DOI: 10.1111/cdev.14099

EMPIRICAL ARTICLE

CHILD DEVELOPMENT

Language development beyond the here-and-now: Iconicity and displacement in child-directed communication

Yasamin Motamedi[1] | Margherita Murgiano[1] | Beata Grzyb[1] | Yan Gu[1,2] | Viktor Kewenig[1] | Ricarda Brieke[1] | Ed Donnellan[1] | Chloe Marshall[3] | Elizabeth Wonnacott[4,5] | Pamela Perniss[6] | Gabriella Vigliocco[1]

[1]Department of Experimental Psychology, University College London, London, UK
[2]Department of Psychology, University of Essex, Colchester, UK
[3]Institute of Education, University College London, London, UK
[4]Department of Language and Cognition, University College London, London, UK
[5]Department of Education, University of Oxford, Oxford, UK
[6]University of Cologne, Köln, Germany

Correspondence
Gabriella Vigliocco, Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK.
Email: g.vigliocco@ucl.ac.uk

Funding information
H2020 European Research Council, Grant/Award Number: 743035; Economic and Social Research Council, Grant/Award Number: ES/P00024X/1; Royal Society, Grant/Award Number: WRM\R3\170016

Abstract
Most language use is displaced, referring to past, future, or hypothetical events, posing the challenge of how children learn what words refer to when the referent is not physically available. One possibility is that iconic cues that imagistically evoke properties of absent referents support learning when referents are displaced. In an audio-visual corpus of caregiver–child dyads, English-speaking caregivers interacted with their children (N=71, 24–58 months) in contexts in which the objects talked about were either familiar or unfamiliar to the child, and either physically present or displaced. The analysis of the range of vocal, manual, and looking behaviors caregivers produced suggests that caregivers used iconic cues especially in displaced contexts and for unfamiliar objects, using other cues when objects were present.

INTRODUCTION

Language allows us to communicate about what we cannot experience directly with our senses. We talk about people, events, and objects that occur in different places at different times, and about abstract or imaginary things that do not physically exist. This capability, known as displacement, is considered a fundamental feature of human language (Hockett, 1960), allowing us to transcend the immediate physical environment. It is, of course, very common in everyday use, characterizing not only interactions among adults but also the language that caregivers use toward their young children (Grimminger et al., 2020; Tomasello & Kruger, 1992; Veneziano, 2001)—for example, when talking to their child about a family member who is out at work, or a trip to the playground that happened the day before. Being able to transcend the "here-and-now" allows humans to experience and acquire vast amounts of knowledge that would not be available if we were limited to what is physically present. Thus, a key question is how children learn in displaced contexts—how can children know what is being referred to if they cannot see, hear, or otherwise perceive it? Here, we focus on caregivers' multimodal behaviors and ask whether these behaviors differ across two learning

Abbreviation: CDL child-directed language contexts: concurrent learning, when the referent and label physically co-occur, and displaced learning, when the referent is not physically available. While prior research exists suggesting that children can and do learn in displaced contexts (Tomasello et al., 1996; Tomasello & Barton, 1994; Tomasello & Kruger, 1992), what resources they draw on to infer the referent in these contexts are unknown.

Iconicity and displacement

Most accounts of vocabulary learning focus only on unimodal speech contexts, where linguistic information is primarily encoded in the label itself. While the mapping between label and referent is traditionally assumed to be arbitrary (de Saussure, 1916; Hockett, 1960), iconicity has been found to be present and prevalent in the vocabularies of both spoken and signed languages (Perlman et al., 2018; Perniss et al., 2010; Perry et al., 2017; Taub, 2001). Moreover, human interactions, and particularly interactions between caregivers and children, take place primarily in face-to-face contexts, in which speech (and manual sign in the case of sign languages) is embedded in multimodal behaviors that engage the hands, face, and torso to communicate (Holler & Levinson, 2019). Iconicity can manifest widely across these multimodal behaviors (Murgiano et al., 2021; Perniss et al., 2018; Perniss & Vigliocco, 2014).

We suggest that the adaptable nature of human communication allows caregivers to use language and other communicative cues in face-to-face interaction to support learning, and communication more broadly, in different contexts. In particular, we propose that, in displaced contexts when the referent is not physically accessible, children can exploit iconicity to successfully learn the association between label and referent. Here, we define iconicity as a resemblance relationship between any property of the communicative form (e.g., word, sign, gesture) and conceptual (in particular sensory-motor) properties of the corresponding referents (Perniss et al., 2010, 2018; Perry et al., 2017; Winter et al., 2023). Thus, our hypothesis states that in spoken languages, iconicity in both vocal and gestural modalities can support learning by "bridging the gap" between communicative signals and their conceptual counterpart (Perniss & Vigliocco, 2014). However, for this to be possible, multimodal iconic cues must be available to children as part of the communicative behavior by the caregivers with whom they interact.

In the vocal modality, iconicity is mostly visible in onomatopoeia (e.g., such as in English, *meow* or *drip*) that are known to be common in child-directed language (CDL) and in children's early vocabularies (Laing, 2014; Motamedi et al., 2021; Perry et al., 2015, 2017; Tardif et al., 2008). While onomatopoeia, as well as other sound effects, provides a direct imagistic mapping between properties of speech and sounds in the world, more indirect forms of iconic mappings can also be seen in languages' lexica, with large-scale cross-linguistic analyses of lexica finding some degree of indirect iconic form-meaning correspondences in basic vocabulary items (Blasi et al., 2016; Johansson, Anikin, Carling, et al., 2020; Joo, 2020; Wichmann et al., 2010), deictic terms (Johansson & Zlatev, 2013), color terms (Johansson, Anikin, & Aseyev, 2020), and texture words (Winter et al., 2022). This more indirect iconicity may also contribute to vocabulary (and potentially conceptual) development. Operationalizing iconicity in terms of ratings by native speakers (Winter et al., 2017), Perry et al. (2017) found that iconic words are more often used when talking to a child than to another adult and that more iconic words tend to be more common than less iconic words in the speech of young children. Perry et al. (2021) further showed that caregivers' use of more iconic words in utterances in which they introduced a new label/referent was associated with better retention of the label by their (1.5 to 2 year old) children. Beyond lexical iconicity, prosodic contours can be modified in ways that reflect properties of referents, for example lengthening vowel duration to indicate length ("loooong") or raising or lowering pitch to indicate the upward or downward direction of events (Herold et al., 2012; Nygaard et al., 2009). These modifications have been shown to facilitate referent mapping for both children and adults (Herold et al., 2011; Nygaard et al., 2009; Shintel et al., 2014).

Moving to the manual modality, there is evidence that speakers use iconic gestures differently in adult-directed language and CDL, thus supporting the idea that iconic gestures are available to children for learning. Campisi and Ozyürek (2013) studied Italian-speaking adults demonstrating how to make coffee to an imagined interlocutor, either a child or an adult. Their results showed increased rates of iconic gesture, as well as larger and more informative gestures, when the imagined interlocutor was a child, but not for imagined adults, even if the adult were imagined as a novice. Other research has shown that the type of iconic mapping can differ in adult-directed language versus CDL—in a study with signers of Turkish Sign language, Ortega et al. (2017) found that children and adults signing to children (but not adults signing to other adults) preferred action-based signs (e.g., hand representing holding a toothbrush) over perception-based signs (e.g., the hand representing the toothbrush itself). However, the evidence that iconic gestures specifically (rather than other types of gestures such as points) can impact learning is very limited (Acredolo & Goodwyn, 1988; Özçalışkan & Goldin-Meadow, 2005; Rowe et al., 2008), though both iconic labels (including onomatopoeia) and iconic gestures have been shown to be easier to learn by young children (around 2 years) than their arbitrary counterparts (Goodrich & Hudson

Kam, 2009; Imai et al., 2008; Motamedi et al., 2021; Vogt & Kauschke, 2017; Yoshida, 2012).

Crucially, iconic forms are grounded in real-world experience (Emmorey, 2014; Perniss et al., 2010), building on domain-general cognitive processes relating to perceptual, sensorimotor, affective, and social experience that develop from early infancy (Imai & Kita, 2014; Motamedi et al., 2021; Perniss & Vigliocco, 2014; Rączaszek-Leonardi et al., 2018; Saint-Georges et al., 2013). We propose that iconic cues in caregivers' multimodal communication can support word learning, and displaced learning in particular, by imagistically linking properties of the communicative signal and conceptual properties of referents—that is, bringing properties of the referent to the "mind's eye." Namely, the affordances of multimodal, iconic cues that caregivers use in interactions can allow children to draw on prior knowledge gained through general cognitive and motor development to scaffold their vocabulary learning. For example, caregivers' use of an iconic cue can provide some initial information about an unknown concept (e.g., producing a sound mimicking that of a xylophone) that can facilitate a new label-referent mapping. Caregivers' use of an iconic cue (e.g., using the hand to gesture the up-down motion of a seesaw to refer to the seesaw they saw at the playground) can help bring back to mind the relevant object and therefore facilitate linking it to the label. Of course, these cues can be used by caregivers both in displaced contexts and in contexts in which the objects are physically present. In fact, the alternative hypothesis is that caregivers could use these cues even more when objects are present in order to direct the child's attention toward the specific properties of the concept, thus maximizing chances that children will be able to single out the correct referent and learn the corresponding label.

Perniss et al. (2018) offered some evidence in favor of the first hypothesis. In a study investigating signing used by deaf British Sign Language users with their children, Perniss et al. (2018) asked the caregivers to imagine signing to their child about a set of toys, both with the toys in front of them and when the toys had been removed from the room. They found that caregivers modified iconic signs (along dimensions of enlargement, temporal lengthening, and repetition) more often than noniconic signs, increasing their iconic power. Crucially, these modifications were more common for absent referents; that is, caregivers selectively emphasized the iconicity inherent in signs in displaced learning contexts.

Multimodal cues and word learning

Although an increasing number of studies have addressed the role of multimodal caregiver's behaviors in learning, most studies concerning vocabulary development do not distinguish between concurrent and displaced language. While iconic cues can be used both in displaced and concurrent contexts, caregivers can and do use indexical cues in concurrent contexts (Murgiano et al., 2021). Indexical cues, such as directed eye gaze, manipulations of objects and points, provide a direct visual link to the referent. Previous research suggests that these cues may also play a key role in children's development (Goldin-Meadow, 2007) and points have been reported as the most common gestures used by caregivers with children between 18 and 36 months (Clark & Estigarribia, 2011; Rowe et al., 2008). Caregivers also modify their interactions with objects (Brand et al., 2002, 2007) in ways that children prefer over adult-directed actions (Brand & Shallcross, 2008). Gaze-following behavior documented from early infancy highlights that children understand eye gaze as a social and communicative cue early in their development (Brooks & Meltzoff, 2005; D'Entremont, 2000; Farroni et al., 2002; Senju & Csibra, 2008; von Hofsten et al., 2005), with children's gaze-following behavior predicting their later vocabulary size (Brooks & Meltzoff, 2005; Morales et al., 1998). Indexical cues can help to solve referential ambiguity in concurrent learning contexts by visually isolating a referent when it is present (Perniss et al., 2018).

Previous studies have tended to focus on one iconic cue only (iconicity in the lexicon, or in the prosody or in the gestures) rather than evaluate how different iconic cues are distributed within the same communicative exchange. Thus, for example, studies have focused on either gesture *or* eye gaze (Goodrich & Hudson Kam, 2009; Imai et al., 2008; Namy, 2008; Namy et al., 2004; Yu et al., 2019). Where more than one multimodal cue has been investigated, they have often been analyzed together: for example, bringing together pointing, iconic gestures, and object manipulations as gestural communication (Clark & Estigarribia, 2011; Rowe et al., 2008), without distinguishing between iconic and indexical cues, or distinguishing between those cues that can be used both in concurrent and displaced contexts (i.e., iconic cues) and those that can only be used in concurrent contexts (i.e., indexical cues).

Here, we characterize the wide range of multimodal cues caregivers have at their disposal when communicating with their children, at the same time evaluating how different cues are used in concurrent and displaced contexts. We suggest that different cues have different affordances, which lead them to be suited to different learning contexts (Murgiano et al., 2021)—that is, the "usefulness" of a given cue in context will be a result of the specific affordances offered by mapping (iconic or indexical) and modality (vocal, manual). Thus, understanding in which contexts these different cues appear offers insight into displaced language learning. In particular, we hypothesize that multimodal iconic cues will support learning in displaced learning contexts, by bringing properties of the referent to the mind's eye. If the child already knows the referent, the iconic cues will support the retrieval from memory of the concept's representation. If the child does not know the referent, the iconic cues can support the creation of an initial representation of that referent. In contrast, where indexical cues can provide a direct visual link to objects when they are present, we expect iconic cues to be less frequently used in concurrent contexts.

Developmental trajectories in the use of iconic cues

In addition to different cues having different affordances for word learning, we may also expect that the relative importance of different cues will change over the child's development, with the use of iconic cues by caregivers following a developmental pattern consistent with the development of the child's representational abilities. For example, onomatopoeia and indirect lexical iconicity have been shown to be prominent in infant-directed speech compared to speech directed to adults, and then decrease between the ages of 1.5 to 3 years old (Motamedi et al., 2021; Perry et al., 2017), a decrease which parallels the decrease in onomatopoeia production—and an increase in the production of other words—by children (Laing, 2014; Motamedi et al., 2021; Perry et al., 2017). Onomatopoeia have been argued to support early language development because they offer a simple imitative link between a word and the infant's sensory experience such that the sound made by a referent is mapped to the sound of the words (Motamedi et al., 2021). However, other cues may not be usable by children until later on.

For iconic gestures, Özçalişkan and Goldin-Meadow (2005) found that caregivers predominantly use deictic gestures (especially points and conventional gestures, e.g., OK) with their children aged between 18 and 22 months. Iconic gesture production by caregivers increases from 26 months, coinciding with the increased ability by the child to comprehend iconic gestures (Namy et al., 2004; Tolar et al., 2008) and increased production (Ozcaliskan et al., 2013). However, Rowe et al. (2008) did not find substantial change in gesture usage (including iconic gestures and points) for children and their caregivers between 22 and 34 months, with the number of gestures produced remaining relatively stable. The tight link between parental gestures and child gestures is further supported by the finding that the initial iconic gestures produced by children represent routines they engage in with their caregivers (Acredolo & Goodwyn, 1988). The later onset of iconic gestures in the communication of caregivers is not surprising as their comprehension requires multimodal integration of vocal and manual signals and mapping these to a representation of the referent. We do not know whether and how caregivers use iconic gestures with older children.

Prosodic modulations are a hallmark of CDL (Saint-Georges et al., 2013; Soderstrom, 2007). They have been argued to attract children's attention, communicate emotion and attitudes between the caregiver and the child, and facilitate children's speech perception and word comprehension (Cooper & Aslin, 1990; Fernald, 2000; Stern et al., 1983) and have also been shown to support learning of word-meaning associations (e.g., Graf Estes & Hurley, 2013). On this basis, the prediction is that iconic prosody will decrease with age, along with the general trend observed for other prosodic modulations and features characteristic of CDL. However, there is also a plausible alternative hypothesis, leading to the opposite prediction. Because iconic prosody represents relatively abstract dimensions such as size, time, and speed, we could expect that these modulations behave differently than other aspects of CDL (such as hyperarticulation) and their production in caregiver's speech will increase following a pattern similar to that of iconic gestures. This hypothesis is supported by experimental research conducted by Herold et al. (2011), in which 4-year-olds only successfully mapped iconic prosody to novel word meanings when their attention was specifically directed to prosodic information, while 5-year-olds were able to map prosody to meaning without additional instruction.

The present study

Here, we assess the hypothesis that caregivers use multimodal iconic cues predominantly in displaced contexts using a seminaturalistic corpus of 71 English-speaking caregiver–child dyads who were filmed during interactions in their homes. We focus on children aged 2–4, an age range that sees substantial vocabulary development and in which they are able to understand the communicative relevance of different cues (Csibra, 2010; Iverson & Goldin-Meadow, 2005; Namy et al., 2004). Caregivers were asked to talk to their child about sets of objects (e.g., animals, tools) which we provided and in which half of the objects were known to the child and the other half unknown, in order to tap into genuine learning episodes. Moreover, dyads were asked to talk about the objects in a concurrent learning context when these were set on the table in front of them (hence present and manipulable) as well as a displaced learning context when the objects were absent. For each caregiver, we examine the distribution of different iconic cues (onomatopoeia, lexical iconicity, iconic prosody, and iconic gestures) to understand whether caregivers use them more often in displaced contexts and for unknown objects. We further examine the distribution of indexical cues (pointing gestures, object manipulations, and looks to the objects) in situated contexts to establish their frequency in communication about known and unknown objects. Figure 1 illustrates the setup for recording and the design of the study.

Overall, we expect that caregivers will use both iconic and indexical cues more often when the objects are unknown to the child, using diverse multimodal cues to provide a richer representation that can help

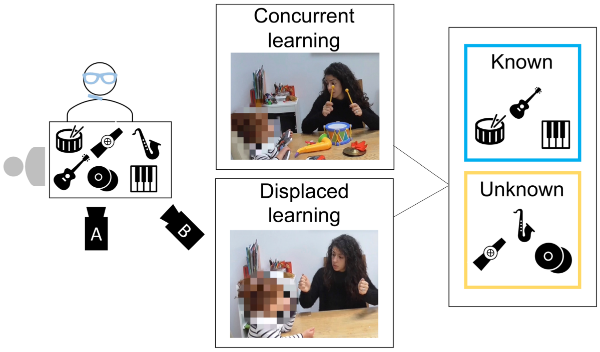

FIGURE 1 Study design. Left-hand panel shows the recording setup. Caregiver (white figure) and child (gray figure) sit at a table at 90° to each other with toys for each category on the table. Two cameras record the interaction: camera A focuses on the caregiver and camera B on the interaction space. Items colored light blue indicate the Tobii eye-tracking glasses and a lapel microphone used with a subset of participating dyads (N=37). Middle panels show the view from camera A when toys are present (concurrent learning context) and when toys are absent (displaced learning context). For each of 4 toy categories, half of the toys are known to the child, half unknown. The right-hand panel shows known (guitar, drums, keyboard) and unknown (saxophone, cymbals, kazoo) toys for the musical instruments category for the participants shown in the middle panel.

the child link label to referent. Crucially, we predict that in displaced learning contexts, caregivers will use iconic cues more often than when the objects are present. We also expect that the relative frequency of different cues will change over development. In line with previous work, we should find that onomatopoeia's use is common in the communication to younger children but then rapidly decreases. A similar decrease is also expected for more general lexical iconicity (Perry et al., 2017). This, however, should not be the case for iconic gestures. As these are only exploited by children later in development, we expect to observe that they are increasingly being used by caregivers (or at least at the same rate) within the developmental window we examine. For iconic prosody, we have two different predictions, as spelled out above, namely, we can find it may decrease in the age window considered as it is the case for other prosodic aspects of CDL, alternatively, we may find it to increase (or at least not to decrease) as children become more capable of mentally representing abstract notions such as duration and speed between ages 2 and 4.

METHODS

Participants

Seventy-one caregiver–child dyads based in the Greater London area (UK) participated in the study. Children were aged between 24 and 52 months (median age = 38 months). The language used between caregiver and child was British or American English. Ethical approval was obtained from the University College London Ethics board. We ran the study with two different cohorts of participants, henceforth cohort A and cohort B; funding for each cohort came from separate sources and at different times (cohort A June 2016–July 2017, cohort B September 2018–February 2020) and each cohort focused on different age groups. Thirty-four caregivers and children took part in cohort A (children: 20 female, adults: 29 female), with children aged between 24 and 42 months (median age = 30 months). Thirty-seven caregivers and children took part in cohort B (children: 18 female, adults: 36 female), with children aged between 36 and 52 months (median age = 42 months). Further demographic information for each cohort can be found on the project's OSF page. All children who participated were typically developing, with no developmental delays reported by caregivers. Receptive vocabulary scores were within the typical range for the age ranges contained in each cohort. Cohort A was assessed using the Oxford Communicative Development Inventory (Hamilton et al., 2000), which gives a measure of the percentage of common words understood by children (N=29, range = 65.55–100, M=93.63, SD = 9.30). Cohort B was assessed with the British Picture Vocabulary Scale, an appropriate measure for the older age range, which gives a measure of receptive vocabulary, standardized based on the child's age (N=37, range = 85–125, M=107.14, SD = 10.64). Data from one child in cohort A were eventually excluded from analysis due to missing demographic (namely, age) data.

Materials

Toys presented in the study were from 4 categories: foods, musical instruments, animals, and tools. These categories are common for children of this age range and were chosen for the availability of toys and for the opportunities these categories offer for vocal and manual iconicity. We used two different sets of toys for each cohort due to the suitability of toys for different age ranges. For each dyad, we used 6 toys from each category such that each set contained approximately 3 toys known to the child, and 3 unknown, based on parental reports of the child's knowledge. Individual toys were used in a roughly equal number of testing sessions across participants. A full list of the toys used for each cohort, and the toys used for each dyad can be found at the project's OSF page.

Procedure

Testing sessions took place in the families' homes. For both cohorts, caregivers were contacted prior to the session with a full list of toy names and asked to indicate whether their child knew each object and each word.

Two experimenters visited the family and recorded interactions with two video cameras, one of which focused on the caregiver, the other of which focused on the interaction space. For cohort B, in addition to the two video cameras, we also recorded caregiver speech, using a unidirectional lavalier (lapel) microphone and caregiver gaze fixations using Tobii Pro Glasses 2. Caregivers in each cohort were given largely the same instructions (The full text shown to each set of caregivers is available on our OSF page):

> You and your child will be recorded while talking about toy objects (sets of animals, tools, musical instruments and foods) as you would normally do. You will be talking about one set at a time and in some cases, we will leave the objects with you while you are talking while in other cases we will take the objects away. We would like to carry out the study in the nursery or in your home as we believe this may be more comfortable and convenient.
>
> Your interaction will be recorded using two cameras: one camera will be on you, and one camera will be on your child. As we are especially interested in how caregivers communicate to their children, we are particularly interested in the video of your communication (rather than your child), however, we need to capture the entire interaction in order to make sure we correctly annotate your communication.

During recording, for cohort A, one experimenter monitored the correct working of the video cameras, while the other brought and removed toy sets to and from the interaction space. For cohort B, one experimenter monitored the correct working of the equipment, but left the room while recording took place. The second experimenter brought and removed toy sets to and from the interaction space. The interactions were carried out at a table, with caregiver and child sitting at 90 degrees from each other, with the exception of 1 participating dyad, who did not have a suitable table (in this case, caregiver and child sat on the floor). Caregivers were asked to interact with their child in a natural way, but to try to talk about each of the objects provided. A drawing depicting the toys (cohort A) or a typed list of the toys (cohort B) was given to the caregivers at the beginning of each category round to help the caregivers remember all the toys in the set during the toy absent manipulation. The order of toy categories and the concurrent-displaced manipulation was counterbalanced across participants.

Where the concurrent learning context came first, the experimenter brought a set of 6 toys to the table from one category (e.g., animals) and left the room. The caregiver and child talked about these objects (3–5 min), before the experimenter entered the room and removed the toys from the interaction space, by asking the child to help them tidy the toys away. The experimenter left the room with the toys for the displaced learning context, asking the caregiver to continue talking about the toys they had just seen (3–5 min). The experimenter then brought in a new set of 6 toys, repeating this procedure for all 4 toy categories. When the displaced learning context came first, the caregiver was asked to begin talking about the toys that would next appear (based on the aforementioned drawings or list). After 3–5 min, the experimenter brought in the set of toys for that category, and left the room, leaving the caregiver and child to interact with the objects present. After 3–5 min, the experimenter entered the room to remove the toys and the procedure continued for all 4 toy categories. The full recording session lasted approximately 35–45 min.

Data processing

For cohort A, recordings of each session were split up into sections by members of the research team, defined by toy category and displacement (e.g. tools absent-tools present), giving 8 sections per dyad. Audio-visual data were used to code caregiver communicative behavior according to the parameters described below. The same parameters were used to code caregiver's behavior for cohort B, except that audio (speech), video (gestures) and eye gaze fixations were coded separately, and then combined. We analyzed data from 70 participants for all measures apart from eye gaze, where data from 33 participants were included (explained below in "Indexical cues" section).

Speech coding

Our unit of analysis is the multimodal utterance. We are interested in understanding how often caregivers produce different cues; for any given cue, we want to assess what proportion of the communicative interaction contains such a cue. Therefore, the multimodal utterance we use as our baseline contains both speech information (based on the utterance coding described in the following paragraph) and information about where manual cues (iconic gestures, points, and object manipulations) occur. Manual cues that occur without speech are coded as a multimodal utterance using a placeholder tag.

Caregiver speech was initially transcribed by utterance, defined as a unit that expresses a single situation (activity, event or state) with an implicit or explicit predicate. For example, "I eat the bread," or the utterance with an implicit predicate, "Red," meaning "It is red." These utterances are then combined with information from the manual cues to create the multimodal utterance.

For each multimodal utterance, we coded the topic, noted as the specific toy (or multiple toys) that each utterance referred to, whether or not an explicit label was produced. For example, the topic would be noted as "saxophone," if the caregiver said "Can you play the saxophone?" or "Can you play it?" when referring to the saxophone, or if they pointed to the saxophone and said "What's that?". The topic coding provided our basis for designating utterances as known or unknown. In cases where there was more than one topic (if the caregiver was communicating about more than one toy, or the topics communicated by speech and manual cues differed), all topics were coded and known-unknown status was assessed on the basis of all topics. For example, if all were known items, the utterance would be coded as known; if topics were mixed, we could not ascertain known-unknown status and these utterances were excluded from analysis. Utterances that did not focus on our toy referents were coded as "other" and excluded from analysis. Utterance level coding for both cohorts was carried out by trained members of the research team. For cohort A, this was carried out in ELAN (Sloetjes & Wittenberg, 2008); for cohort B, all speech coding and processing was carried out using Praat (Boersma & Weenink, 2019).

In addition, we coded separately explicit mentions of the label (i.e., the caregiver saying the word "saxophone"). For cohort A, this was done by expert coders. For cohort B, this was done automatically. Speech utterances were segmented using WebMaus (Kisler et al., 2017) and then manually corrected by expert coders. Toy referents were then extracted from the transcriptions using a Python script.

Iconic cues

Onomatopoeia

We noted cases of onomatopoeia as sounds produced by caregivers which imitate real-world sounds. This accounts for both lexicalized onomatopoeia (e.g., in the phrase "The dog goes *woof"*) as well as sound effects produced by the mouth (e.g., making a barking noise). Onomatopoeia coding was carried out by trained members of the research team.

Lexical iconicity

All words in multimodal utterances were queried against the iconicity norm scores in Winter et al. (2023) receiving a score between 0 and 7. Note that we excluded fillers (e.g., "hmm") and interjections (e.g., "ah," see OSF page for details). For each multimodal utterance the word with the highest iconicity score was selected as the maximally iconic word in that utterance. If the maximally iconic word scored >5, the utterance was considered as highly iconic, and if the maximally iconic word score <5, the utterance was considered not highly iconic.

Iconic prosody

We subjectively coded iconic prosody along the semantic dimensions of size, duration, speed, direction (position), and loudness. Prosodic modulation was either in speaking rate (duration and speed), pitch (duration, size, speed, direction and loudness), vowel duration (duration and speed), or intensity (size, loudness) (e.g., "The Toucan has a loooooong beak").

Iconic gestures

We coded the use of iconic gestures that represent properties of referents. This may be through depicting the shape or size of an object (e.g., hands moving apart to represent the long legs of the ostrich) or how the object is manipulated (e.g., a hammering gesture in which the shape of the hand represents how a hand would hold a hammer).

Indexical cues

Points

We coded points as manual gestures that single out a referent through deixis. This may be a canonical index finger point or, for example, an open palm gesture indicating a specific object. Points are directed at particular objects; as such, they can only occur in the concurrent learning context. Though we acknowledge that points can appear in displaced contexts, we assert that in this case points have a different representational characterization. Points to a present object allow the communicator to create a visual link to an object. Points that occur in the absence of an object (e.g., to a location the object previously occupied, or to an object that acts as a stand-in for another object) cannot establish the same visual link, and so do not have the same representational function.

Object manipulations

Object manipulations are defined as actions or movements performed while holding or manipulating an object. For example, this could be the caregiver holding an object with the intention of making it visually salient to the child, or performing an action on an object, such as tapping the keys on the keyboard to indicate how it is used, or highlighting a feature of the object (such as the anteater's snout). Object manipulations must be meaningful with respect to the object (i.e., we do not count actions carried out while incidentally holding an object). Object manipulations were only coded for the toys we provide; as such, they can only occur in the concurrent learning context.

Eye gaze

Eye gaze data were collected for cohort B only (N=367). Eye gaze raw recordings were first processed using Python (Tobii_glasses2_utils, 2015/2017), in order to mark gaze position in the caregiver point-of-view recording obtained from the eye tracking glasses. Then a member of the research team noted which toy was the video was annotated by an expert coder manually, noting the specific toy that was the focus of each gaze fixation. The gaze fixation was annotated and coded for gazes that lasted for 3 or more consecutive video frames. As gaze was annotated and coded for fixation on specific objects, it can only be done so coded for concurrent learning contexts. We excluded gaze data from 3 participants due to low gaze fixations; this was determined such that the mean proportion of gaze fixations for the excluded participants was less than the overall mean proportion of 1 SD (calculated across all participants).

Annotation protocols

For cohort A, all manual cues were identified according to the above criteria by trained members of the research team, using ELAN. Manual cues were annotated as descriptions of the cue, including which toy the cue made reference to. For cohort B, a subset of participants (N=22) had manual cues coded using a mix of expert and crowdsourced coding. To this end, the expert coders (trained members of the research team) marked the beginning and end of segments containing any of our manual cues (representational gestures, object manipulations, and points). Then, these video segments were extracted and uploaded to an online coding system designed in and hosted by Gorilla (Anwyl-Irvine et al., 2020) and coders were crowdsourced using Prolific (www.prolific.co). Online coders were shown a few introductory video clips of parent–child interactions and were given instructions about how to code them. Each online coder saw a total of approx. 150 video clips, each lasting less than 30 s. For each video clip, the online coder was asked (1) to classify the parent's hand gesture into either representational gesture, object manipulation, point, or other (if it did not fall into any of these categories) and (2) to identify the object that the gesture referred to (to select one of the available objects, or none if they could not identify the object). The procedure took approximately 30 min, and participants were paid £3.75 for participation. Each video clip was seen and coded by at least 7 online coders. The answer of the majority was taken as the selected answer (i.e., more than 60% of online coders had to agree on the gesture or object identity). Finally, all the gesture types or objects that were not agreed on by online coders were subsequently coded manually by expert coders. The rest of the corpus (N=15) was coded by expert coders following the same coding scheme as study A.

Multimodal integration for cohort B

For cohort B, audio, video, and eye-tracking data were coded separately in Praat (speech), and in ELAN (manual gestures and eye-gaze), and then combined in ELAN. To do this, first the multimedia files were aligned—the signal for the beginning of recording was the clapperboard snap, which was easily detectable in video and audio channels—and then the codings were offset based on the start point of the recording and imported to ELAN.

Reliability analysis

For cohort A, one section of the interaction for each participant (counted as the combination of presence–absence and toy category, e.g., toy present—musical instruments) was randomly selected for reliability coding relating to onomatopoeia, iconic gesture, pointing, and object manipulations, representing approximately 12% of the total interaction. For cohort B, 10% of the total duration for the interaction was randomly selected for each participant for reliability coding for manual cues (object manipulation, points and iconic gestures) and a similar selection process was used to reliability code 19 participants for onomatopoeia and 16 participants for eye gaze fixations. We analyzed data from these annotations by computing the Spearman's correlation between main and reliability coders for each variable, relating to (i) the number of annotations created and (ii) the time in milliseconds those annotations covered. All measures showed strong correlations between main and reliability coders (see Table 1).

TABLE 1 Spearman's correlation coefficients for main and reliability coding for onomatopoeia, gesture, points, and object manipulations.

	Cohort A		Cohort B	
Cue	N annotations	Time covered	N annotations	Time covered
Onomatopoeia	r_s=.80	r_s=.69	r_s=.97	r_s=.97
Iconic gesture	r_s=.91	r_s=.91	r_s=.84	r_s=.85
Points	r_s=.87	r_s=.84	r_s=.60	r_s=.58
Object manipulations	r_s=.99	r_s=.97	r_s=.81	r_s=.82
Eye gaze	-	-	r_s=.87	r_s=.64

TABLE 2 Onomatopoeia model results.

Effect	*β*	SE	*z*	*p*
Age	−.04	.02	−2.76	.006
Familiarity	−.11	.08	−1.26	.21
Presence	−.58	.09	−6.70	<.001
Cohort	.41	.23	1.76	.08
Category	.01	.01	1.18	.24
Age:familiarity	−.02	.01	−2.31	.02
Age:presence	.009	.01	0.81	.42
Familiarity:presence	−.31	.13	−2.36	.02
Age:familiarity:presence	.04	.01	2.97	.003

The subjective coding for iconic prosody was confirmed as highly reliable in two ways. First, through an intersubjective reliability coding of 10% of the coded utterances (99.9% agreement between coders). Second, on the objective prosodic properties of a subset of predefined seed words related to the semantic dimensions of interest (e.g., long, big, quickly). Specifically, we found that those seed words coded as having iconic prosody did in fact exhibit significantly more extreme prosodic properties compared to corresponding seed words that were not coded as having iconic prosody (for example, “long” vs. “loooooong”).

Analytical procedure

Coded data were exported from ELAN in .csv format for analysis (processing scripts and procedure can be found at the project's OSF page). Data for label production as well as each iconic and indexical cue of interest were recoded into binary variables that noted the presence or absence of a cue in each row (representing an utterance). We ran mixed effects models using R (R Core Team, 2013) and lme4 (Bates et al., 2015).

To model the presence of iconic cues (onomatopoeia, prosody and representational gesture), we use a logistic mixed effects model with the following model structure:

$$\text{glmer}(\text{cue use} \sim \text{age} \times \text{familiarity} \times \text{displacement} \\ + \text{category} + \text{cohort} \\ + (\text{familiary} \times \text{displacement} | \text{participant}), \\ \text{family} = ' \text{binomial}')$$

Familiarity and displacement were coded as binary variables and deviation coded and the fixed effect of age was centered, using the same fixed effects structure as for the utterance frequency model. We also included a by-participant random intercept, with random slopes for toy presence, familiarity and the two-way interaction term. Note that for all models we do not include by-item random effects. This is because identifying individual items (here the toy being referenced) at the utterance level is difficult; caregivers may talk about several toys at the same time within the same utterance. We do not necessarily expect *items* to differ systematically across our sets of toys; rather, we expect that different categories of toys will have different affordances (e.g., musical instruments as a category will afford more onomatopoeia than, e.g., foods). For this reason, we include the toy category as a deviation-coded control variable in our models. We also include cohort as a control variable in the model (except for gaze models, which only include data from cohort B), to account for differences in how data were collected for each cohort. Full model details can be found on the project's OSF page. For the model analyzing lexical iconicity (excluding onomatopoeia), iconicity is modeled similarly as a binary variable (iconic or not), using the same model structure with additional controls—for the maximally iconic word in each utterance, we obtained concreteness values (Brysbaert et al., 2014), word frequency (Zipf score from the SUBTLEX-UK CBeebies corpus, van Heuven et al., 2014) and word length in phonemes (via SCOPE, Balota et al., 2007; Gao et al., 2022), as these are known to independently predict lexical iconicity.

The models for indexical cues are largely similar, with the exception that we only analyze data from the concurrent learning context and therefore do not include displacement as an effect in the models.

$$\text{glmer}(\text{cue use} \sim \text{age} \times \text{familiarity} + \text{category} + \text{cohort} \\ + (\text{familiary} | \text{participant}), \text{family} = ' \text{binomial}').$$

To deal with convergence issues, we ran the models with different optimizers using the allFit function from the lme4 package, reporting models using optimizers that allow convergence. Where no optimizers allow convergence, we removed the correlations between slopes and intercepts. Data files in .csv format and all analysis scripts can be found at the project's OSF page.

RESULTS

Below, we report separate logistic or linear mixed effects models for each cue of interest to analyze how often caregivers used cues dependent on (i) the learning context (displaced-concurrent), (ii) the child's familiarity with the object (known-unknown), and (iii) the child's age. Learning context was not included in models for indexical cues (or any associated interactions), as they cannot occur when toys are absent (see methods for full description of our coding scheme). We use the multimodal utterance as our unit of analysis, such that, we analyzed each primary cue of interest (label use, onomatopoeia, iconic prosody, iconic gesture, points, object manipulations, and looks to the object) as a proportion of the multimodal utterance. We focus here on the range of iconic and indexical cues we identify in the interactions—full model details and additional analyses relating to the speech content can be found in the methods section and on our OSF page.

Iconic cues

Results pertaining to caregiver onomatopoeia production are shown in Table 2 and illustrated in Figure 2a. Our model revealed a main effect of age, and toy presence, as well as several interactions. The interaction between age and familiarity indicates that caregivers use more onomatopoeia for unknown objects when interacting with younger children. We also find an interaction between familiarity and toy presence such that caregivers produce more onomatopoeia for unknown objects when they are displaced, and a three-way interaction between age, familiarity, and toy presence, which suggests that this effect is stronger for younger children in our sample.

We also analyzed utterances which contained words high in lexical iconicity but which were non-onomatopoeic, the results of which are reported in Table 3 and illustrated in Figure 2b. Our findings suggest that lexical iconicity is more common in utterances when toys are present, while the interaction between age and familiarity suggests that use of highly iconic words is maintained across ages when toys are unknown, but reduces across development for known contexts. This model further shows that caregivers tend to use more concrete language in displaced than concurrent contexts.

Results for iconic prosodic modifications are shown in Table 4 and illustrated in Figure 2c. We find a main effect of familiarity, as well as a main effect of toy presence and an interaction between the child's age and toy familiarity, with more iconic prosodic modifications when toys are absent, and for unknown objects especially for the younger children.

Results concerning iconic gesture production are shown in Table 5 and Figure 2d. We find a main effect of toy presence, as well as a main effect of familiarity. Caregivers produce more iconic gestures in displaced learning contexts, and when talking about unknown objects.

Indexical cues

Analysis of caregivers' points are shown in Table 6 and illustrated in Figure 3a. We found no evidence for a main effect of familiarity, nor for the interaction between age and familiarity. The effect of cohort suggests that caregivers in cohort B used more pointing gestures than those in cohort A.

Findings relating to caregiver direct manipulation of objects are shown in Table 7 and illustrated in Figure 3b. Our analysis revealed a main effect of familiarity, such that caregivers engage in more object manipulations when toys are unknown to the child, and an effect of cohort, suggesting that caregivers in cohort B performed object manipulations less often than those in cohort A.

We analyzed caregivers' looks to objects in our toy sets, for a subset of our data ($N=37$, with three participants excluded from analysis). In Table 8, we present data showing the proportion of looks to objects overall, as well as looks to objects that co-occur with object manipulations (see Figure 4).

When analyzing all gaze fixations, we found no evidence that child's age or toy familiarity affected fixations on targets across utterances with labels and all utterances. We also analyzed gaze fixations that co-occurred with object manipulations. In this case the model revealed a significant effect of familiarity on gaze fixations—caregivers show a higher proportion of fixations for unknown toys, compared to known ones. We did not find a main effect of age, or an interaction between age and familiarity in this case.

DISCUSSION

Despite displacement being a fundamental feature of language in general (Hockett, 1960), and common in CDL in particular (Tomasello & Kruger, 1992; Veneziano, 2001), very little prior work exists that investigates how caregivers communicate in displaced learning contexts. Moreover, previous studies have tended to investigate multimodal communication by either focusing on just one cue in addition to speech (e.g. gesture or object interactions or eye gaze; Brand et al., 2002; Gogate et al., 2000; Iverson et al., 1999; Jo & Ko, 2018; Perry et al., 2017; Rowe et al., 2008; Yu et al., 2019) or by grouping together multiple (manual) cues with different

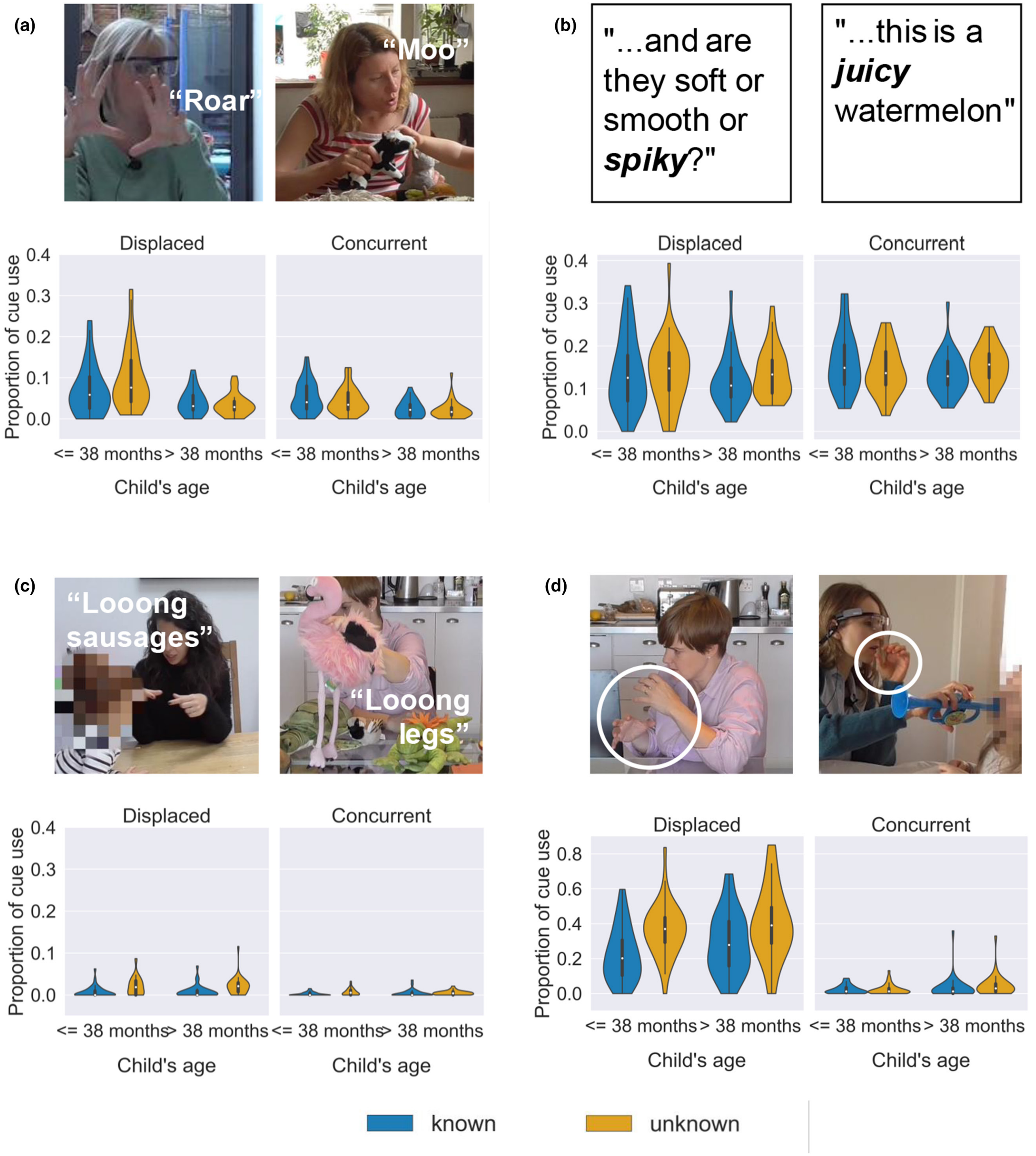

FIGURE 2 Iconic cue results. Each panel (a–d) shows examples of each cue in the displaced context (left) and when the concurrent context (right). Plots show the proportion of utterances that contain onomatopoeia (a), lexical iconicity (b), iconic prosody (c) and iconic gesture (d). Plots show separate panels for displaced and concurrent learning contexts. All plots show proportions across utterances. Age differences are shown on the x-axis using a median split (38 months). Different colors represent known (blue) and unknown (yellow) toys respectively. For lexical iconicity (b), word in bold shows the maximally iconic word in the utterance.

affordances, such as gesture and object manipulations or iconic and pointing gestures (Clark & Estigarribia, 2011; Rowe et al., 2008). Our work brings together multiple cues from different modalities in the same interactional context to understand their distribution across different learning contexts.

We found that displacement and familiarity affected all caregiver behaviors. Caregivers talked more about toys when these were present and more about toys that were unfamiliar to the children. They also produced the label for the object more often when the object was present and when it was known to the child (see OSF

TABLE 3 Lexical iconicity model results.

Effect	β	SE	z	p
Age	−.006	.009	−0.65	.51
Familiarity	.04	.06	0.63	.53
Presence	.20	.06	3.50	<.001
Cohort	−.09	.14	−0.68	.50
Category	.02	.02	1.09	.28
Concreteness	.22	.02	11.20	<.001
Word frequency	−.53	.02	−25.07	<.001
Word length	−.49	.02	−26.72	<.001
Age:familiarity	.02	.008	2.49	.01
Age:presence	.003	.008	0.43	.67
Familiarity:presence	−.08	.07	−1.05	.29
Age:familiarity:presence	.006	.01	0.60	.55

TABLE 4 Prosody model results.

Effect	β	SE	z	p
Age	0.01	.02	0.75	.45
Familiarity	1.23	.23	5.44	<.001
Presence	−1.20	.19	−6.30	<.001
Cohort	−0.11	.26	−0.44	.66
Category	−0.19	.02	−7.90	<.001
Age:familiarity	−0.06	.02	−2.76	.006
Age:presence	−0.02	.02	−0.81	.42
Familiarity:presence	−0.41	.41	−1.00	.32
Age:familiarity:presence	−0.008	.04	−0.20	.84

TABLE 5 Iconic gesture model results.

Effect	β	SE	z	p
Age	.001	.02	0.05	.96
Familiarity	.42	.14	3.13	.002
Presence	−3.33	.16	−20.54	<.001
Cohort	−.35	.32	−1.09	.27
Category	.20	.007	26.65	<.001
Age:familiarity	.02	.02	1.19	.23
Age:presence	.02	.02	1.14	.25
Familiarity:presence	−.30	.25	−1.22	.22
Age:familiarity:presence	.04	.03	1.47	.14

TABLE 6 Pointing model results.

Effect	β	SE	z	p
Age	.005	.02	0.30	.76
Familiarity	.12	.11	1.16	.25
Cohort	−.60	.26	−2.26	.02
Category	.10	.009	11.04	<.001
Age:familiarity	.009	.01	0.66	.51

project page). While it is unclear why this may be the case, this finding suggests that caregivers did not interpret the study setting as a "teaching experiment" in which their task is to teach their children the novel words; rather, they felt free to interact in a more naturalistic manner.

We hypothesized that iconic cues that provide imagistic representations would be more frequently used in displaced contexts (Perniss et al., 2018), where they

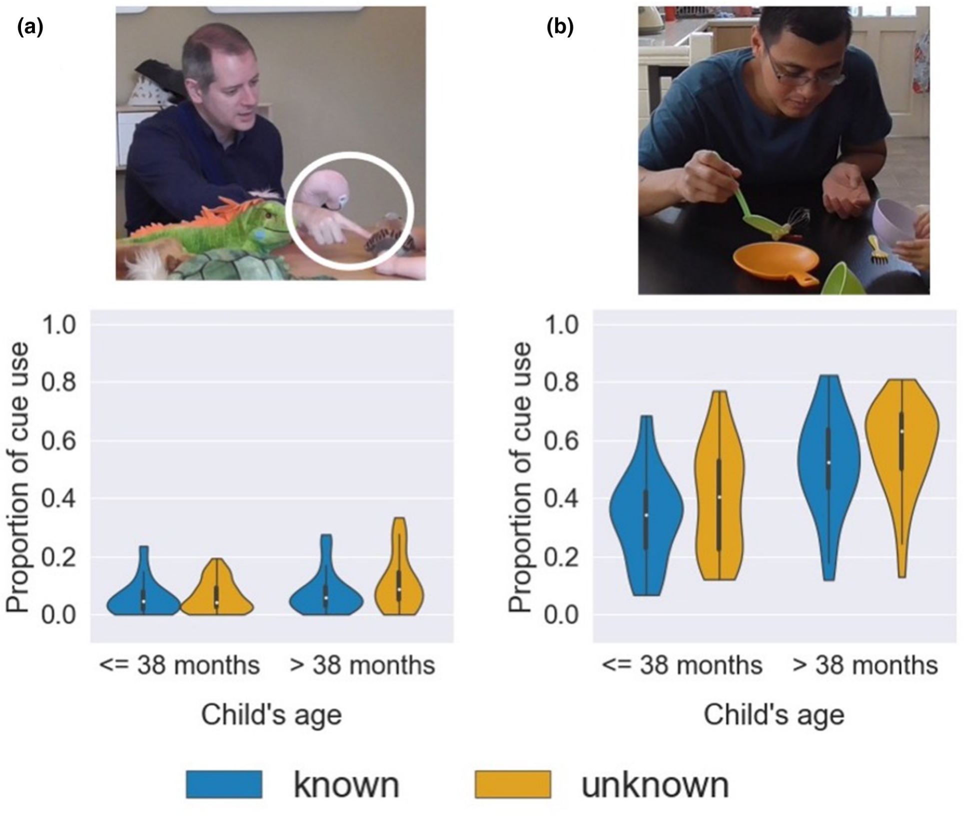

FIGURE 3 Indexical manual cue results: (a) Pointing gestures. A caregiver points to a toy bird the child is holding (top); plot shows proportion of utterances that contain a pointing gesture (bottom). (b) Object manipulations. A caregiver pretends to lift food from a pan using a spatula (top); plot shows proportion of utterances that contain an object manipulation (bottom). Plots show proportions across utterances, with age differences on the *x*-axis shown using a median split (38 months). Different colors represent known (blue) and unknown (yellow) toys respectively. Note that interactions with objects cannot occur without the objects present, so we only analyze the concurrent learning context for indexical cues.

TABLE 7 Object manipulation model results.

Effect	*β*	SE	*z*	*p*
Age	−.03	.01	−1.77	.08
Familiarity	.24	.05	4.37	<.001
Cohort	−1.19	.22	−5.36	<.001
Category	−.09	.005	−17.03	<.001
Age:familiarity	−.01	.007	−1.43	.15

TABLE 8 Gaze model results.

Effect	*β*	SE	*z*	*p*
Gaze overall				
Age	−.10	.02	−0.64	.52
Familiarity	.08	.07	1.28	.20
Category	−.03	.01	−2.30	.02
Age:familiarity	.02	.01	1.12	.26
Gaze with object manipulations				
Age	−.02	.01	−1.34	.18
Familiarity	.26	.07	3.86	<.001
Category	−.16	.01	−10.87	<.001
Age:familiarity	.003	.02	0.18	.86

can bring properties of referents to the mind's eye. We further hypothesized that all cues would be used more frequently in learning contexts, namely when the toy is unfamiliar to the child, suggesting that they do provide useful referential information rather than having a purely social-interactive function.

Indeed, our findings suggest that, in displaced learning contexts, caregivers provide rich representational input, in the form of iconic cues, that can offer support in linking labels to referents. Even in concurrent contexts, where the referent is present, caregivers combine speech with rich multimodal cues, especially indexical cues that provide a visual link to the object, helpful in a physical environment with multiple possible referents. Furthermore, we find that for all cues except pointing gestures, the child's existing knowledge of the object being referred to modulated caregivers' frequency of use, with caregivers producing cues more frequently when toys were unfamiliar to the child. Most caregiver–child interactions take place in face-to-face contexts; our study provides the first overarching picture of the wide range of iconic and indexical cues caregivers use in such interactions.

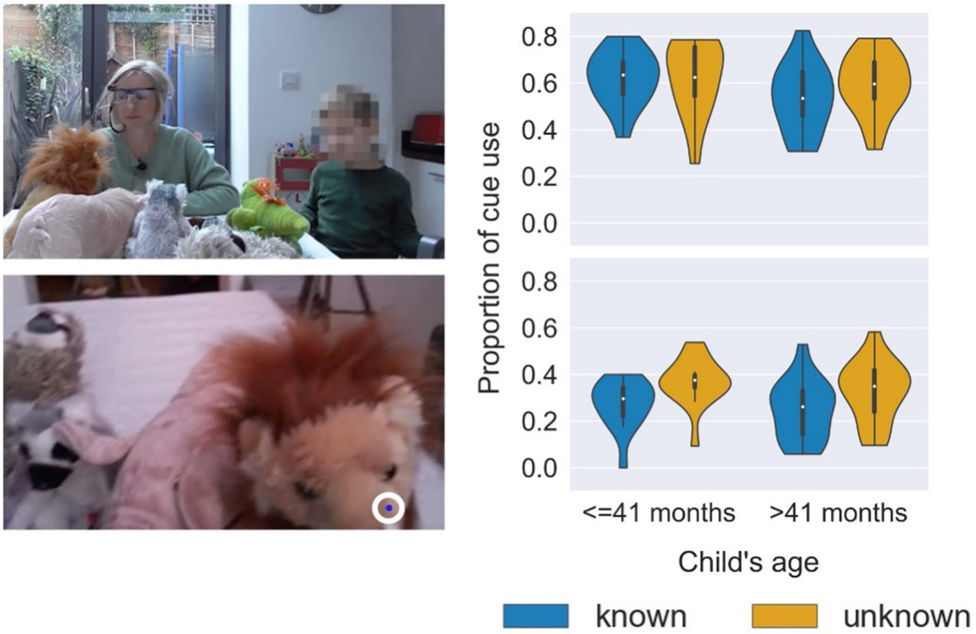

FIGURE 4 Gaze fixations on objects. The left hand panel provides an example of a gaze fixation on the lion toy, as captured by the main camera view (top) and by the Tobii eye tracking glasses (bottom, blue fixation dot circled in white). Plots in the right hand panel show the proportion of looks to objects for all gaze fixations (top) and for gaze fixations that co-occur with object manipulations (bottom). Plots show proportions across utterances, with age differences on the x-axis shown using a median split (41 months in this subset). Different colors represent known (blue) and unknown (yellow) toys respectively. As with other indexical cues, looks to objects are only analyzed in the concurrent learning context.

Iconic cues in displaced versus concurrent contexts

Our results are generally consistent with our hypothesis concerning the role of iconicity in displaced learning and provide further constraints to the hypothesis. Caregivers use vocal *and* gestural iconic cues (onomatopoeia, iconic prosody, and iconic gesture) more frequently in displaced contexts, namely when children cannot directly associate labels with referents in the environment and especially when the objects are unknown to the child, underscoring their importance in genuine learning episodes. There are large disparities in the overall frequency with which the different cues are used: roughly 24% of all utterances produced in displaced contexts and 3% of those in concurrent contexts were accompanied by iconic gestures; 4% of utterances produced in displaced and 3% of those produced in concurrent contexts contained onomatopoeia and, finally, 1% of the utterances in displaced and 0.45% of those in concurrent contexts contained iconic prosody. Thus, overall iconic gestures are the most common iconic cues present and the one that shows the greater difference between displaced and concurrent contexts. This may be explained in terms of the much larger iconic potential of the hands in comparison to the voice as there are many more imagistic links one can create in the visual than auditory modality (Taub, 2001). In particular, iconic prosody can only convey a very small number of semantic dimensions (duration, direction, and speed of motion primarily). Even with this in mind, it is somewhat surprising to see how rarely iconic prosody is used by caregivers. Of course, even very rare cues can be useful when present and it remains to be established whether and to what extent children between 2 and 4 can use iconic prosody in learning.

We have also assessed iconicity in non-onomatopoeic words produced by caregivers using available word-level iconicity ratings (Winter et al., 2023). Here, first, we did not find any difference between known and unknown objects' contexts, suggesting that the use of non-onomatopoeic iconic words is not necessarily more associated with learning episodes. Moreover, in contrast to our general hypothesis, we found that caregivers used iconic words more often in concurrent than displaced contexts. This is an interesting and unexpected finding for which we can only offer some speculations. We propose that a key distinction for interpreting this result is the one between *direct* and *indirect* iconicity. As already mentioned, the multimodal iconic cues that are the main focus of this paper all provide a *direct* and unimodal imagistic resemblance to properties of objects and events. This is not the case for the mappings provided by iconicity in the word form. For example, there is no direct, unimodal association between the phonological form of "up" or "look" (among the most common and most highly iconic in our corpus) and properties of the event (except mediated via iconic prosody for "up" and possibly orthography for "look"). Thus, it may be the case that in the more difficult displaced communicative settings caregivers prefer to use simpler, more direct forms of iconicity to support learning (as these cues are also more common for unknown than known referents) while this is not the case for indirect lexical iconicity. While this plausibly explains why we do not find higher rates of iconic words in displaced contexts, it does not account for why, instead, the rates are higher in the concurrent condition. Future research will need to establish the extent to which higher rates of iconic words in concurrent contexts relates to general communicative strategies, or perhaps simply to specific words that tend to occur more often in interacting with objects (as in the current study). Future research will also need to establish if the previously reported finding that verbs and adjectives tend to be more iconic than nouns (Imai & Kita, 2014; Perniss et al., 2018; Perry et al., 2015) also holds in our dataset.

Indexicality in caregivers' communication

Indexical cues—pointing, object manipulations, and looks to objects—were frequently used when toys were present, providing direct, visual links to objects. While we did not find that caregivers used eye gaze by itself differentially, we did find that looks to objects co-occurring with object manipulations were more frequent for unfamiliar toys. Previous work has shown that caregiver gaze fixations are often coupled with manual actions, coordinating behavior with the child on an object (Yu & Smith, 2013). Therefore, the role of gaze may lie in coordinating interactions between caregiver and child, rather than establishing the link between label and referent directly.

Crucially, the high frequency of use of these indexical cues in concurrent contexts indicates that finding larger numbers of iconic cues in displaced than concurrent is not simply a consequence of caregivers doing *more* in displaced contexts because they are more cognitively demanding for the children; rather it reflects caregivers using *qualitatively different* cues depending upon the context.

Developmental trajectories

We further found developmental patterns in the use of different cues. As we predicted and in line with previous work, we found that caregivers use onomatopoeia more frequently with younger children (Jo & Ko, 2018; Motamedi et al., 2021; Perry et al., 2017). We also observe that lexical iconicity decreases with age, though only for known referents, while the use of iconic words remains relatively stable across the age span for unknown referents, in line with findings by Perry et al. (2021), who reported greater iconicity in the words used by caregivers in utterances with novel labels. Iconic gestures, instead do not show any change within the developmental window under investigation, indicating that, once the ability to integrate information across modalities and to map this onto mental representations is acquired, iconic gestures can provide useful information in displaced contexts. For iconic prosody, we found that the use of this cue in the presence of unknown referents decreases with age just like other prosodic aspects of CDL and production of onomatopoeia. We do not find any clear developmental changes in the use of indexical cues in our data, except a trend toward an increase with age for object manipulations. While it is possible that the cross-modal nature of these cues in combination with speech, like iconic gestures, may require further cognitive development to be fully exploited (Marentette & Nicoladis, 2011; Namy, 2008), this explanation is at odds with the understanding that points appear very early in development (Woodward & Guajardo, 2002) and with our finding that iconic gestures do not increase with the age of the children. Fully understanding these trends in relation to children's linguistic abilities is beyond the scope of the current work, but one possibility is that these cues increase with increasing complexity in the interaction itself (e.g., points come to have anaphorical reference and can be used to indicate both whole objects and parts of objects; object manipulations can comprise complex illustrations of the use of novel objects). Therefore, it may be that use and understanding of these cues develops beyond the age range we cover in this study.

Multimodal cues in learning and communication

We have documented here the use of multimodal cues by caregivers in interaction with their child and discussed them primarily in terms of the potential use of these cues in supporting children's learning (Jo & Ko, 2018; Ma et al., 2011; Ota et al., 2018; Perniss et al., 2018; Rowe et al., 2008). These multimodal cues are found in communicative interactions in general, not only when the addressee is presented with a new concept or word. In our data, we see that *all* cues are substantially used both with unfamiliar (hence when there is learning potential) and familiar topics: overall, 60% of utterances about unknown referents contained some vocal or gestural cue versus approximately 52% of utterances about known referents. This fact suggests that multimodal cues can play a general role in communication, regardless of whether the content is known or unknown by the addressee. This is not surprising and suggests that these cues contribute to the building of situational models for what is being discussed. Going a step further, in other work, we have proposed that cues such as iconic gestures are produced in contexts in which they can support successful communication, namely when what the speaker is about to say is less predictable from the context (Grzyb et al., 2022).

Considering specifically *iconic* cues, while languages differ with respect to the extent they embed iconicity in their linguistic form, all languages exploit iconicity to some degree, and, as discussed in this paper, iconicity is present in communication beyond the linguistic form (namely, in iconic prosody and gestures; Murgiano et al., 2021). Our findings underscore the utility of iconic forms in imagistically evoking properties of referents, particularly in instances of displacement. The potential of iconicity to support displaced learning and displaced communication in general offers a novel hypothesis for why iconicity would be maintained across languages (Perniss & Vigliocco, 2014).

CONCLUSIONS

Our study offers important insight into the multimodal nature of child-directed communication. We show that children receive input from a wide range of representational sources that are used dynamically in an interactive system that supports learning (Rączaszek-Leonardi et al., 2018). Prior research has often considered these cues marginal in the face of arbitrary speech that is considered the primary linguistic input, suggesting that, for example, onomatopoeia are a minor part of our vocabularies (Newmeyer, 1992). Indeed, though we find that caregivers modify onomatopoeia and prosody use based on displacement, toy familiarity, and their child's age, the overall usage of these cues is relatively low. However, when taken together, these cues, in addition to other cues in speech and beyond (e.g., facial expression), collectively offer a rich representational input that can be dynamically exploited in context to support learning. That is, the quantity of individual cues may be secondary to the overall quality of the input children receive (Cartmill et al., 2013).

If a primary goal of developmental research is to understand how children learn from the input they receive, then a comprehensive understanding of what that input *is* is the first key step to attaining that goal. Avenues for future study should aim to understand the diversity of multimodal communication cross-linguistically (Perniss, 2018; Perniss et al., 2018), to understand the interaction between these cues in combination and (noniconic) caregiver speech, and to assess the causal links between the multimodal cue use shown here and children's learning outcomes (Motamedi et al., 2021; Vosoughi et al., 2010; Yu et al., 2019).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The annotated data (i.e., not audiovisual files) and analytic code necessary to reproduce the analyses presented here are publicly accessible at the project's OSF page. The analyses presented here were not preregistered.

STATEMENT ON CONFIRMATORY VERSUS EXPLORATORY RESEARCH

The study reported here represents a mixture of confirmatory and exploratory research. While we had specific hypotheses regarding the presence of iconic cues in concurrent and displaced contexts, we took a more exploratory view with respect to developmental effects related to each cue, such that we did not make strong predictions about whether the nonvocal cues would be present more often in interactions with younger or older children.

ORCID

Yasamin Motamedi https://orcid.org/0000-0001-8846-5387

REFERENCES

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development*, *59*(2), 450–466.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

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

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (6.1.02) [Computer software]. http://www.praat.org

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1), 72–83. https://doi.org/10.1111/1467-7687.00211

Brand, R. J., & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science*, *11*(6), 853–861. https://doi.org/10.1111/j.1467-7687.2008.00734.x

Brand, R. J., Shallcross, W. L., Sabatos, M. G., & Massie, K. P. (2007). Fine-grained analysis of Motionese: Eye gaze, object exchanges, and action units in infant-versus adult-directed action. *Infancy*, *11*(2), 203–214. https://doi.org/10.1111/j.1532-7078.2007.tb00223.x

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Campisi, E., & Ozyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14–27. https://doi.org/10.1016/j.pragma.2012.12.007

Cartmill, E., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the*

National Academy of Sciences of the United States of America, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Clark, E. V., & Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture*, *11*(1), 1–23. https://doi.org/10.1075/gest.11.1.01cla

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595. https://doi.org/10.2307/1130766

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

de Saussure, F. (1916). *Cours de linguistique générale*. Payot.

D'Entremont, B. (2000). A perceptual–attentional explanation of gaze following in 3- and 6-month-olds. *Developmental Science*, *3*(3), 302–311. https://doi.org/10.1111/1467-7687.00124

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130301. https://doi.org/10.1098/rstb.2013.0301

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9602–9605. https://doi.org/10.1073/pnas.152159999

Fernald, A. (2000). Speech to infants as Hyperspeech: Knowledge-driven processes in early word recognition. *Phonetica*, *57*(2–4), 242–254. https://doi.org/10.1159/000028477

Gao, C., Shinkareva, S. V., & Desai, R. H. (2022). SCOPE: The South Carolina psycholinguistic metabase. *Behavior Research Methods*, *55*, 2853–2884. https://doi.org/10.3758/s13428-022-01934-0

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

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language–and creating language. *Child Development*, *78*(3), 741–745. https://doi.org/10.1111/j.1467-8624.2007.01029.x

Goodrich, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science*, *12*(1), 81–87. https://doi.org/10.1111/j.1467-7687.2008.00735.x

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824. https://doi.org/10.1111/infa.12006

Grimminger, A., Rohlfing, K. J., Lüke, C., Liszkowski, U., & Ritterfeld, U. (2020). Decontextualized talk in caregivers' input to 12-month-old children during structured interaction. *Journal of Child Language*, *47*(2), 418–434. https://doi.org/10.1017/S0305000919000710

Grzyb, B., Frank, S. L., & Vigliocco, G. (2022). Communicative efficiency in multimodal language directed to children and adults. *PsyArXiv*. https://doi.org/10.31234/osf.io/a9wt3

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory: Lower scores in the UK than the USA. *Journal of Child Language*, *27*(3), 689–705. https://doi.org/10.1017/S0305000900004414

Herold, D. S., Nygaard, L. C., Chicos, K. A., & Namy, L. L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology*, *108*(2), 229–241. https://doi.org/10.1016/j.jecp.2010.09.005

Herold, D. S., Nygaard, L. C., & Namy, L. L. (2012). Say it like you mean it: Mothers' use of prosody to convey word meaning. *Language and Speech*, *55*(3), 423–436. https://doi.org/10.1177/0023830911422212

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 88–96.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesture in mother–child interactions. *Cognitive Development*, *14*, 57–75.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Jo, J., & Ko, E.-S. (2018). Korean mothers attune the frequency and acoustic saliency of sound symbolic words to the linguistic maturity of their children. *Frontiers in Psychology*, *9*, 2225. https://doi.org/10.3389/fpsyg.2018.02225

Johansson, N., Anikin, A., & Aseyev, N. (2020). Color sound symbolism in natural languages. *Language and Cognition*, *12*(1), 56–83. https://doi.org/10.1017/langcog.2019.35

Johansson, N., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*, *24*(2), 253–310. https://doi.org/10.1515/lingty-2020-2034

Johansson, N., & Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: A typological study of 101 languages. *Public Journal of Semiotics*, *5*(1), 1–20. https://doi.org/10.37693/pjos.2013.5.9668

Joo, I. (2020). Phonosemantic biases found in Leipzig-Jakarta lists of 66 languages. *Linguistic Typology*, *24*(1), 1–12. https://doi.org/10.1515/lingty-2019-0030

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347.

Laing, C. E. (2014). A phonological analysis of onomatopoeia in early word production. *First Language*, *34*(5), 387–405. https://doi.org/10.1177/0142723714550110

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Marentette, P., & Nicoladis, E. (2011). Preschoolers' interpretations of gesture: Label or action associate? *Cognition*, *121*(3), 386–399. https://doi.org/10.1016/j.cognition.2011.08.012

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, *21*(2), 373–377. https://doi.org/10.1016/S0163-6383(98)90014-5

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), 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), 1–18. https://doi.org/10.5334/joc.113

Namy, L. L. (2008). Recognition of iconicity doesn't come for free. *Developmental Science*, *11*(6), 841–846. https://doi.org/10.1111/j.1467-7687.2008.00732.x

Namy, L. L., Campbell, A. L., & Tomasello, M. (2004). The changing role of iconicity in non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development*, *5*(1), 37–57. https://doi.org/10.1207/s15327647jcd0501_3

Newmeyer, F. J. (1992). Iconicity and generative grammar. *Language*, *68*(4), 756–796. https://doi.org/10.2307/416852

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*(1), 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x

Ortega, G., Sumer, B., & Ozyürek, A. (2017). Type of iconicity matters in the vocabulary development of signing children. *Developmental Psychology*, *53*, 89–99.

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why choo-choo is better than train: The role of register-specific words in early vocabulary growth. *Cognitive Science*, *42*(6), 1974–1999. https://doi.org/10.1111/cogs.12628

Ozcaliskan, S., Gentner, D., & Goldin-Meadow, S. (2013). Do iconic gestures pave the way for children's early verbs? *Applied Psycholinguistics*, *35*, 1143–1162.

Özçalişkan, Ş., & Goldin-Meadow, S. (2005). Do parents lead their children by the hand? *Journal of Child Language*, *32*(3), 481–505. https://doi.org/10.1017/S0305000905007002

Perlman, M., Little, H., Thompson, B., & Thompson, R. L. (2018). Iconicity in signed and spoken vocabulary: A comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in Psychology*, *9*, 1433. https://doi.org/10.3389/fpsyg.2018.01433

Perniss, P. (2018). Why we should study multimodal language. *Frontiers in Psychology*, *9*(June), 1–5. https://doi.org/10.3389/fpsyg.2018.01109

Perniss, P., Lu, J. C., Morgan, G., & Vigliocco, G. (2018). Mapping language to the world: The role of iconicity in the sign language input. *Developmental Science*, *21*(2), e12551. https://doi.org/10.1111/desc.12551

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*(DEC), 1–15. https://doi.org/10.3389/fpsyg.2010.00227

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *369*, 20130300. https://doi.org/10.1098/rstb.2013.0300

Perry, L. K., Custode, S. A., Fasano, R. M., Gonzalez, B. M., & Savy, J. D. (2021). What is the buzz about iconicity? How iconicity in caregiver speech supports children's word learning. *Cognitive Science*, *45*(4), e12976. https://doi.org/10.1111/cogs.12976

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS One*, *10*(9), 1–17. https://doi.org/10.1371/journal.pone.0137147

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2017). Iconicity in the speech of children and adults. *Developmental Science*, *21*(3), 1–8. https://doi.org/10.1111/desc.12572

R Core Team. (2013). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing. http://www.R-project.org/

Rączaszek-Leonardi, J., Nomikou, I., Rohlfing, K. J., & Deacon, T. W. (2018). Language development from an ecological perspective: Ecologically valid ways to abstract symbols. *Ecological Psychology*, *30*(1), 39–73. https://doi.org/10.1080/10407413.2017.1410387

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

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M.-C., & Cohen, D. (2013). Motherese in interaction: At the cross-road of emotion and cognition? (a systematic review). *PLoS One*, *8*(10), e78103. https://doi.org/10.1371/journal.pone.0078103

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Shintel, H., Anderson, N. L., & Fenn, K. M. (2014). Talk this way: The effect of prosodically conveyed semantic information on memory for novel words. *Journal of Experimental Psychology. General*, *143*(4), 1437–1442.

Sloetjes, H., & Wittenberg, P. (2008). Annotation by category—ELAN and ISO DCR. *Proceedings of the 6th international conference on language resources and evaluation*.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, *10*(1), 1–15. https://doi.org/10.1017/s0305000900005092

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929–938. https://doi.org/10.1037/0012-1649.44.4.929

Taub, S. (2001). *Language from the body: Iconicity and metaphor in American sign language*. Cambridge University Press.

Tobii_glasses2_utils. (2017). *Python*. https://github.com/pearsonlab/tobii_glasses2_utils. (Original work published 2015)

Tolar, T. D., Lederberg, A. R., Gokhale, S., & Tomasello, M. (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education*, *13*(2), 225–240. https://doi.org/10.1093/deafed/enm045

Tomasello, M., & Barton, M. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, *30*(5), 639–650. https://doi.org/10.1037/0012-1649.30.5.639

Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, *19*(2), 311–333. https://doi.org/10.1017/S0305000900011430

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, *23*(1), 157–176. https://doi.org/10.1017/S0305000900010138

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Veneziano, E. (2001). Displacement and informativeness in child-directed talk. *First Language*, *21*(63), 323–356. https://doi.org/10.1177/014272370102106306

Vogt, S., & Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *Journal of Child Language*,*44*(6), 1458–1484. https://doi.org/10.1017/S0305000916000647

von Hofsten, C., Dahlström, E., & Fredriksson, Y. (2005). 12-month-old infants' perception of attention direction in static video images. *Infancy*, *8*(3), 217–231. https://doi.org/10.1207/s15327078in0803_2

Vosoughi, S., Roy, B. C., Frank, M. C., & Roy, D. (2010). Contributions of prosodic and distributional features of caregiver's speech in early word learning. In S. Ohlsson (Ed.), *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 1822–1827).

Wichmann, S., Holman, E. W., & Brown, C. H. (2010). Sound symbolism in basic vocabulary. *Entropy*, *12*(4), 844–858. https://doi.org/10.3390/e12040844

Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. (2023). Iconicity ratings for 14,000+ English words. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02112-6

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic? *Interaction Studies*, *18*(3), 430–451. https://doi.org/10.1075/is.18.3.07win

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, *12*(1), 1035. https://doi.org/10.1038/s41598-021-04311-7

Woodward, A. L., & Guajardo, J. J. (2002). Infants' understanding of the point gesture as an object-directed action. *Cognitive Development*, *17*(1), 1061–1084. https://doi.org/10.1016/S0885-2014(02)00074-6

Yoshida, H. (2012). A cross-linguistic study of sound symbolism in children's verb learning. *Journal of Cognition and Development*, *13*(2), 232–265. https://doi.org/10.1080/15248372.2011.573515

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. https://doi.org/10.1371/journal.pone.0079659

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1), e12735. https://doi.org/10.1111/desc.12735

How to cite this article: Motamedi, Y., Murgiano, M., Grzyb, B., Gu, Y., Kewenig, V., Brieke, R., Donnellan, E., Marshall, C., Wonnacott, E., Perniss, P., & Vigliocco, G. (2024). Language development beyond the here-and-now: Iconicity and displacement in child-directed communication. *Child Development*, *00*, 1–19. https://doi.org/10.1111/cdev.14099