Vos, 2012). These modalities operate in concert and share deictic information (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Talmy, 2018). For example, imagine a speaker telling an interlocutor: "*The cup you are looking for is over there, on that table!*". She produces the sentence while performing a pointing gesture towards the intended referent (i.e., cup/table); and during this communication she first looks at the interlocutor, then at the referent, and finally at the interlocutor again. The spatial information conveyed by the demonstratives *there* and *that* is consistent with the vectoral and manual trajectory marked by the communicative pointing gesture, as well as with eye-gaze contact (Fricke, 2014; Iverson, Tencer, Lany, & Goldin-Meadow, 2000; Kita, 2003). In the case of the pointing gesture, the extension of the arm and index finger draws a line originating from the speaker, ending at the location of the target (i.e. the cup on the table). The same deictic path may also be drawn by the vectoral visual trajectory established by gaze alternation during interactions, when directing the interlocutor's focus or 'back-checking' on the previously shared reference.

The frequent co-occurrence of demonstratives, pointing gestures and eye-gaze supports the hypothesis that both verbal and nonverbal modalities are used to establish and manipulate joint attention (Kita, 2003). It has been suggested that this function is

crucial in the early stages of language acquisition (Clark & Sengul, 1978; Diessel, 2006). However, how these modalities are actually organized during deictic communication has not thus far been investigated. Our main goal was to investigate whether joint attention is a *precondition* for deictic communication, or whether deictic communication may also be used to more broadly orient the attention of a disengaged hearer. Moreover, given that joint attentional behaviours are pivotal in early language acquisition (Carpenter et al., 1998; Kita, 2003), we examined how children coordinate verbal and nonverbal aspects of communication, and specifically whether they convey deictic information in a unimodal or multimodal manner. For multimodal deictic productions, we further aimed to establish whether verbal and nonverbal components are synchronized during production, either appearing at the same time, or operating independently from one another.

In typically developing children, joint attentional behaviours stabilize between 6 and 18 months (Bruinsma, Koegel, & Koegel, 2004; Mundy et al., 2007). During early caregiver-child interactions, caregivers typically manipulate the focus of attention of a child by pointing to a referent (initiating a joint attentional behaviour), with the child shifting gaze to follow the caregiver's pointing gesture (Shaw, Bryant, Malle, Povinelli, & Pruett, 2017). These early interactions are often multimodal, with caregivers typically using pointing gestures, to disambiguate verbally conveyed referents, and eye-gaze contact, to control for the establishment (and maintenance) of joint attention (Iverson & Capirci, 1994). Children also combine eye-gaze contact with pointing gestures, representational words and additional deictic expressions quite early in language acquisition (Bates, Camaiori, & Volterra, 1975; Clark, 1978; Diessel, 2006; Iverson & Capirci 1994).

Individually, modalities follow a similar developmental trajectory, each progressing from sharing, to following and later to directing attention and behaviour (Carpenter et al., 1998). At the same time children's interactions develop from being mainly dyadic to triadic. During dyadic interaction the child interacts with either an object/event or an adult in isolation; during triadic interactions the child interacts with an adult and an object/event simultaneously. Capirci, Iverson and Volterra (1996) assessed the progressive interplay of speech and gesture from single to multi-element combinations between 1;04 and 1;08, during triadic interactions. They provide detailed analysis regarding how pointing gestures and verbal production coordinate information, but Capirci et al. do not provide any characterization of eye-gaze, other than suggesting it as a precondition for effective communicative intent (Tomasello, 1995, 2003; Diessel, 2013).

'Usage-based' approaches to language have offered a dynamic theory of how children development communication in which they are assumed to combine different modalities to progress from multi-elements to adult-like complex utterances (Dąbrowska, 2015; Ellis, 2019; Lieven, 2010; Tomasello & Bates, 2001). Before getting to adult-like complex utterances, initially children's multi-element productions go through 'pivot schemas' and 'item-based constructions'.[1] These constructions, extended from a few words to a whole class of lexical expressions, are commonly used to refer to things for which children do not yet have a label (Diessel, 2013; Tomasello, 2003). Both pivot schemas and item-based constructions can convey and combine information. For instance, the deictic information enclosed in referential and attributive constructions can take place through individual modalities (Figure 1a-c), or through the bi- or

[1] Pivot schemas are characterized by a partially fixed and partially schematic slots (e.g., '*More juice'* → '*More X'*). Item-based constructions are characterized by a more abstract use of nouns in referential (identificational) or attributive forms (e.g., 'It is a/the X', 'This is a/the X', or 'It is X' - 'This is X', respectively; see Tomasello, 2003, for detailed description).

multimodal interaction of speech, pointing gesture and eye-gaze (Figure 2). Bimodal combinations can involve speech and pointing gesture, pointing gesture and eye-gaze or speech and eye-gaze (Figure 2a, 2b, 2c), whereas multimodality refers to the use of more than two modalities to convey information (Figure 2d) (Carpenter et al. 1998; Kita, 2003; Yoshida & Smith, 2008; Yu & Smith, 2013; 2017).

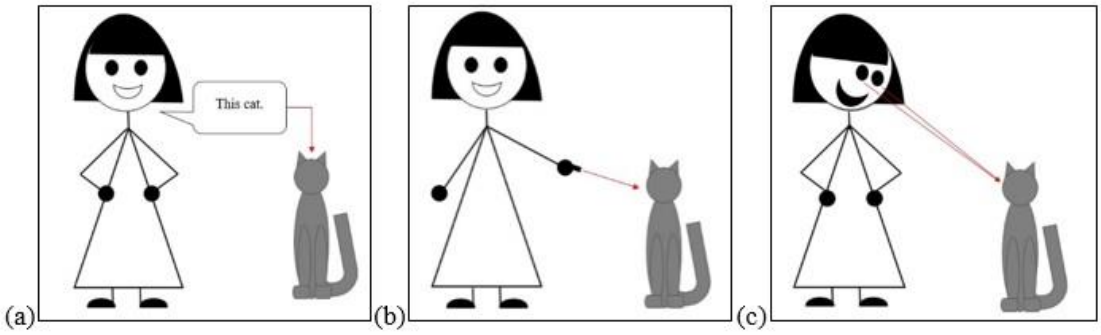


Figure 1(a-c). Examples of unimodal deictic production. In (a), the speaker produces a deictic verbal production. In (b), the speaker produces a nonverbal deictic pointing gesture. In (c), the speaker produces a nonverbal deictic eye-gaze. The cat is the referent across modalities. [COLOUR IMAGE ONLINE ONLY]

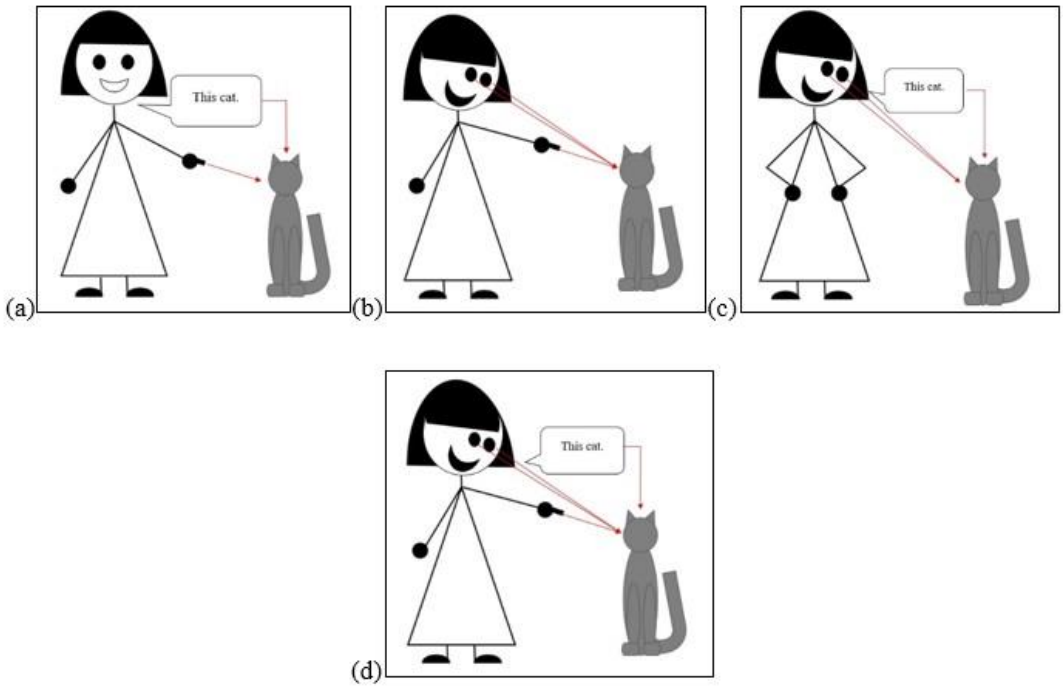


Figure 2(a-d). Examples of bimodal/multimodal deictic production. In (a), the speaker produces both a deictic verbal production and a pointing gesture. In (b), the speaker produces both a nonverbal deictic pointing gesture and eye-gaze. In (c), the speaker produces a nonverbal deictic eye-gaze and a verbal production. In (d), the speaker conveys deictic information though the three modalities: verbal, pointing gesture and eye-gaze. The cat is the referent across modalities. [COLOUR IMAGE ONLINE ONLY]

Studies conducted by Yoshida and Smith (2008), and Yu and Smith (2013, 2017) have investigated the integration of nonverbal modalities in caregiver-child early interaction using both head mounted cameras (Yoshida & Smith, 2008) and eye-trackers (Yu & Smith, 2013; 2017). These methods bring strong evidence of a hand-following pattern, starting from age 0;11 to 2;00 and characterized by following the hand to reach joint attention (Brooks & Meltzoff, 2008; Triesch, Teuscher, Deák, & Carlson, 2006). Specifically, since the coordination of eye-gaze and hand-following provides redundant information, the possibility of relying on a single pattern led by hand-following might reduce the input load during joint attention in complex tasks, such as triadic interactions. Moreover, following the hand to see what the focus of joint attention is (Yoshida & Smith, 2008) might explain how children coordinate their gaze and hand action to experience smooth social interaction, joint attentional episodes and the shift from dyadic to triadic interactions (Yu & Smith, 2017). However, this approach leaves out the analysis of the linguistic input, a pivotal component on which interaction solidly relies upon in early phases of typical language acquisition.

Although usage-based approaches present a domain-general approach that underlines the parallel development and use of verbal and nonverbal modalities in the early phases of language acquisition, they present an incomplete picture of how those modalities are coordinated during interaction *after* they are aligned. To that end, we studied the

relationship between speech, pointing gesture and eye-gaze in deictic communication during joint attentional behaviours. We analysed the deictic communication of caregiver-child dyads, exploring *how* deictic communication takes place (i.e. unimodal, bimodal, multimodal, and synchronization/asynchronization of modalities where applicable), and the role eye-gaze plays immediately prior to and during deictic communication events. With triadic communication, it is often assumed that both participants in the interaction are looking in the same direction in order for a communicator to direct the attention of the conspecific to a particular part of the spatial array and therefore joint attention may be regarded as a precondition of deictic communication (Diessel, 2006; Kita, 2003). Yet there are some languages (e.g. Turkish) which have unique demonstratives (purportedly) used to grab the attention of an interlocutor that person is looking elsewhere (Küntay & Özyürek, 2006). Although most languages (including Italian, our focus below) do not have a specific demonstrative with that function, speakers of languages without a specific explicit demonstrative distinction may nevertheless be affected by that parameter when using the language available to them (Coventry, Griffiths & Hamilton, 2014). To test the idea that joint attention is a precondition for deictic communication, our analyses aimed to identify the extent of three differentiable joint attentional behaviours prior to a deictic episode: shared object attention, coordinated attention and disjoint attention. In shared object attention (Figure 3a) both interactive partners look at the same external referent prior to the production of a deictic event. In coordinated attention (Figure 3b) the two partners look at each other before the production of a deictic event (exhibiting social referential behaviour). Finally, disjoint attention (Figure 3c) is the case where the two partners (or one of them) look away from the shared referent/each other before a deictic production breaking attentional engagement.

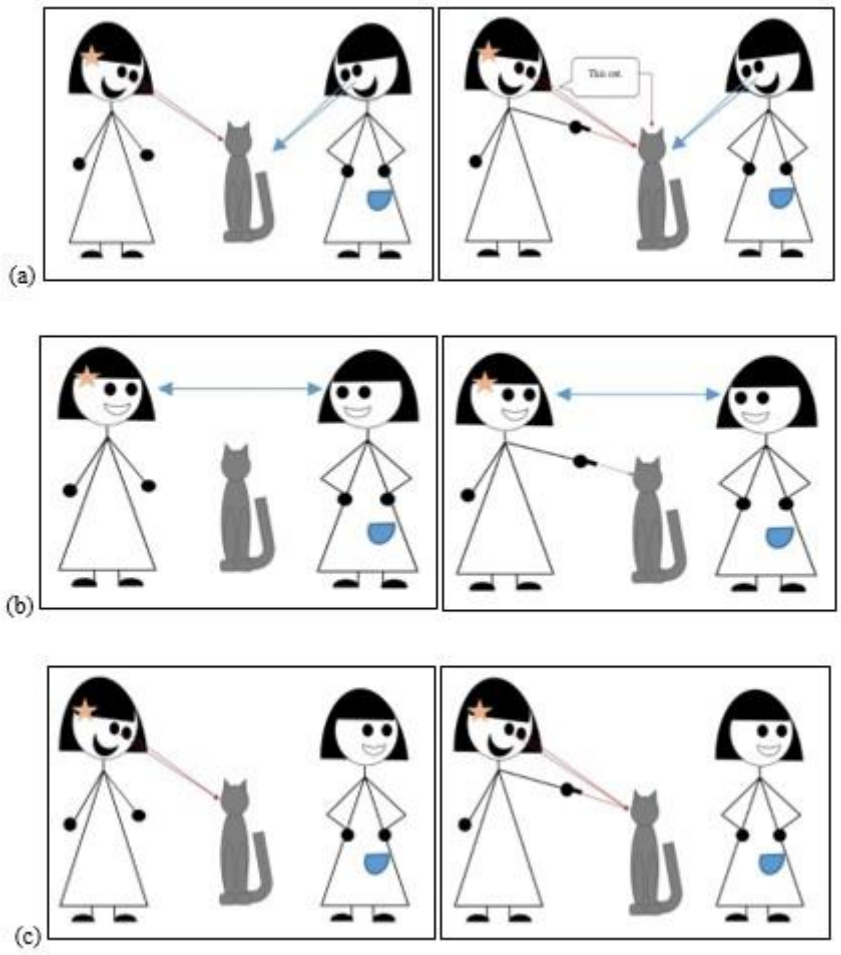


Figure 3. Examples of joint attentional behaviours. In (a), both interactive partners look at the same external referent prior to the production of a deictic event. In (b), both partners look at each other before the production of a deictic event. In (c), one of the partners looks away from the shared referent. All images on the left report the joint attentional behaviour prior to deictic event, while all the images on the right report joint attentional behaviour during the deictic event. [COLOUR IMAGE ONLINE ONLY]

In addition to establishing attentional patterns immediately preceding and during a deictic event, we analysed the organization of speech, pointing gestures and eye-gaze *during* deictic events to assess whether deictic information is conveyed uni-modally, bi-modally or multi-modally. For bi-modal/multi-model deictic events, we further assessed whether the verbal and nonverbal components of the deictic event are synchronized or

independent of one another during production. By doing so, we additionally considered Yu & Smith's (2013; 2017) hypothesis that the hand-following pathway is preferred to eye-gaze following in triadic interactions.

In summary, the study reported below aimed to address the following issues: whether joint attention is a precondition of deictic communication during triadic caregiver-child interaction and *how* deictic events are conveyed across modalities (whether they are unimodal, bimodal or multimodal, and whether modalities are synchronized).

Method

The present study builds upon referential communicative paradigms previously used to investigate verbal and nonverbal strategies for coordinating attention in interaction (Bakeman & Adamson, 1984), and to investigate eye-gaze in shared reading (Guo & Feng, 2013). Data were previously collected for a Master's thesis on language delay at the University of Pisa in collaboration with the IRCCS Stella Maris in Calambrone (Italy). All research procedures were conducted according to institutional guidelines for the protection of human participants, as set out in Legislative Decree No196/2003.

Participants and Materials

Children (5 males and 3 females, mean age 2;02.12, SD 4.7; see Table 1) and caregivers (mothers) participated in a picture-story book-reading task. Children had no history of hearing, vision, or cognitive impairment and presented with a typical language developmental profile (according to their last paediatric review). To assess the complexity of the linguistic structures used by the sample for chronological age, we measured their Mean Length of Utterance on intelligible words (MLU-w; Rice,

Redmond, & Hoffman, 2006; Table 1) and their morpho-syntactic constructions (*GALS; Griglia d'analisi del linguaggio spontaneo*: Spontaneous language analysis form; Cipriani, Chilosi, Bottari, & Pfanner, 1993; Table 1) during spontaneous interaction with their caregivers. Children were all speakers of standard Italian living in a dilalia context (Berruto, 2005); all had passive exposure to Procidano, (an insular variety of Neapolitan dialect), since their parents were speakers of standard Italian/Procidano {AE#13a}. The number of years of education of the caregivers in our sample ranged from 13 to 20 years (median = 13 years), placing participants in a middle SES bracket (following the criteria in Pettinati, Gherardi, Bertelli, & Bilancia, 2007). None of the children attended nursery school at the time of the study.

Table 1.

Demographic Information for the sample

Children sample	Gender	Age	MLU-w	GALS level
1	F	1;08	1.25	2
2	F	1;08	1.33	1
3	M	1;10	1.7	3
4	M	2;03	3.31	4
5	M	2;05	2.3	0.5
6	M	2;06	1.46	2.5
7	M	2;06	2.09	3
8	F	2;07	2.06	3.5
Mean	-	2;02.12	1.93	2.43
SD	-	4.70	0.67	1.20

Note. The table reports information about the sample together with the mean and SD for Age, MLU-w (the mean of the total intelligible word produced divided by the number of statements) and GALS (Griglia d'analisi del linguaggio spontaneo, Spontaneous language analysis form: calculated on the 0-6 level of morpho-syntactic complexity).

An age-appropriate story book ('La spesa di Peppa Pig', Giunti Kids© 2003) was provided by the researcher and used during the experiment to elicit deictic communicative acts and to assess pre-deictic joint attentional states.

We analysed the production of pointing gestures and verbal demonstratives in storytelling. The Italian demonstrative system employs two terms; *quest-* (this) and *quell-* (that) (Bonfiglioli, Finocchiaro, Gesierich, Rositani, & Vescovi, 2009; Jungbluth & Da Milano, 2015). *Quest-* (this) maps onto the proximal domain, identifying a position near the speaker, while *quell-* (that) maps onto a distal domain, identifying a position far from the speaker. In both cases the speaker is usually considered the *origo* (deictic centre). The same binary and symmetric division can be applied to the spatial adverbs *qui/qua* (here) and *lì/là* (there), as displayed in Table 2 (adapted from Jungbluth & Da Milano, 2015).

Table 2.

Demonstratives in Standard Italian

Italian demonstratives' system	Singular form		Plural form		Adverbial form
	M	F	M	F	
Proximal pronoun/adjective	*questo* 'this'	*questa* 'this'	*questi* 'these'	*queste* 'these'	-
Distal pronoun/adjective	*quello* 'that'	*quella* 'that'	*quelli* 'those'	*quelle* 'those'	-
Proximal adverb	-	-	-	-	*qui/qua* 'here'
Distal adverb	-	-	-	-	*lì/là* 'there'

Note. The table sketches the demonstrative system in standard Italian and specifically refers to the types of demonstratives analysed in the present study.

Procedure

Caregivers were informed that their children were participating in a study on nonverbal communication development. They agreed to be videotaped at home (using a Samsung Smart Camcorder) for approximately 30 minutes during their normal play that

included the book reading session. Caregivers were asked to read the story as they generally would, but to hold the book so that the child's hands were free to produce potential gestures[2]. Moreover, caregivers were encouraged to include a series of questions about objects appearing in the book before turning each page. The additional questions could deal with the location of the objects, such as *'Where is the X?'* or with the physical appearance of the object, '*What's the colour of the X?*', in line with the plot of the story. Caregivers were not informed that we were interested in particular types of gestures, expressions or words, such as pointing gestures, demonstrative pronouns, determiners or spatial adverbial forms, so that they would not to be primed, and to allow interaction to flow as naturally as possible.

Children were free to choose the reading position they generally assume during this routine, while allowing the researcher to clearly record the location of looking of both participants during the whole session. The frontal position of the recording afforded the detection of the head movements of the participants, alone or accompanied by eye-gaze movements, as well as brief glances on a vertical shift (cf. Yoshida & Smith, 2008).

Data coding

The whole dataset was annotated in ELAN 5.0 (https://tla.mpi.nl/tools/tla-tools/elan/), an open source tool for the manual creation of annotations to audio and/or video files developed by the Max Planck Institute of Psycholinguistics in Nijmegen. We first identified all instances of deictic communication in the dataset.

Deictic occurrences were categorized into unimodal or bimodal productions across three broad categories: Verbal alone (henceforth, V), Proximal Pointing gesture alone

[2] As we expected, caregivers naturally held the book with their non-preferred hand, on their legs or on another surface in order to produce potential gestures. For this reason, it was not necessary to provide them with additional instructions.

(henceforth, P) - and combinations of the two. When a combination of a verbal expression and pointing gesture occurred, they were further categorized according to how verbal expression and pointing gesture were synchronized as follows: pointing gesture followed by deictic verbal expression P→V, deictic verbal expression followed by pointing gesture V→P and concurrent deictic verbal expression and pointing gesture P+V. The 'V' label refers to each deictic expression without pointing gestures: demonstrative adjectives and pronouns ('this'/'that') + noun, locative adverbs ('here'/'there') and predicative nouns (Clark & Sengul, 1978; 'referential words' in Capirci et al., 1996). A predicative noun refers to a noun or a noun phrase used in the predicate position with a copula verb, having the same referent of the subject of the copula verb and appearing in the frame of the same deictic event (e.g. This is a banana = 'is' is the copula verb and 'a banana' is the predicative noun phrase referring to 'this'). The timing of co-occurrence was considered synchronous (i.e. P+V) if the intelligible vocalization appeared at the peak of the pointing gesture (McNeill, 1992; Kendon, 1980). In the P→V condition a pointing gesture preceded a verbal production, with the verbal production occurring after the point when the hand begins to retract from the peak of the pointing gesture (Table 3 for detailed examples). We additionally coded for false starts and repetitions for the same deictic event. In the case of a false start, either the child or the caregiver started to convey a piece of information, but the production was suppressed before the message was actually conveyed and then started again in the same deictic event (i.e. 'th--- [V]', this banana [P+V]). When restarting, child or caregiver could either use the same modality to convey that piece of information or change it. In the case of repetition, either the child or the caregiver conveyed an entire piece of information, and then repeated it, using the same or different modalities (i.e. 'this banana' [V]/[P+V]' 'this banana'[V]') in the same deictic event. We counted a

false start and repetition as one occurrence and excluded them from the analysis, unless they presented the following characteristics: a) a change in the modality conveying the information; b) they were not adjacent (i.e., they were followed by additional pieces of discourse); c) use of synonymic verbal forms (i.e. 'this banana' and 'this fruit'). Moreover, the occurrences of pointing presenting the finger tapping or producing a circular movement around the referent were considered a P→V combination or a P alone.

Table 3.

Deictic Productions Coding Scheme

Grammatical function	**Examples**
Proximal pronoun [V] *quest-* (this):	*Che cos'è quest-o?* What is this-M:SG? 'What is this?'
Proximal determiner [V] (adjectival form) *quest-* (this):	*Peppa Pig sta prend-endo quest-a carot-a.* Peppa Pig is tak-ing 3:PRS:PROG this- F:SG carrot-F:SG. 'Peppa Pig is taking this carrot.'
Proximal spatial adverbial [V] *qua/qui* (here):	*Guarda qua!* 'Look here!'
Predicative noun [V] (answer to a proximal pronoun):	A: *Cos'é quest-a?* What is this- F:SG. 'What is this?' B: *É un-a banan-a!* (It) is –a F:SG banan-a F:SG 'It is a banana!'
Deictic pointing gesture [P] (stroke of the pointing):	

Note. In the coding adopted, V stands for verbal production and P for deictic pointing gesture. The combined forms deriving from the synchronous or subsequent production of the V and P modalities, give the following realizations: P+V, when the

onset of the stroke of the pointing gesture coincide with the verbal production; V→P, when the verbal production comes before the stroke of the pointing gesture and P→V, when the stroke of the pointing gesture comes before the verbal production. [COLOUR IMAGE ONLINE ONLY]

Additionally, we coded each verbal and nonverbal deictic event according to timing, that is whether the person exhibiting the deictic act was *Initiating* an interaction or *Responding* to an interaction initiated by the other person. In the case of *Initiating* behaviour, one of the partners typically started the interaction using a pointing gesture to a specific element in the book, and by asking a question (i.e., *'Where is the X?*'), by using an imperative to establish or restore shared object state after a disjoint event (*'Look'/ 'Look here!'*), or by producing a deictic noun to which the conversational partner would engage. All the initiating verbal forms could be optionally accompanied by a pointing gesture. In the *Responding* behaviour, one of the partners answered a question using a pointing gesture, verbal production or a combination of pointing gesture and verbal production.

To assess the attentional state just prior to and during a deictic event, we measured three levels of joint attentional behaviours: shared joint attention, coordinated attention and disjoint attention. In shared joint attention (Figure 4), both the caregiver and the child were looking at the book before deictic production. In coordinated attention (Figure 5), the caregiver and the child were looking at each other before deictic production. In disjoint attention (Figure 6a-c), the caregiver or the child showed disengagement from the book reading task before deictic production.

Figure 4. An example of Shared object behaviour. Here both the partners look at the book immediately preceding a deictic production (Moore & Dunham, 1995). [COLOUR IMAGE ONLINE ONLY]

Figure 5. An example of Coordinated pre-deictic behaviour. Here the two partners are looking at each other just before the deictic occurrence. [COLOUR IMAGE ONLINE ONLY]

(a)  (b)  (c) 

Figure 6(a-c). Examples of Disjoint pre-deictic behaviour. In (a), the caregiver and the child are in disjoint behaviour; in (b), the child is showing a disengaged behaviour and in (c), the disjoint behaviour is performed by the mother. [COLOUR IMAGE ONLINE ONLY]

To check the accuracy of the classifications used, two native speakers of Italian, unaware of the purpose of the study, were given 10% of the deictic episodes (randomly selected) to categorize. The level of agreement of both the additional raters ranged from 87% to 90% (Cohen's kappa = .7).

Results

We first considered deictic episodes, and then what happens to eye-gaze contact immediately preceding and during those episodes.

Analyses of Deictic Episodes

We collated the frequencies of proximal and distal deictic events for children and caregivers separately, as a function of timing; that is whether the events were initiations or responses (see above). Due to the low frequencies of occurrence of distal deictic forms (1.38% of all deictic events for children, range 0 – 3, SD = .74; 1.68% of all deictic events for caregivers, range: 0 – 5, SD = 1.0), we focused on proximal productions. We analysed the first 6 minutes of the experimental sessions of each dyad (a total of 2870 seconds of sessions) to afford comparability across the sample and to ensure that children were fully engaged in the sessions being analysed.

Both children and caregivers produced a range of types of proximal deictic communications, with a total of 514 deictic episodes (Table 4). We first considered unimodal (verbal or pointing gesture) and bimodal (verbal and pointing gesture) productions. Children's productions account for 85 initiating events and 132 responding events; while caregivers productions accounted for 268 initiating events and 29 responding events. Notably, the children were fairly equal in the numbers of deictic events produced, with more variability across the caregivers in the numbers of deictic events produced (see rightmost column in Table 4).

Table 4.

Frequency of Occurrence of the Dyads Proximal Deictic Productions

	V - P (demonstratives and predicative nouns in combination with pointing gesture)						V (demonstratives and predicative nouns)		P		Total per subject's production
	P→V		V→P		P+V		V		P		
Children	I	R	I	R	I	R	I	R	I	R	total
1	2	1	0	0	7	0	0	16	5	2	33
	(.92)	(.46)	0	0	(3.22)	0	0	(7.37)	(2.30)	(.92)	(15.19)
2	5	2	1	0	6	2	2	2	1	4	25
	(2.30)	(.92)	(.46)	0	(2.76)	(.92)	(.92)	(.92)	(.46)	(1.84)	(11.52)
3	2	5	0	1	9	1	1	12	1	1	33
	(.92)	(2.30)	0	(.46)	(4.15)	(.46)	(.46)	(5.53)	(.46)	(.46)	(15.20)
4	2	3	0	0	21	2	1	3	0	2	34
	(.92)	(1.38)	0	0	(9.67)	(.92)	(.46)	(1.38)	0	(.92)	(15.66)
5	0	2	0	0	5	0	3	12	0	2	24
	0	(.92)	0	0	(2.30)	0	(1.38)	(5.52)	0	(.92)	(11.06)
6	0	2	0	0	0	2	1	10	0	13	28
	0	(.92)	0	0	0	(.92)	(.46)	(4.60)	0	(5.99)	(12.90)
7	1	1	2	1	4	0	0	6	3	0	18
	(.46)	(.46)	(.92)	(.46)	(1.84)	0	0	(2.76)	(1.38)	0	(8.29)
8	0	0	0	1	0	1	0	13	0	7	22
	0	0	0	(.46)	0	(.46)	0	(5.99)	0	(3.22)	(10.13)
Caregivers	I	R	I	R	I	R	I	R	I	R	Total
1	9	1	7	2	13	6	1	0	3	0	42
	(3.03)	(.33)	(2.35)	(.67)	(4.38)	(2.02)	(.34)	0	(1.01)	0	(14.14)
2	3	0	0	0	9	0	5	4	0	0	21
	(1.01)	0	0	0	(3.03)	0	(1.68)	(1.34)	0	0	(7.07)
3	6	0	3	0	19	0	11	3	1	0	43
	(2.02)	0	(1.01)	0	(6.4)	0	(3.7)	(1.01)	(.33)	0	(14.47)
4	2	0	0	0	1	0	5	2	0	0	10
	(.67)	0	0	0	(.33)	0	(1.68)	(.67)	0	0	(3.37)
5	36	5	4	0	20	1	11	1	0	1	79
	(12.12)	(1.68)	(1.34)	0	(6.73)	(.33)	(3.7)	(.33)	0	(.33)	(26.6)
6	11	0	0	0	12	0	7	2	4	0	36
	(3.7)	0	0	0	(4.04)	0	(2.36)	(.67)	(1.35)	0	(12.12)
7	15	0	1	0	17	0	2	1	3	0	39
	(5.05)	0	(.33)	0	(5.72)	0	(.67)	(.33)	(1.01)	0	(13.13)
8	9	0	1	0	15	0	1	0	1	0	27
	(3.03)	0	(.33)	0	(5.05)	0	(.33)	0	(.33)	0	(9.09)

Note. Frequency of occurrence of proximal deictic production by type for children (top half of the Table) and caregivers (bottom half). The numbers in brackets represent the percentage of each cell of the total number of deictic events for children and caregivers. 'I' refers to initiations, and 'R' to responses.

The data were analysed using a multilevel multinomial generalized mixed model in SPSS (v.25) using a logit link function to examine whether deictic communication type was associated with the participant's role (caregiver or child) and/or the timing (whether the deictic events were initiations or responses). This method was selected due to the two-level structure of the data (multiple responses within each dyad plus multiple responses within each participant) thus accounting for the variance appropriately. The data within the deictic communication type revealed multiple cells with small numbers which rendered multivariate analysis less appropriate (see Table 5) and so the decision was taken to collapse the categories that prioritised pointing gesture (P, pointing gesture alone, and P→V, pointing gesture before verbal production) and verbal communication (V, verbal alone, and V→P, verbal before pointing gesture (see Figure 7).

Table 5.

Frequency of Different Types of Proximal Deictic Productions.

		P	V	P→V	V→P	P+V
Initiating	Child	10 (4.6)	8 (3.68)	12 (5.52)	3 (1.38)	52 (23.96)
	Caregivers	12 (4.04)	43 (14.47)	91 (30.63)	16 (5.38)	106 (35.69)
Responding	Child	31 (14.28)	74 (34.1)	16 (7.37)	3 (1.38)	8 (3.68)
	Caregiver	1 (0.33)	13 (4.37)	6 (2.02)	2 (0.67)	7 (2.35)

Note. Frequency of occurrence of proximal deictic production by type and timing (initiating and responding) for children (top half of the Table) and caregivers (bottom half). The numbers in brackets represent the percentage of each cell to the total number of deictic events for children and caregivers.

The model was run with the data structured at the first level within the dyads and at the second level within the participants (SPSS Syntax in Appendix). Main effects were entered for participant's role (caregiver or child) and the timing of the interaction (initiating/responding), as well as for the two-way interaction participant*timing, with caregivers, initiated and P+V as the reference.

The model correctly predicted 55.1% of the deictic productions (50.3% of the P/ P→V, 52.5% of the V/ V→P and 62.4% of the P+V). The interaction between participant and timing was significant, F (2, 506) = 4.247, p = 0.015, as was timing F (2, 506) = 11.905, $p < 0.01$. Participant role was non statistically significant F (2, 506) = 0.561, p = 0.571. The interaction occurs as children were more likely to use P/ P→V and V/ V→P as responders than as initiators, contrasting with caregivers who were more likely to use V/ V→P categories as initiators than as responders (displayed in

Figure 7); children and caregivers were both more likely to use P+V as initiators than as responders (Fixed effects and coefficients are presented in Table 6 A and B).

Table 6.

Fixed Effects and Coefficient for the Multinomial Analysis

(A) Fixed Effects

	F	df1	df2	Sig.
Corrected Model	14.01	6	506	<0.001
Initiating_Responding	11.905	2	506	<0.001
Child_Caregiver	0.561	2	506	0.571
Initiating_Responding*Child_Caregiver	4.247	2	506	0.015

Probability distribution: Multinomial Link function: Generalized logit

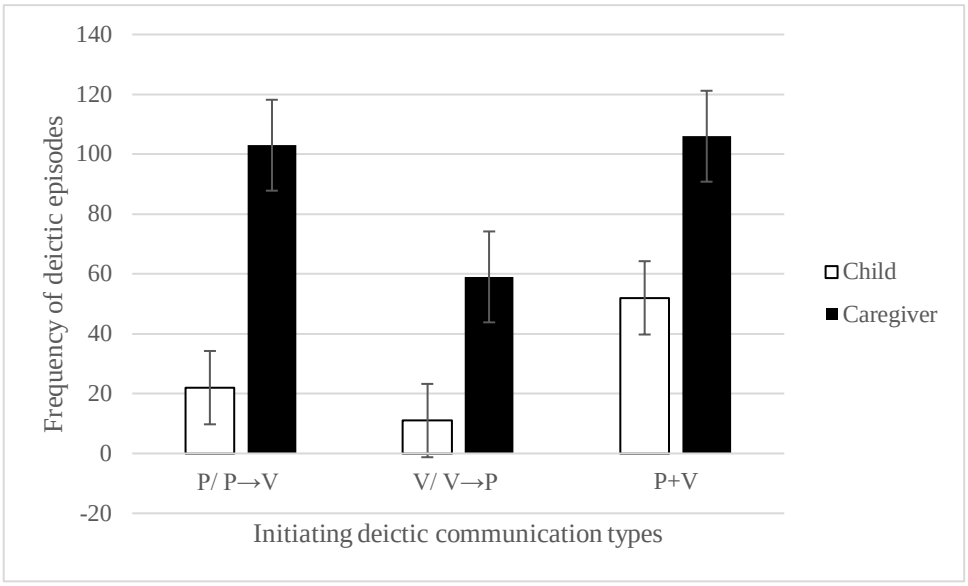
(B) Fixed Coefficient

		Coefficient	95% CI		Std. error	t	Sig.	Exp (Coef.)	95% CI for Exp (Coef.)	
			Lower	Upper					Lower	Upper
Pointing gesture alone or P→V	Intercept	1.747	1.329	2.165	0.2129	8.205	<0.001	5.737	3.776	8.716
	Initiating/Responding = Initiating	-2.543	-3.311	-1.774	0.3913	-6.499	<0.001	0.079	0.036	0.170
	Child/Caregiver = Caregiver	-1.866	-3.781	0.048	0.9743	-1.916	0.056	0.155	0.023	1.049
	Initiating*Caregiver	2.496	0.591	4.401	0.9695	2.575	0.010	12.138	1.807	81.544
Verbal alone or V→P	Intercept	2.279	1.327	3.232	0.4848	4.702	<0.001	9.770	3.770	25.324
	Initiating/Responding = Initiating	-3.797	-5.075	-2.519	0.6503	-5.839	<0.001	0.022	0.006	0.081
	Child/Caregiver = Caregiver	-1.548	-3.436	0.339	0.9606	-1.612	0.108	0.213	0.032	1.403
	Initiating*Caregiver	2.516	0.462	4.570	1.0454	2.407	0.016	12.379	1.588	96.528

Probability distribution: Multinomial. Link function: Generalized logit

Reference categories: Responding/Child

Note. Table 6 (A) reports fixed effects while Table 6 (B) reports fixed coefficient values related to the multinomial modelling analysis. Link function: generalized logit.



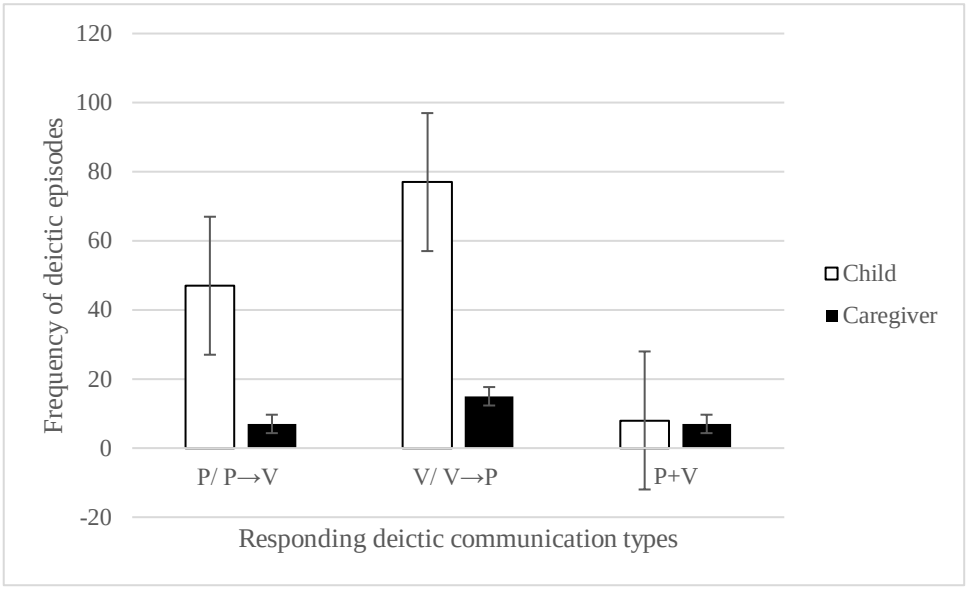


Figure 7. Cropped deictic communicative types (P/P→V; V/V→P; P+V) and timing (initiating/responding) by group (caregiver/child) for initiating (upper panel) and responding (lower panel) deictic communications. The children's data is shown in the white bar; caregivers in the black bar. [COLOUR IMAGE ONLINE ONLY]

We also examined the relationship between types of deictic communication in caregivers and children separately. For children, the number of deictic initiations

produced correlated significantly (Spearman correlations) with the number of multimodal deictic episodes, r_s = .970, p < .001, but not with the number of unimodal initiations, r_s = .278, p = .505. In other words, children producing larger numbers of deictic communications did so as a result of greater use of multimodal deictic productions. For caregivers the same pattern was the case, with a significant correlation between the total number of deictic episodes initiated and multimodal deictic episodes r_s = .929, p < .001, but not unimodal deictic episodes, r_s= .135, p = .750.

Overall the results show that children and caregivers, when initiating deictic episodes, do so in different ways. While both caregivers and children use multimodal synchronous deictic communication frequently, caregivers also use P→V deictic communication frequently while children do not. Intuitively, a pointing gesture could be used by the adult to get the attention of the child, drawing it towards the reference object of focus (Kita, 2003). Thus, a pointing gesture may be used to concretely move the eye-gaze of the child in the direction of the object in the book (cf. hand following pathway; Yu & Smith, 2017). In contrast, children's deictic initiations assume that the caregiver is already attending to the reference object, hence synchrony is a more informationally efficient means with which to communicate.

Next we examined the possible relationship between language level as measured by MLU-w or GALS, and the extent of deictic communication produced by the children. Neither of these language level measures correlated significantly with the number of deictic episodes produced by children, either multimodal or unimodal (p > .05) (but see footnote 3).

Coordination Preceding Deictic Episodes

We categorized eye-gaze immediately preceding deictic events. The number of disjoint (one partner or both partners looking away), coordinated (partners looking at each other) and shared object behaviours are shown in Figure 8 for each dyad. The majority (M = 54.25) of deictic communications were preceded by shared attention on an object, in contrast to relatively few instances of disjoint attention (M = 9.625) or coordinated attention (M = 1.25). A one-way ANOVA revealed that the number of shared attention events preceding deictic communication was significantly greater than the number of other types of attentional events, F (1.1, 7.69) = 44.093, $p < .0001$, partial $\eta^2 = .863$ (both contrasts, $p < .0001$). Analysis of attentional behaviour preceding both initiations and responses showed overwhelmingly that shared object attention was the preferred state in both groups. However, we note that deictic communication was not exclusively preceded by shared attention; for each dyad there were also multiple instances of disjoint attention preceding deictic episodes.

Measurement of the duration of shared object attention prior to deictic communication commencing revealed that the length of shared object attention was remarkably consistent both within and across dyads, with a Mean shared attention time of 2.149 seconds (range = 1.03 - 3.38, SD = 0.771).[3]

[3] We also considered the possible relationship between the length and complexity of the linguistic structures used by the sample and pre-deictic events. There were significant correlations between the average duration of shared object attention pre-deictic episodes and both measures of the length of the linguistic structures used by the sample MLU-w ($r = .728$, $p = .041$) and GALS ($r = .830$, $p = .011$). The only other reliable correlation was between the number of disjoint attention episodes and GALS ($r = .843$, $p = .009$). Given the limited size of the sample, replication with a larger sample will be necessary to confirm these results.

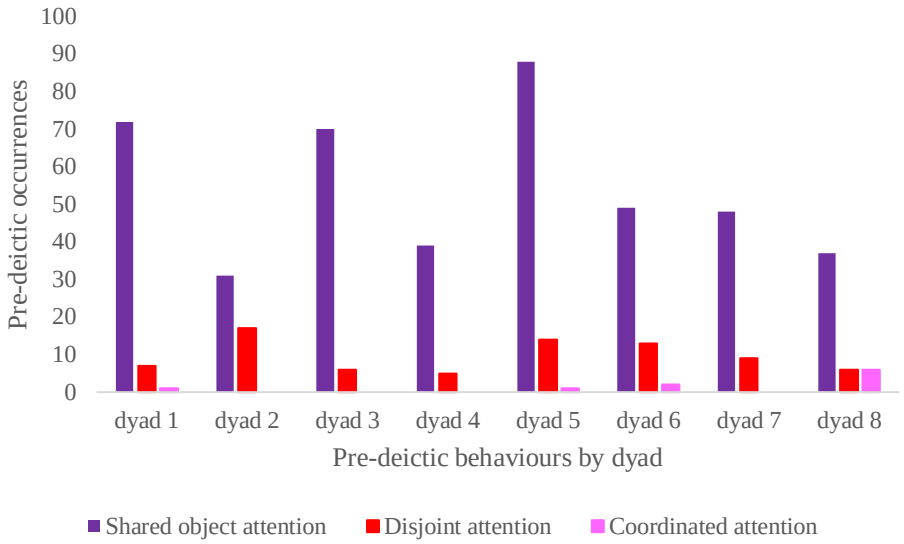


Figure 8. The graph shows the number of pre-deictic behaviours across dyads and reports the frequency of shared object attention (violet/left bar), disjoint attention (red/middle bar) and coordinated attention (pink/right bar) occurring before a deictic event. [COLOUR IMAGE ONLINE ONLY]

Coordination During Deictic Episodes

We next examined eye-gaze synchrony during the production of deictic events. Here, we report percentages of the three joint attentional states analysed (i.e. shared object attention, coordinated attention and disjoint attention) computed on the total number of occurrences per group of participants (children and caregivers) during non-overlapping deictic episodes (Table 7).

Table 7.

Joint attentional behaviour during deictic event

Attention During Deictic Production Across Dyads			
Shared object attention	coordinated attention	disjoint attention	total
374 (89.9)	13 (3.12)	29 (6.97)	416

Note. Table 7 presents the data related to the joint attentional states accompanying the performance of the deictic event. We report the total number of non-overlapping deictic episodes according to the three joint attentional states analysed with reference to the two groups (children and caregivers). We report both raw data and percentages (in brackets), calculated on the total number per group.

As can be seen in Table 7, the predominant joint attentional behaviour during deictic episodes is shared object attention, both in children (91.22%) and caregivers (95.17%). This state is followed by children's disjoint attention (5.61%) and by coordinated attention, both in children (3.17%), and finally by caregivers' disjoint attention state (1.53%). These findings highlight that shared object attention is not only the dominant precondition for deictic communication to be conveyed, but also the preferred state during the production of a deictic event.

Discussion

Deixis is a fundamental part of communication. Spatial demonstratives are the main linguistic vehicle for deictic communication and one of the few classes of non-content words to appear during the one-word stage in language acquisition (Clark, 1978;

Diessel, 2006; 2013). The present study analysed deictic episodes in children with typical language development ranging from age 1;08 to 2;07. Following Carpenter et al. (1998), Capirci et al. (1996) and Yu and Smith (2013, 2017), we assumed that children by this age have already progressed to engagement in triadic interactions. Our analyses examining a corpus of 514 naturally occurring deictic episodes revealed a range of new findings regarding how deictic communication takes place. Acknowledging that multimodality in communication relies on the co-occurrence of different verbal and nonverbal tools and strategies (De Pablo, Murillo, & Romero, 2019; Hall & Knapp, 2013; Matsumoto, Hwang, & Frank, 2015), we focused on verbal production, pointing gestures and eye-gaze synchrony.

To begin with what happens *during* a deictic event, we found an abundance of proximal deictic episodes in the data, with differences in the type of deictic communication form used as a function of initiating versus responding, and participant (child vs. caregiver). As one would expect, children initiated deictic episodes less frequently than adults, but when they did so they overwhelmingly used synchronous bimodal deictic communication – pointing gesture plus demonstrative/predicative word – to direct the attention of the caregiver and to refer to a particular object in the book. Moreover, pointing gestures and verbal production were almost invariably accompanied by a continuous eye-gaze synchrony on the shared object throughout the duration of the deictic episode, which made their proximal deictic productions basically multimodal.

Caregivers also used synchronous multimodal deictic communication when initiating, but in addition, and in contrast to children, they also used pointing gestures to direct the attention of the child prior to speech. This latter pattern, rare in children, may be a means of first directing the attention of the child to the intended reference. Moreover, the tendency to accompany verbal productions with pointing gestures might

reduce the likelihood of any child misinterpretation related to a referent only verbally conveyed, a hypothesis that is supported also by Iverson et al. (1994) in their analyses of early caregiver-child interactions.

In contrast, when responding to a deictic initiation, both children and caregivers were more likely to respond using a verbal production or a pointing gesture alone, while maintaining eye gaze on the shared object. Given that initiations were invariably multimodal, with attention already drawn to a reference and shared attention maintained on that referent, the production of verbal production or pointing alone may serve an anaphoric function in the frame of the deictic episode, referring back to the multimodal deictic initiation. The *forms* of deictic communication produced by children were not linked to the complexity of their linguistic structures.

The analyses of what happens immediately *before* a deictic communication showed overwhelming evidence for the importance of shared attention on an object immediately preceding the initiation of a deictic communication. Children and caregivers were engaged in shared joint attention - looking at the wider context in which the to-be-communicated reference occurs (i.e. the book) - prior to the production of pointing gesture and/or speech to draw attention to a specific object in that wider visual context. This finding reinforces the view that demonstratives 'serve to coordinate [manipulate] the interlocutors' joint attentional focus' (Diessel, 2006: 469; see also Levinson, 2003; Kita, 2003). However, although significantly less frequent, there was also evidence that deictic communication occurred when one of the interlocutors was looking away (disjoint attention state), with synchronous multimodal deictic communication affording a more general attentional function to draw the interlocutor back into the focus of attention for the initiator. This may be related to the specific task of the picture-story book-reading, during which a sharing attention state is a preferred state.

The timing of pre-deictic shared object attention was remarkably consistent across dyads. However, the language measures did not correlate with the numbers of deictic communication episodes (unimodal or bimodal), nor with the frequency of shared object attention episodes.

In order to extend the present findings, there are two points to note. First, there is somewhat mixed evidence that languages differ in the extent to which gestures are utilized. Italian has been regarded as a high gestural language (Kendon, 2004) with a 'high-context' communication style (Cattani et al., 2019; Hall, 1976), and as such one might argue that our results might not generalize to low gestural languages. However, while Cattani et al. (2019) found Italian children outperform English children in the production of representative and pointing gestures (but curiously not Australian children), Liszkowski et al., (2012) found no differences in the production of pointing gestures across seven different languages. Clearly further work is needed to establish specifically whether means of deictic communication differ across languages.

The final point to note is that the present data were collected with only a small number of dyads. Moreover, informing the caregivers that the purpose of the study was the development of nonverbal communication might have boosted the use of gesture, even if they were told to read the story as they usually do. Future work would do well to exclude any effect of instructions and to record deictic communication in a much wider range of communicative settings, and with larger and more diverse samples.

Conclusions

We have provided the first evidence that deictic communication is preceded by shared object attention, and only sometimes by disjoint attention. Children's (1;08 – 2;07) use of deictic communication is overwhelmingly synchronously multimodal. They use speech and pointing gesture to direct attention to a particular part of the visual world together with a synchronous eye-gaze on the object.

References

Bakeman R., & Adamson L. B., (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development, 55(4),* 1278–1289. doi:10.2307/1129997.

Bates, E., Camaiori L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly, 21,* 205-226. Retrived from: https://www.jstor.org/stable/23084619.

Berruto, G. (2005). *Fondamenti di sociolinguistica.* Roma-Bari, Laterza.

Bonfiglioli, C., Finocchiaro, C., Gesierich, B., Rositani, F., & Vescovi, M. (2009). A kinematic approach to the conceptual representation of this and that. *Cognition, 111,* 270-274. doi: 10.1016/j.cognition.2009.01.006.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modelling study. *Journal of Child Language, 35,* 207–220. doi:10.1017/S030500090700829X.

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews, 10,* 169–175. doi: 10.1002/mrdd.20036.

Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache.* Jena: Fischer.

Capirci, O., Iverson, M. J., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language, 23,* 645- 673. doi: 10.1017/S0305000900008989.

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),* 1-143. doi: 10.2307/1166214.

Cattani, A., Floccia, C., Kidd, E., Pettenati, P., Onofrio, D., & Volterra, V. (2019). How do language and culture shape the use of gestures and words by toddlers? OASIS Summary of Cattani et al. *Language Learning.* Retrieved from: https://oasis-database.org.

Cipriani, P., Chilosi, A. M., Bottari, P., & Pfanner, L., (1993). *L'acquisizione della morfosintassi in italiano: fasi e processi*. Padova: Unipress.

Clark, E. V. (1978). From gesture to word: On the natural history of deixis in language acquisition. In Bruner, J.S. & Garton, A. (Eds.), *Human growth and development: Wolfson College lectures 1976*. (pp. 85-120) Oxford: Oxford University Press. Retrieved from: https://www.researchgate.net/publication/306203701_From_gesture_to_word_On_the_natural_history_of_deixis_in_language_acquisition.

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language, 5(3),* 457-475. doi: 10.1017/S0305000900002099.

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive Psychology, 69,* 46-70. doi: 10.1016/j.cogpsych.2013.12.001.

Dąbrowska, E. (2015). What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology, 6,* 852. doi: 10.3389/fpsyg.2015.00852.

De Pablo, I., Murillo, E., & Romero, A. (2019). The effect of infant-directed speech on early multimodal communicative production in Spanish and Basque. *Journal of Child Language,* 1-15. doi: 10.1017/S0305000919000412.

Diessel, H. (2006), Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics, 17(4*), 463–489. doi: 10.1515/COG.2006.015.

Diessel, H. (2013). Where does language come from? Some reflections on the role of deictic gesture and demonstratives in the evolution of language. *Language and Cognition, 5,* 239-249. doi:10.1515/langcog-2013-0017.

ELAN (Version 5.0) [Computer software]. (2018). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/

Ellis, N. C. (2019). Essentials of a Theory of Language Cognition. *The Modern Language Journal, 103, Supplements*, 39-60. doi: 10.1111/modl.12532.

Fricke, E. (2014). Deixis, gesture, and embodiment from a linguistic point of view. In Müller, C., et al. (Eds.), *Body, Language, Communication* (pp. 1803-1823). Berlin/Boston: De Gruyter. doi: 10.1515/9783110302028.1803.

Guo, J. & Feng, G. (2013). How Eye-gaze Feedback Changes Parent-Child Joint Attention in Shared Storybook Reading? An Eye-Tracking Intervention Study. In Nakano, Y., Conati, C. and Bader, T. (Eds), *Eye-gaze in Intelligent User Interfaces,* (pp. 9-21). Springer-Verlag London. doi: 10.1007/978-1-4471-4784-8_2.

Hall, E. T. (1976). *Beyond culture*. New York, NY: Doubleday.

Hall, J. A., & Knapp, M. L. (2013). *Nonverbal communication*. De Gruyter Mouton. doi: 10.1515/9783110238150.

Iverson, J. M., & Capirci, O. (1994). From Communication to Language in Two Modalities. *Cognitive Development, 9,* 23-43. doi: 10.1016/0885-2014(94)90018-3.

Iverson, J. M., Tencer, H. L., Lany, J., & Goldin-Meadow, S. (2000). The relation between gesture and speech in congenitally blind and sighted language-learners. *Journal of Nonverbal Behavior, 24 (2),* 105-130.

Jungbluth, K., & Da Milano, F. (2015). *Manual of Deixis in Romance Languages.* Walter de Gruyter GmbH & Co KG. doi: 10.1515/9783110317732.

Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In Key, M. R. (Ed.) *The Relationship of Verbal and Nonverbal Communication*, (pp. 207-227) The Hague: Mouton & Co.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511807572.

Kita, S. (2003). *Pointing: where language culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers. doi: 10.4324/9781410607744.

Küntay, A. C., & Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal. *Journal of Child language*, *33*, 303-320. doi: 10.1017/s0305000906007380.

Levinson, S. (2003). *Space in Language and Cognition.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511613609.

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua, 120,* 2546–2556. doi: 10.1016/j. lingua.2010.06.005.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A Prelinguistic Gestural Universal of Human Communication. *Cognitive Science, 36,* 698–713. doi: 10.1111/j.1551-6709.2011.01228.x.

Matsumoto, D., Hwang, H. C., & Frank, M. G. (2016). *APA Handbook of Nonverbal Communication.* American Psychological Association, Washington DC. doi: 10.1037/14669-000.

McNeill, D. (1992). *Hand and Mind: What Gesture Reveal about Thought.* University of Chicago Press. doi: 10.2307/1576015.

Moore, C., & Dunham, P. J. (1995). *Joint attention: Its origins and role in development*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development, 78,* 938–954. doi: 10.1111/j.1467-8624.2007.01042.x.

Pettinati, B., Gherardi, P., Bertelli, B., & Bilancia. G. (2007). Indicatori predittivi precoci dello sviluppo del linguaggio: confronto tra metodologie di analisi della componente lessicale. *Psichiatria dell'infanzia e dell'adolescenza, 74,* 707-718.

Rice, M. L., Redmond, S. M., & Hoffman. L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language and Hearing Research, 49(4),* 793-808. doi: 10.1044/1092-4388(2006/056).

Shaw, J. A., Bryant, L. K., Malle, B. F., Povinelli, D. J., & Pruett, J. R. (2017). The relationship between joint attention and theory of mind in neurotypical adults. *Consciousness and Cognition, 51,* 268–278. doi: 10.1016/j.concog.2017.02.012.

Stukenbrock, A. (2015). Intercorporeal Phantasms: Kinesthetic Alignment with Imagined Bodies in Self-Defense Trainings. Retreived from: https:// www.researchgate.net/publication/282319901.

Talmy, L. (2018). *The Targeting System of Language*. Cambridge: MIT press. doi: 10.1515/cog-2018-0061.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103-130). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Tomasello, M. (2003). *Constructing a language. A Usage-based theory of language acquisition*. Cambridge MA: Harvard University Press.

Tomasello, M. & Bates, E. (2001). *Language Development: The Essential Readings.* Malden, MA: Blackwell Publishers.

Triesch, J., Teuscher, C., Deák, G., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science, 9*, 125–147. doi: 10.1111/j.1467-7687.2006.00470.x.

Yoshida, H. & Smith, L. B. (2008). What's in View for Toddlers? Using a Head Camera to Study Visual Experience. *Infancy, 13(3),* 229-248. doi: 10.1080/1525oooO802037.

Yu, C., & Smith, L.B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE 8(11)*: e79659. doi: 10.1371/journal.pone.0079659.

Yu, C., & Smith, L. B. (2017). Hand–Eye Coordination Predicts Joint Attention. *Child Development, 88(6),* 2060–2078. doi: 10.1111/cdev.12730.

APPENDIX

SPSS Syntax

```
GENLINMIXED
/DATA_STRUCTURE SUBJECTS =pairing*id
/FIELDS TARGET = interaction_3 TRIALS =none OFFSET =none
/TARGET_OPTIONS REFERENCE ='3' DISTRIBUTION =MULTINOMIAL LINK
=LOGIT
/fixed EFFECTS = initate_respond child_parent initate_respond*child_parent
USE_INTERCEPT =true
/RANDOM USE_INTERCEPT=TRUE SUBJECTS =Pairing*id
COVARIANCE_TYPE = VARIANCE_COMPONENTS
/BUILD_OPTIONS TARGET_CATEGORY_ORDER =ASCENDING
INPUTS_CATEGORY_ORDER =DESCENDING MAX_ITERATIONS =100
CONFIDENCE_LEVEL =95
DF_METHOD =RESIDUAL COVB =ROBUST
/EMMEANS_OPTIONS SCALE =ORIGINAL PADJUST =lsd.
```