



# Infant-directed communication: Examining the many dimensions of everyday caregiver-infant interactions

Jessica E. Kosie<sup>1,2</sup>  | Casey Lew-Williams<sup>1</sup>

<sup>1</sup>Department of Psychology, Princeton University, Princeton, New Jersey, USA

<sup>2</sup>School of Social and Behavioral Sciences, Arizona State University, Phoenix, Arizona, USA

## Correspondence

Jessica E. Kosie, School of Social and Behavioral Sciences, Arizona State University, Phoenix, Arizona 85306, USA.  
Email: [jkosie@asu.edu](mailto:jkosie@asu.edu)

## Funding information

National Science Foundation SBE Postdoctoral Research Fellowship, Grant/Award Number: 2004983; National Institute of Child Health and Human Development, Grant/Award Numbers: F32HD103439, R01HD095912

## Abstract

Everyday caregiver-infant interactions are dynamic and multidimensional. However, existing research underestimates the dimensionality of infants' experiences, often focusing on one or two communicative signals (e.g., speech alone, or speech and gesture together). Here, we introduce "infant-directed communication" (IDC): the suite of communicative signals from caregivers to infants including speech, action, gesture, emotion, and touch. We recorded 10 min of at-home play between 44 caregivers and their 18- to 24-month-old infants from predominantly white, middle-class, English-speaking families in the United States. Interactions were coded for five dimensions of IDC as well as infants' gestures and vocalizations. Most caregivers used all five dimensions of IDC throughout the interaction, and these dimensions frequently overlapped. For example, over 60% of the speech that infants heard was accompanied by one or more non-verbal communicative cues. However, we saw marked variation across caregivers in their use of IDC, likely reflecting tailored communication to the behaviors and abilities of their infant. Moreover, caregivers systematically increased the dimensionality of IDC, using more overlapping cues in response to infant gestures and vocalizations, and more IDC with infants who had smaller vocabularies. Understanding how and when caregivers use all five signals—together and separately—in interactions with infants has the potential to redefine how developmental scientists conceive of infants' communicative environments, and enhance our understanding of the relations between caregiver input and early learning.

## KEYWORDS

caregiver-infant interaction, infancy, multimodal communication

## Research Highlights

- Infants' everyday interactions with caregivers are dynamic and multimodal, but existing research has underestimated the multidimensionality (i.e., the diversity of simultaneously occurring communicative cues) inherent in infant-directed communication.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Developmental Science* published by John Wiley & Sons Ltd.



- Over 60% of the speech that infants encounter during at-home, free play interactions overlap with one or more of a variety of non-speech communicative cues.
- The multidimensionality of caregivers' communicative cues increases in response to infants' gestures and vocalizations, providing new information about how infants' own behaviors shape their input.
- These findings emphasize the importance of understanding how caregivers use a diverse set of communicative behaviors—both separately and together—during everyday interactions with infants.

## 1 | INTRODUCTION

Everyday caregiver-infant interactions are dynamic and multimodal (e.g., Frank et al., 2013; Holler & Levinson, 2019; Perniss, 2018; Schatz et al., 2022; Suarez-Rivera et al., 2022; Yu & Smith, 2012). A variety of studies have demonstrated that caregivers use and modify a diverse set of communicative behaviors—including speech, action, gesture, emotion, and touch—while interacting with infants (e.g., Brand et al., 2002; Chong et al., 2003; Fernald et al., 1989; Iverson et al., 1999; Stack & Muir, 1990), and these infant-directed modifications are linked to infants' attention and learning (e.g., Abu-Zhaya et al., 2017; Dimitrova & Moro, 2013; Gogate et al., 2000; Koterba & Iverson, 2009; Lew-Williams et al., 2019; Moses et al., 2001; O'Neill et al., 2005; Rowe, 2000; Seidl et al., 2015; Stack & Muir, 1992; Tamis-LeMonda et al., 2008; Tincoff et al., 2019; Williamson & Brand, 2014; Wu et al., 2021). However, these behaviors are frequently treated as separable components of communication, and most research focuses on one or two dimensions at a time (e.g., speech alone, or speech and gesture together), which results in an oversimplification of how communication unfolds during natural caregiver-infant interactions. Caregiver-infant communication is highly multidimensional—it is easy to imagine a caregiver saying “Wow, look!” while pointing at an object and showing surprise on their face. In this case, speech, gesture, and emotion would all be occurring simultaneously and working together to create meaning. A richer characterization of this natural multidimensionality is necessary for the field of developmental science to construct a comprehensive framework for understanding how communicative cues are used and integrated during everyday caregiver-infant interactions.

The goal of the current study is to address this limitation by investigating “infant-directed communication” (IDC): the suite of communicative signals from caregivers to infants including speech, action, gesture, emotion, and touch. We first review existing literature on how each of these dimensions, independently, are used and modified in caregiver-infant interactions. Next, we examine 10-min, at-home free play sessions to characterize and quantify caregivers' use of this suite of multidimensional cues. With this comprehensive characterization of caregiver-infant communication, we take a new approach to studying turn-taking to better understand not only whether caregivers respond to infant vocalizations and gestures, but also the extent

to which they increase the dimensionality of their communication (i.e., providing combinations of cues across multiple modalities) at key moments in dynamic interactions. Finally, we explore how caregivers' use of multidimensional communication is modulated with infants' growing language skills. Understanding how and when caregivers use infant-directed communication provides a more complete picture of infants' real-world learning environment, deepens our understanding of caregiver input and early learning, and enables the creation of ecologically-valid theories of infants' learning from natural, everyday interactions.

### 1.1 | Infant-directedness across modalities

*Infant-directed speech.* It has been well-established that many caregivers (and older children; e.g., Shatz & Gelman, 1973) modify their speech when interacting with infants (e.g., Ferguson, 1964; Fernald et al., 1989; Hilton et al., 2022; Kuhl et al., 1997; Piazza et al., 2017; Snow & Ferguson, 1977). Infant-directed speech (IDS) features higher and more variable pitch, shorter utterances, increased repetition, and simplified vocabulary, among other characteristics that make it distinct from the type of speech that is typically directed to adults (Fernald et al., 1989). These adaptations to IDS have a number of benefits for young learners. For example, there is robust evidence that infants prefer to listen to infant- over adult-directed speech (ManyBabies Consortium, 2020; Cooper & Aslin, 1990; Fernald, 1985), which is likely driven by optimized neural entrainment to its moment-to-moment dynamics (Nencheva & Lew-Williams, 2022). Additionally, the characteristics of IDS appear to promote early word learning by, for example, helping infants discriminate speech sounds and segment words out of continuous speech (Golinkoff et al., 2015; Graf Estes & Hurley, 2013; Ma et al., 2011; Soderstrom, 2007; Thiessen et al., 2005; Trainor & Desjardins, 2002).

IDS is clearly a key component of infant-directed communication (IDC), but speech does not occur in isolation. Communication is *much* more than just speech alone; it involves a suite of infant-directed communicative signals across multiple modalities. These important communicative signals include (but are not limited to): action, gesture, emotion, and touch. While each of these signals is typically studied



in isolation, there is independent evidence that caregivers modify their behaviors in each of these dimensions when interacting with infants and that these modifications promote successful caregiver-infant communication. In many cases, these signals may be overlapping, simultaneous, or even synchronous or redundant with one another. Below, we describe each of these signals and their benefits for early development.

**Infant-directed action.** When interacting with infants, caregivers act on objects as they talk about them (e.g., Karmazyn-Raz & Smith, 2023; Meyer et al., 2011; Schatz et al., 2022; Suanda et al., 2016; Suarez-Rivera et al., 2022), and there is evidence that caregivers modify these object-directed actions in ways that are analogous to the modifications observed in IDS. Action demonstrations directed to infants (vs. adults) are more enthusiastic, repetitive, simplified, use a larger range of motion, and are performed closer to the infants' space (Brand et al., 2002). Infants prefer to look at infant-directed over adult-directed action, and they are more likely to imitate and explore objects that have been demonstrated using these infant-directed features (Brand & Shallcross, 2008; Koterba & Iverson, 2009; Meyer et al., 2022; Williamson & Brand, 2014). Furthermore, caregivers frequently manipulate objects as they simultaneously make a verbal reference to that same object (Gogate et al., 2000; Messer, 1978), suggesting that caregivers may use action to reinforce their speech. Additionally, just as IDS seems to help infants segment words from speech, recent research suggests that infant-directed action may help infants segment individual action units from unfolding activity streams (Kosie & Baldwin, 2023). Action, too, thus appears to be an important feature of everyday caregiver-infant communication.

**Infant-directed gesture.** Gesture is another common and widely investigated feature of caregiver-infant interactions (e.g., Goldin-Meadow, 2005; Morganstern, 2024; Rowe et al., 2008; Schmidt, 1996; Vigliocco et al., 2019). Like speech and action, there is evidence that caregivers modify their gestures when interacting with infants and that caregivers' use of gestures changes with infant development. For example, gestures to infants are initially simple and concrete (e.g., frequent pointing and reference to the immediate environment), but both the amount and complexity of gestures increase with infants' age and abilities (Dimitrova & Moro, 2013; Gogate et al., 2000; Iverson et al., 1999; O'Neill et al., 2005; Shatz, 1982). As with infant-directed speech and action, caregivers' gesture use impacts infants learning and behavior. For example, it has been demonstrated that caregiver gesture directs infants' attention to relevant features of the environment and has been linked to more accurate word learning (Booth et al., 2008; Vogt & Kauschke, 2017; Zukow, 1991). Additionally, caregivers' gesture use is positively predictive of infants' gesture use, which in turn predicts their later vocabulary size (Rowe, 2000; Rowe & Goldin-Meadow, 2009). Given this evidence, it is clear that any comprehensive conceptualization of communication should necessarily include gesture.

**Infant-directed emotion.** Emotion, too, is a primary source of information that infants use in everyday communication with caregivers (for a review, see Wu et al., 2021). While emotion can be conveyed in multiple modalities including facial expression, prosody, and movement dynamics, we chose to operationalize "emotion" as facial expressions

that caregivers often use and modify (e.g., through exaggeration) to express emotional content. For example, Chong et al. (2003) described the form and meaning of three facial expressions that are frequently used and exaggerated during caregiver-infant interactions to express love/concern, joy, and surprise. Caregivers use facial expressions to communicate emotional content, and infants use them to gather information that guides their exploration and learning. When conditions are uncertain, such as activities involving interaction with novel toys or traversing seemingly hazardous paths (e.g., a "visual cliff"), infants use their caregivers' facial expressions as one source of information to guide their own actions. Specifically, infants are more likely to avoid novel toys or situations when their caregiver shows negative emotions and approach the toys/situations when the caregiver shows neutral or positive emotions (Moses et al., 2001; Tamis-LeMonda et al., 2008). In addition to speech, action, and gesture, facial expressions conveying emotion are yet another way that caregivers provide communicative signals to infants.

**Infant-directed touch.** Caregivers' use of touch plays an important role in everyday interactions with infants as well (Anisfeld et al., 1990; Feldman et al., 2010; Ferber et al., 2008; Franco et al., 1996; Hertenstein, 2002; Jean et al., 2009; Stack & Arnold, 1998; Stack & Muir, 1990). Existing research on the role of touch in early language and auditory learning suggests that caregivers' touch can both enhance infants' ability to find words in speech and to learn new words (Abu-Zhaya et al., 2017; Lew-Williams et al., 2019; Seidl et al., 2015; Seidl et al., 2024; Stack & Muir, 1990; Tincoff et al., 2019). For example, Lew-Williams et al. (2019) found that combining touch and auditory information promoted infants' learning of auditory patterns, while infants failed to learn these same patterns in the absence of redundant touch cues. Touch, too, is a common source of information that promotes successful communication between caregivers and their infants.

While the research summarized above provides important information about caregivers' use of multiple communicative cues during interactions with infants, these studies often focus on a single dimension at a time (or sometimes a single non-speech dimension alongside speech). To our knowledge, prior work has not aimed to provide a comprehensive view of communication with infants, yet an integration across domains is important for understanding the real multidimensionality of everyday interactions. As a result, the field is currently working from an impoverished picture of how communication naturally unfolds between caregivers and infants. By examining multiple communicative behaviors in parallel, we aim to characterize the full richness of infants' everyday social environments, and in doing so, improve frameworks for understanding the varied contexts of early learning.

## 1.2 | Multimodal signals support learning

Overlap between different communicative behaviors is likely to support infants' learning. For example, Bahrack et al. proposed that the use of redundant cues across modalities (which they refer to as



intersensory redundancy) serves to direct infants' attention to features of a situation or stimulus that have been redundantly specified and attenuate infants' attention to features that are not redundant (Bahrick et al., 2004; Bahrick & Lickliter, 2000). More generally, the use of multimodal cues has been shown to aid learning in a variety of structured in-lab tasks across dimensions that include perception and discrimination of rhythm, tempo, and affect, learning of abstract rules, and mapping words or sounds to objects (Bahrick et al., 2002; Booth et al., 2008; Flom & Bahrick, 2007; Frank et al., 2009; Gogate et al., 2000; Gogate & Bahrick, 1998).

Multimodality promotes attention and learning in more natural interactions as well. For example, mothers' multimodal engagement with objects (e.g., simultaneously touching and talking about an object) during play extends both the duration of play bouts and infants' attention to relevant objects (in contrast to instances that do not involve multimodality; Schatz et al., 2022; Suarez-Rivera et al., 2019; Suarez-Rivera et al., 2022). Caregivers also use multimodal cues to initiate bouts of joint attention and may tailor their use of these cues to the needs of their child (Depowski et al., 2015; Gabouer et al., 2018, 2020). For example, while there is substantial variation in how caregivers use multimodal cues with both hearing and deaf children, caregivers of deaf children may be more likely to accommodate their child's needs by incorporating the use of more tactile cues overall (Abu-Zhaya et al., 2019; Gabouer et al., 2020). However, less is known about how caregivers use and adapt a broad suite of multimodal cues in response to infants as dynamic interactions unfold across time.

### 1.3 | Infants' role in shaping early input

As interactions unfold, caregivers and infants engage in reciprocal turn-taking in which one individual initiates a behavior and the other responds contingently (Bornstein et al., 1992, 2015; Gratier et al., 2015; Holler et al., 2015; Levinson, 2006; Sacks et al., 1978; Stern, 1985). Infants use a variety of behaviors to elicit responses from caregivers (e.g., Bates et al., 1975; Begus & Southgate, 2012; Gros-Louis et al., 2006; Kovács et al., 2014; Lucca & Wilbourn, 2019). For example, vocalizations directed at objects or speech-like babbling by infants are behaviors likely to receive a caregiver response (Albert et al., 2018; Elmlinger et al., 2019; Goldstein et al., 2010) and infants use gestures, like pointing, to request information from others (Kovács et al., 2014; Lucca & Wilbourn, 2019). Caregivers' contingent responses to infants have been linked to the development of attachment, cognitive skills, vocal behavior (i.e., the maturity of infant vocalizations), and language across infancy and early childhood (Baumwell et al., 1997; Goldstein et al., 2003; Goldstein & Schwade, 2008; Jaffe et al., 2001; Nicely et al., 1999; Rollins, 2003; Tamis-LeMonda et al., 2001). Contingent caregiver responses include speech and vocalizations as well as non-verbal behaviors such as facial expressions, gestures, actions on objects, and touch (e.g., Deak, 2018; Gros-Louis et al., 2006, 2014; Nicely et al., 1999). While we might expect caregivers to also employ more multimodal cues in response to infants' vocalizations and gestures, the coding of caregivers' responses is frequently binary (e.g.,

simply whether they responded or not) and focused on one single communicative dimension at a time. It is currently unknown whether caregivers increase the dimensionality of their communicative acts in response to their infant, which may be ideal for infant engagement and learning.

### 1.4 | Characterizing infant-directed communication (IDC)

Existing evidence suggests that: (1) caregivers use a variety of multimodal cues in communicative interactions with infants, (2) multimodal cues support early learning, and (3) multimodal cues may be tailored to infants' behavior. However, the majority of research on the dynamics of caregiver-infant interaction has investigated each multimodal dimension independently, and much of our understanding of how multimodal cues support learning comes from structured, in-lab tasks. We know of no work to date that investigates how caregivers dynamically use and modulate *all five* multimodal cues—both independently and as they overlap with one another—during dynamic interactions with their infants. Thus, the aim of the current study is to expand beyond a dominant focus on IDS to more fully capture the range of communicative behaviors in natural caregiver-infant interactions.

To better understand how IDC occurs in everyday caregiver-infant interactions, we video-recorded caregivers and their 18- to 24-month-old infants during free play at home (via Zoom). Interactions were coded continuously for caregivers' use of each of the five dimensions of IDC—speech, actions on objects, gesture, emotion (operationalized as facial expressions), and touch on the infants' body—as well as infants' vocalizations and gestures. Because of the degree of complexity introduced by our highly detailed coding scheme, we were limited in the amount of data that was feasible to code and opted for brief, 10-min recordings. Though these recordings represent only a tiny sample of infants' lives, there is existing evidence that brief, at-home recordings like these do capture moments within infants' everyday experiences that closely match periods of high interaction with caregivers (Bergelson et al., 2019; Manning et al., 2020; Tamis-LeMonda et al., 2017). These data were analyzed with the following objectives:

1. First, we sought to characterize the extent to which caregivers used IDC during 10-min interactions with their infants. We examined how each dimension of IDC was used independently as well as how frequently dimensions overlapped. We predicted that caregivers would use dimensions of IDC throughout the entire interaction, but that their use of IDC would vary, both across dimensions and across caregivers. We also expected that speech would rarely occur in isolation and that—even when speech was not occurring—caregivers would still communicate via other non-speech dimensions.
2. We then asked how caregivers modulate their use of IDC in response to infant behaviors, namely, vocalizations and gestures. Specifically, we tested whether the dimensionality of communication (operationalized as the number of cues used simultaneously by caregivers) increased in the period surrounding an infant

vocalization or gesture, relative to regions not surrounding an infant vocalization or gesture. If infants respond to higher dimensionality in caregivers' communication by vocalizing or gesturing, then we might expect the number of overlapping cues used by caregivers to be higher in the seconds *before* an infant behavior. In contrast, if caregivers respond to infants by increasing the dimensionality of their communication, then the number of overlapping dimensions would be higher in the seconds *after* an infant vocalization or gesture.

3. Finally, we engaged in exploratory analyses to investigate relations between caregivers' use of IDC and infants' language skill, as measured by the MacArthur-Bates Communicative Development Inventory (MCDI; Fenson et al., 2007).

If our goal as developmental scientists is to understand how infants learn in the natural, everyday context in which this learning occurs, we need to better understand and characterize how multidimensional, dynamic, communicative caregiver-infant interactions unfold in infants' everyday, real-world experience.

## 2 | METHOD

### 2.1 | Participants

A total of 44 caregivers (41 mothers, 3 fathers) and their 18- to 24-month-old infants ( $M_{age} = 20.48$ ,  $SD_{age} = 1.87$ , 21 female) were recruited from Central New Jersey and the surrounding area. This sample size was determined based on the number of families we were able to recruit during one summer and is similar to other recent studies examining measures of caregiver-infant communication (which typically include approximately 40 caregiver-infant dyads; e.g., Schatz et al., 2022; Suarez-Rivera et al., 2022). All infants were monolingual (i.e., exposed to at least 80% English in their daily lives), full-term (born after 37 weeks gestation), and had no reported developmental delays. Infants were predominantly white ( $n = 34$  White;  $n = 2$  Asian; and  $n = 8$  more than one race) and came from higher-SES households. All mothers had graduated high school, and most had completed college ( $n = 14$ ; 32%) or graduate school ( $n = 27$ ; 61%).

Given these participant characteristics, we ask readers to keep in mind two points. First, as with many studies of early development, we are using a restricted convenience sample (see Kidd & Garcia, 2022; Singh et al., 2023). All of our participants come from relatively higher-SES, predominantly white, English-speaking families in the Northeast of the United States. Thus, while some of our findings may reflect characteristics of caregiver-infant interactions that apply across communities and cultures, we do not intend to make broad generalizations beyond this particular group. Second, we strive to consider and appreciate the substantial variation that exists even across families in this restricted sample. Our goal in describing variation is not to place value on families who fall on either end of the spectrum. Instead, we



**FIGURE 1** Example image depicting the Zoom recording setup.

encourage appreciation of the natural variation that exists even in this relatively homogenous sample.

### 2.2 | Procedure

Caregivers and infants were recorded during 10 min of free-play over Zoom. Before joining the session, caregivers were asked to identify a place in their home where they could play with their infant alone, gather some toys to play with, and turn off TV, music, or other distractions. Before joining the Zoom chat, caregivers were asked to adjust their camera based on a sample image depicting the recording setup (i.e., both caregiver and infant in view, caregiver facing the camera, and infant seated nearby). Caregivers also completed the MCDI vocabulary questionnaire via WebCDI (deMayo et al., 2021) prior to their appointment. At the start of the session, the experimenter joined the Zoom chat, greeted the caregiver, and explained the procedure (instructing the caregiver to play with their infant just as they normally would). She also helped the caregiver adjust their camera so that both the caregiver and the infant were fully in view and that the camera setup was as similar as possible across dyads (see Figure 1). The experimenter informed the caregiver that she would turn off her camera and audio to minimize distractions and would not be watching or listening as the caregiver and infant played together. After answering any questions from the caregiver, the experimenter turned off her video and audio and minimized and muted the Zoom chat. Approximately 10 min later, the experimenter returned to the Zoom chat to let the caregiver know that the play session was finished and thanked the family for their time. After participating, families were emailed a \$10 Amazon gift card. Ten months after the play session, caregivers were asked to complete a second MCDI vocabulary questionnaire, again via WebCDI, and a total of 39 caregivers (89%) completed this second vocabulary measure. All instructions provided to parents (via email and over Zoom) and other procedural details are available on the OSF



page associated with this research ([https://osf.io/u5avz/?view\\_only=f18f90b5592842b6ae1e38347e610dfe](https://osf.io/u5avz/?view_only=f18f90b5592842b6ae1e38347e610dfe)).

### 2.3 | Coding infant-directed communication

Trained researchers coded all videos for caregivers' use of each of the dimensions of IDC (speech, action, gesture, emotion, and touch) as well as infants' vocalizations and gestures. Videos were coded continuously using ELAN, an open-source software designed for annotating video and audio files (Wittenburg et al., 2006). Researchers coded one dimension at a time, watching the video in its entirety and identifying the onset and offset of each occurrence of speech, action, gesture, emotion, or touch as well as infants' vocalizations and gestures. Our comprehensive coding manual and all related materials are available on the OSF and select session videos are available (for the 21 dyads from which we received caregivers' permission to share) to authorized Databrary users (<https://nyu.databrary.org/volume/1589>). We defined five dimensions of IDC as follows:

1. *Speech*: speech directed from the caregiver to the infant including communicative utterances such as "hmmm" or animal noises. The beginning and ending of each utterance were identified following existing practices for the annotation and transcription of child-directed speech (i.e., using cues such as prosodic contour, turn-taking, and syntactic structure; MacWhinney, 2014; Soderstrom et al., 2021) and only periods between the onset and offset of utterances counted as caregiver speech, ignoring brief pauses that occurred between utterances.
2. *Action*: interacting with an object to change its state, location, or spatial orientation. While previous studies have coded object-directed actions whenever the caregiver's hand was touching an object (e.g., Karmazyn-Raz & Smith, 2023; Schatz et al., 2022; Suanda et al., 2016; Suarez-Rivera et al., 2022), our definition of action additionally required the presence of movement. Simply holding or touching an object was not considered action, but activities like stacking blocks, rolling a truck across the floor, or pretending to feed a doll with a toy spoon would all involve movement and thus were considered actions.
3. *Gesture*: a movement, usually of the body or limbs, that has a communicative intent. Our gesture coding was adapted from Iverson et al. (1999) and included deictic (e.g., pointing), conventional (e.g., shaking the head "no"), representational (e.g., extending and retracting the index finger for "worm"), and emphatic gestures (e.g., bobbing side to side while singing). To differentiate gestures and action, gestures did not involve interactions with objects.
4. *Emotion*: caregiver production of an exaggerated facial expression depicting emotion. These were most frequently the emotions defined by Chong et al. (2003)—joy, surprise, and care/concern—but other exaggerated communicative facial expressions (e.g., frowning the brow in an exaggerated way to indicate confusion) were included as well.

5. *Touch*: non-accidental touches that the caregiver provides on any part of the infants' body (following Abu-Zhaya et al., 2017). We included the use of hands or objects to directly touch the child (e.g., using a stuffed animal to pretend to "kiss" the child on their cheek).

Because each dimension was coded independently, researchers did not have to make decisions about which one of the five dimensions each event depicted; or whether a single event should be classified as one, two, or more dimensions of IDC. That is, a researcher might identify an event as an action when doing the action coding, and the same event as an example of touch when doing the touch coding. For example, a caregiver "driving" a car up and down an infants' leg would be coded as action (moving the car toy to change its state or location) and touch (touching the infants' leg with an object). Additionally, just as we separated utterances in our coding of speech, each individual occurrence of a non-speech behavior (e.g., action, gesture, emotion, and touch) was segmented separately. For example, if a caregiver repeated a pointing gesture in rapid succession, these would be coded as two different gesture events. Similarly, if the caregiver's gesture turned into action (e.g., pointing at a toy car and then picking it up with the same hand), the gesture event would end the moment the caregiver began moving their arm to pick up the car, and an action event would begin (though both gesture *and* action could be occurring at the same time if the caregiver pointed with one hand while simultaneously rolling the car with the other hand). Finally, because only a single caregiver and infant were present during the recording, any speech, action, gesture, emotion, or touch by the caregiver was considered infant-directed (e.g., we did not focus solely on moments that were highly exaggerated and/or repetitive).

In addition to coding the five dimensions of IDC, researchers also coded the occurrence of infant vocalizations and gestures. This was again conducted in separate passes through the video. Infants' vocalizations occurred any time the infant made a communicative vocalization (including babbles, laughs, or squeals, but not vegetative sounds like coughing or burping). As with caregiver gestures, infant gestures were defined as any movement of the infants' body or limbs that had a communicative intent (e.g., pointing or nodding).

Recording via Zoom enabled us to use only one camera angle, which is a limitation of this method of remote data collection, and it was sometimes the case that caregivers or infants were not visible. They occasionally went off-screen, turned away from the camera, or were otherwise blocked from the camera's view. Thus, for each dimension, researchers coded "NC" (not codable) for any segment of time in which it was not possible to see whether the caregiver (for non-speech dimensions of IDC) or the infant (for infant gesture) was or was not exhibiting each behavior. To avoid over-estimating the amount of each behavior present in the interactions, a behavior was deemed "not codable" if there was any question at all about whether the behavior could be properly coded. On average, 9% of seconds in each recording met this criterion. However, this varied across the five dimensions, with more not-codable seconds occurring in the Emotion dimension ( $M = 21\%$ ,  $SD = 18.36\%$ )—as caregivers simply had to turn their head for the emotion dimension to be considered "not codable"—and the least occurring

in the Touch dimension ( $M = 1.14\%$ ,  $SD = 2.51\%$ ; Action:  $M = 10.16\%$ ,  $SD = 10.79\%$ ; Gesture:  $M = 10.48\%$ ,  $SD = 10.23\%$ ; Infant Gesture:  $M = 2.62\%$ ,  $SD = 3.95\%$ ). We found no significant correlations between the proportion of the interaction deemed “not codable” for each dimension and the proportion of the interaction during which each dimension occurred (Action:  $r = -0.03$ ,  $p = 0.858$ ; Gesture:  $r = 0.04$ ,  $p = 0.818$ ; Emotion:  $r = -0.19$ ,  $p = 0.220$ ; Touch:  $r = 0.04$ ,  $p = 0.785$ ; Infant Gesture:  $r = 0.12$ ,  $p = 0.433$ ).

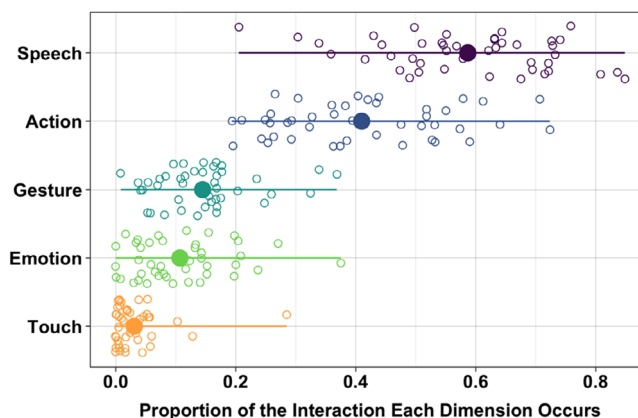
## 2.4 | Assessing reliability

Approximately 20% of the total number of videos ( $n = 9$ ) were fully coded by two independent researchers to assess inter-observer reliability (i.e., whether two independent coders observed the same behavior at the same point during the interaction). We first assessed reliability for our coding of caregivers' use of IDC (speech, action, gesture, emotion, and touch) using both Cohen's kappa (which corrects for agreements expected by random chance; Cohen, 1960) and percent agreement (i.e., the percent of total seconds for which our two coders agreed that an event had or had not occurred). These reliability metrics were computed overall as well as separately for each dimension and each video. Across all five dimensions, we found a Cohen's kappa of 0.88 (% agreement = 95.21%). When examined separately, Cohen's kappa was high, at or above 0.77 (e.g., Bakeman & Quera, 2011; Fleiss, 1981), and percent agreement was at or above 91.49% for all dimensions (Speech:  $\kappa = 0.89$ , % agreement = 94.39%; Action:  $\kappa = 0.84$ , % agreement = 91.49%; Gesture:  $\kappa = 0.79$ , % agreement = 95.41%; Emotion:  $\kappa = 0.81$ , % agreement = 96.87%; Touch:  $\kappa = 0.77$ , % agreement = 98.19%) and all subjects included in the reliability analysis (min:  $\kappa = 0.85$ , % agreement = 93.57%; max:  $\kappa = 0.92$ , % agreement = 96.59%). We also calculated an intraclass correlation (ICC) to assess the extent to which the two coders agreed on the proportion of the interaction during which each dimension occurred and found that this reliability metric was high as well (ICC = 0.99,  $p < 0.001$ ).

This process was repeated for our coding of infant behaviors (i.e., vocalizations and gestures). The overall Cohen's kappa was 0.88 (% agreement = 96.32%). When examined separately, Cohen's kappa was high, at or above 0.70, and percent agreement was at or above 94.19% for both infant vocalizations ( $\kappa = 0.86$ , % agreement = 94.19%) and infant gestures ( $\kappa = 0.70$ , % agreement = 98.71%). Again, we found strong agreement in our coding of the proportion of the interaction during which each dimension occurred (ICC = 0.99,  $p < 0.001$ ).

## 3 | RESULTS

All analyses were performed in RStudio (RStudio Team, 2020) using R Statistical Software (R version 4.1.3 (2022-03-10), R Core Team, 2022). Mixed-effects analyses were performed using the lme4 package (Bates et al., 2015), and  $p$ -values were calculated using the lmerTest package (Kuznetsova et al., 2017). Though it is not straightforward to estimate



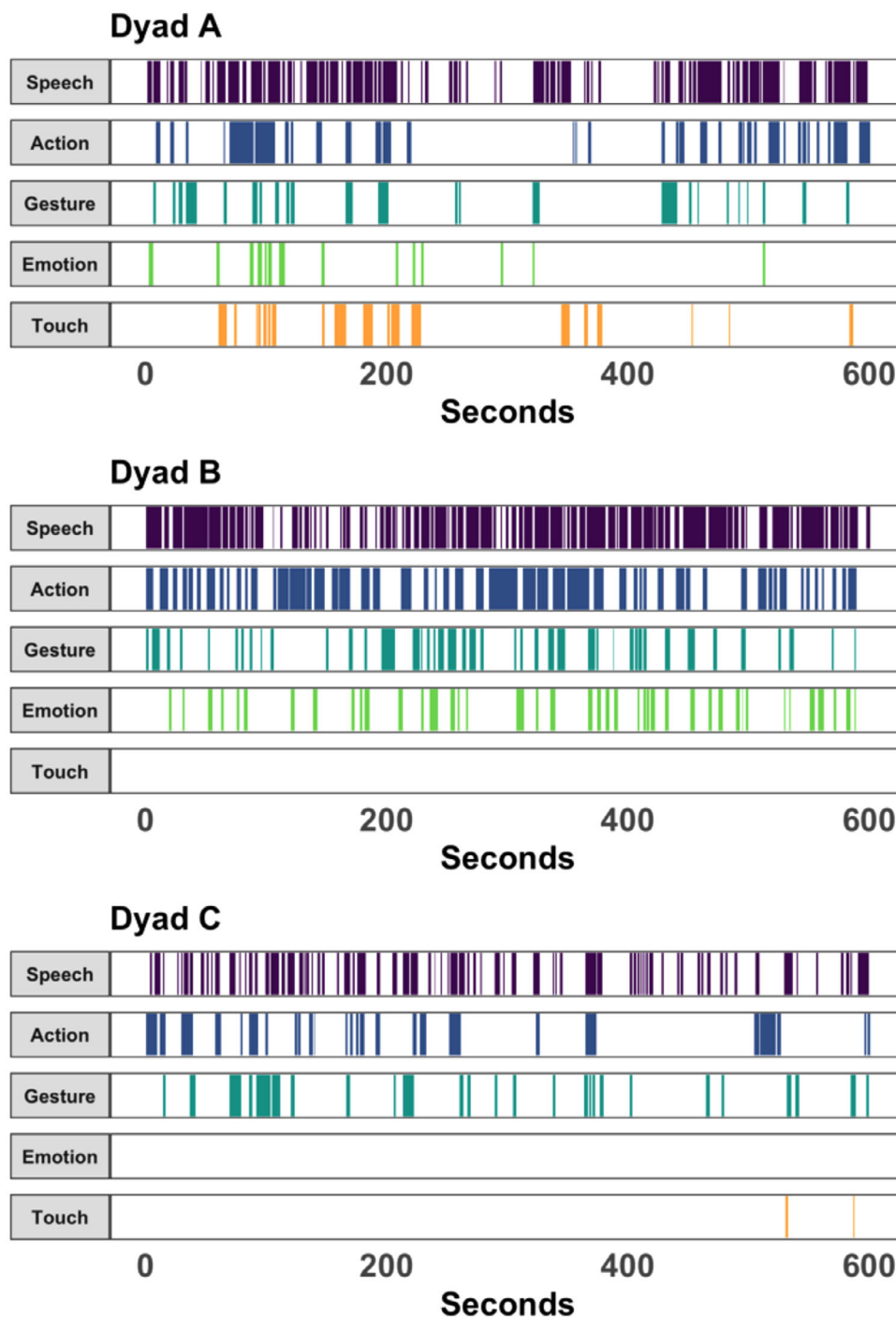
**FIGURE 2** Proportion of the 10-min interactions during which caregivers used each of the five dimensions of IDC. Averages are represented by larger, solid circles, and individual data points are represented by smaller, open circles. The solid lines span the range of scores for each dimension.

standardized effect sizes for mixed-effects models (e.g., Brysbaert & Stevens, 2018), we used the effectsize package (Ben-Shachar et al., 2020) to approximate partial eta-squared and associated confidence intervals for each overall model. Pairwise comparisons and their associated effect sizes (approximated Cohen's  $d$ ) were calculated using the emmeans package (Lenth, 2018), which uses a Tukey adjustment for multiple comparisons (Tukey, 1949).

## 3.1 | Characterizing caregivers' use of IDC

**Variation in caregivers' use of IDC.** The proportion of the interaction during which speech, action, gesture, emotion, and touch occurred varied across caregivers and across dimensions. For example, as indicated by the individual data points in Figure 2, the caregiver who spoke the most did so for 84.80% of the interaction (or 8.48 min) while the caregiver who spoke least did so for 20.50% of the interaction (or 2.05 min). The caregiver who gestured the most did so for 36.83% of the interaction (or 3.68 min) while the caregiver who gestured least did so for 0.83% of the interaction (or 0.08 min). Additionally, a linear mixed-effects model (predicting the proportion of the interaction during which caregivers used each dimension from a fixed effect of dimension and a random intercept for subjects) confirmed that caregivers' use of IDC varied across dimensions,  $F(4, 172) = 223.12$ ,  $p < 0.001$ ,  $\eta^2_p = 0.84$ , 95% CI: [0.81, 1.00] (see Figure 2). While the proportion of the interaction during which caregivers were using emotion and gesture did not differ from one another ( $p = 0.429$ ), differences between all other pairs were significant ( $ps < = 0.007$ ).

Caregivers who used one dimension frequently did not necessarily use all other dimensions frequently, but we did find correlations between the use of some dimensions of IDC. In particular, as caregivers' use of speech increased, so did their use of emotion ( $r = 0.32$ ,  $p = 0.035$ ) and gesture ( $r = 0.54$ ,  $p < 0.001$ ). However, caregivers tended to gesture less if they used more action ( $r = -0.38$ ,  $p = 0.010$ ),



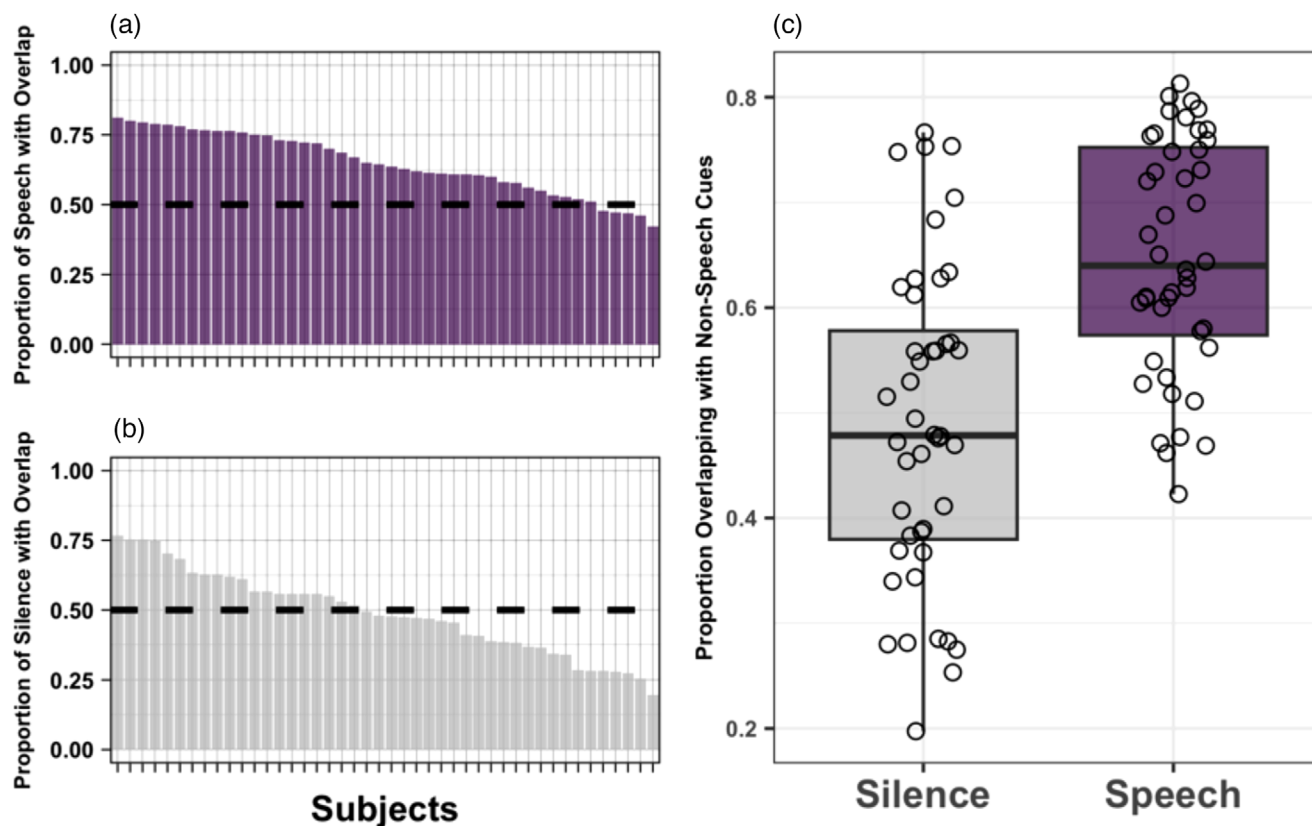
**FIGURE 3** Time-series plots depicting three example caregivers' use of IDC across the 10-min interaction. The x-axis represents seconds as the interaction unfolds and each dimension of IDC is plotted in a different color.

possibly because both gestures and actions frequently (though not always) involve movement of the hands. No other correlations were significant ( $r_s = 0.19$ ,  $p_s = 0.26$ ). See Section S1 for additional details about each pairwise correlation.

*Caregivers use multiple, overlapping dimensions of IDC.* While the previous analyses focused on how much caregivers used each dimension of IDC, this approach does not capture an important feature of infants' everyday experience: that caregiver-infant interactions are dynamic

and unfold across time (see Figure 3). For example, as interactions unfold, caregivers often use multiple multimodal cues simultaneously. In our next analysis, we explored how frequently this occurred. How often does caregivers' use of IDC involve the use of multiple dimensions at the same time?

Underscoring the importance of examining IDC, we found that caregivers' speech overlapped with other cues (e.g., action, gesture, emotion, and/or touch) 65% of the time on average (Figure 4a). However,



**FIGURE 4** Plots A and B show the proportion of speech (a) and silence (b) that overlaps with non-speech dimensions of IDC for each individual caregiver-infant dyad. Plot C combines these two plots to illustrate the average amount of non-speech cues that occur in silence (light gray bar) and with speech (dark purple bar); open circles indicate individual data points.

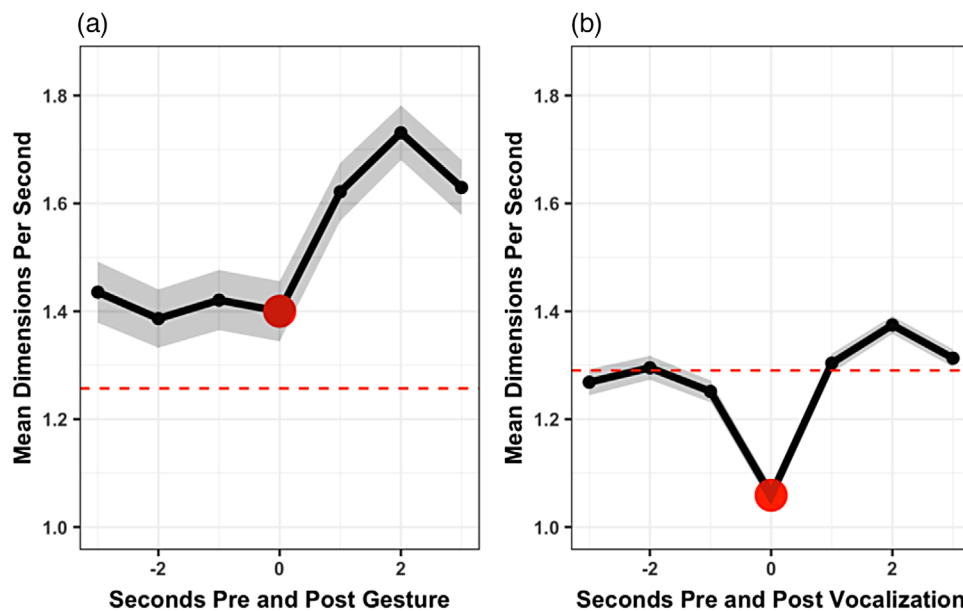
even when caregivers were not speaking, non-verbal communicative cues were used 49% of the time (Figure 4b). To examine whether non-verbal cues occurred more often with speech or in silence, we ran a linear mixed-effects model predicting the frequency of multimodal cues from whether or not speech was also occurring. This model included a random intercept for subjects. Multimodal cues occurred frequently both with speech and also in periods of silence, although multimodal cues accompanied speech ( $M = 64.90\%$ ,  $SD = 11.10\%$ ) significantly more often than they occurred in its absence ( $M = 48.90\%$ ,  $SD = 15.10\%$ ),  $\beta = 0.08$  ( $SE = 0.01$ ),  $p < 0.001$ ,  $\eta^2_p = 0.69$ , 95% CI: [0.56–1.00] (Figure 4c). While prior research has focused predominantly on speech as the driving force of communication (see Perniss (2018) for a review), the present results reveal that rich, multidimensional communication occurs about half the time even in the absence of speech. This means that—even when some might not consider “communication” to be occurring (i.e., in the absence of speech)—it is. Multimodal, multidimensional cues are key features of infants’ everyday communicative interactions.

**Caregivers’ use of multimodal cues is dispersed throughout the interaction.** We next asked whether caregivers’ use of multimodal communicative cues was clustered in time versus spread throughout the interaction. The example time series plots in Figure 3 suggest that multimodal cues occur throughout the recording for all dyads. However, to quantify these visually observed temporal patterns, we used Devia-

tion of Proportion ( $DP$ ), which is a measure of the dispersion of an event (e.g., Gries, 2008).

To generate  $DP$  values, the recording session is first broken into a pre-specified number of parts (in our case, each part was 5 s of the recording). Then, the “observed frequency” is calculated (i.e., how often an event actually occurs in each part of the recording). Next, an “expected frequency” is calculated. This value indicates the expected number of occurrences in each part of the recording if the event was distributed equally throughout. These values are then normalized based on the total number of times the event occurs in the recording and, for each recording part, the absolute value of the difference between the expected and observed frequency is calculated. The resulting  $DP$  statistic is generated by summing these difference values and dividing by two.  $DP$  values range from 0 to 1, with lower values (closer to  $DP = 0$ ) indicating that the event’s occurrence is dispersed relatively equally across the session (e.g., the “expected” and “observed” frequencies do not differ much), and higher values (closer to  $DP = 1$ ) indicating that the event occurs more sporadically (e.g., the “expected” and “observed” frequencies differ substantially).

To calculate the  $DP$  value for each dyad, we split the 10-min recordings into 120 5-s parts<sup>1</sup>. Within each 5-s part, we summed the number of seconds during which two or more communicative behaviors occurred simultaneously (resulting in a value of 1 through 5) and followed the steps above to calculate the dispersion of multimodal cues



**FIGURE 5** Average number of dimensions of IDC used by caregivers before and after infant gestures (a) and vocalizations (b). Gesture or vocalization onset is indicated by the large red circle and shaded regions indicate  $\pm 1$  SE.

across the 10-min session. We found that *DP* values were relatively low, both on average and for all dyads ( $M = 0.37$ ,  $SD = 0.10$ ,  $range = 0.21$ – $0.76$ ; for *DP* values by the dyad, see Figure S2.1). A one-sample *t*-test confirmed that these values were significantly lower than 0.5,  $t(43) = -8.29$ ,  $p < 0.001$ , Cohen's  $d = -1.25$ , 95% CI:  $[-1.64 - -0.85]$ . Because this value is closer to 0 than to 1, it suggests that caregivers' use of multimodal cues was more dispersed throughout the 10-min interaction than clustered in time. For complementary analyses on burstiness (e.g., Abney et al., 2018; Slone et al., 2023), see Section S3.

### 3.2 | Caregivers' use of IDC increases in response to infant gestures and vocalizations

We have so far focused on caregivers' behavior during play. However, the caregiver is not acting alone, and their infant may play an important role in shaping caregivers' use of IDC (e.g., Albert et al., 2018; Elmlinger et al., 2019; Goldstein et al., 2010; Kovács et al., 2014; Lucca & Wilbourn, 2019). In our next set of analyses, we asked whether caregivers increased their use of IDC—specifically, the number of overlapping communicative cues—when infants gestured or vocalized.

We first examined caregivers' use of IDC surrounding infant gestures. On average, infants gestured 7.64 times during the 10-min interaction ( $range = 0$  to 28). We quantified how many overlapping multimodal cues caregivers used across three spans of time: (1) the three-second period before the onset of an infant gesture (pre-gesture region), (2) the three-second period after the onset of an infant gesture (post-gesture region), and (3) outside of these three-second regions (outside gesture region). When these regions overlapped, the segment was counted as post-gesture (though results were similar when the segment was removed entirely or counted as pre-gesture, see

Section S4). We ran a linear mixed-effects model predicting the number of overlapping cues used by caregivers from the gesture region, as defined above. This model also included random intercepts and slopes for subjects. As can be seen in Figure 5a, caregivers used multimodal cues differently across the three regions,  $F(2, 35.12) = 25.85$ ,  $p < 0.001$ ,  $\eta^2_p = 0.60$ , 95% CI:  $[0.41 - 1.00]$ . Pairwise comparisons revealed that caregivers used more multimodal cues in post-gesture regions ( $M = 1.66$ ,  $SD = 0.94$ ) relative to pre-gesture regions ( $M = 1.41$ ,  $SD = 0.96$ ),  $p < 0.001$ , Cohen's  $d = 0.33$ , 95% CI:  $[0.18 - 0.48]$ . Caregivers also used more multimodal cues in post-gesture over non-gesture regions ( $M = 1.40$ ,  $SD = 1$ ),  $p < 0.001$ , Cohen's  $d = 0.43$ , 95% CI:  $[0.32 - 0.55]$ . Pre-gesture and non-gesture regions did not differ from each other,  $p = 0.153$ , Cohen's  $d = 0.11$ , 95% CI:  $[0.01 - 0.22]$ . These findings suggest that caregivers increase the dimensionality of communication by using more overlapping multimodal cues in moments that immediately follow infants' gestures.

Next, we focused on caregivers' responses to infant vocalizations. Infant vocalizations were more frequent than gestures; on average, infants vocalized 62.86 times ( $range = 16$  to 105). As with infant gestures, we asked how many overlapping multimodal cues caregivers were using across three spans of time: (1) the three-second period before the onset of an infant vocalization (pre-vocalization region), (2) the three-second period after the onset of an infant vocalization (post-vocalization region), and (3) outside of these three-second regions (outside vocalization region). As with infant gestures, when these regions overlapped, the segment was counted as post-vocalization (but see Section S4 for alternative classification methods). A linear mixed effects model predicting number of overlapping multimodal cues used by the caregiver from vocalization region (and including random intercepts and slopes for subjects) found a significant effect of vocalization region,  $F(2, 43.43) = 5.39$ ,  $p = 0.008$ ,  $\eta^2_p = 0.20$ , 95% CI:  $[0.04$



–1]. As with gesture, pairwise comparisons revealed that caregivers used more multimodal cues in post-vocalization regions ( $M = 1.33$ ,  $SD = 0.89$ ) than in pre-vocalization regions ( $M = 1.27$ ,  $SD = 0.90$ ),  $p = 0.011$ , Cohen's  $d = 0.08$ , 95% CI: [0.03 –0.13]. We again also found that caregivers used more multimodal cues in post-vocalization compared to non-vocalization regions ( $M = 1.29$ ,  $SD = 0.91$ ),  $p = 0.007$ , Cohen's  $d = 0.11$ , 95% CI: [–0.19 to –0.04]. Pre-vocalization and non-vocalization regions did not differ from each other,  $p = 0.458$ , Cohen's  $d = 0.04$ , 95% CI: [–0.09 –0.02] (see Figure 5b). Thus, like infant gestures, caregivers increased the dimensionality of their communication in response to infant vocalizations. Note that, in contrast to gestures, caregivers seem to use fewer dimensions of IDC at the moment that an infant vocalizes. This may reflect turn-taking in that they are pausing in order to encourage and/or listen to an infant vocalization.

### 3.3 | Relations between caregivers' use of IDC and infants' language skill

We performed two final analyses to investigate how caregivers' use of IDC relates to infants' language, specifically their scores on the MacArthur-Bates Communicative Development Inventory (MCDI; Fenson et al., 2007). The MCDI is a parent-report measure of young children's language development.

In the first analysis, we compared the frequency with which caregivers used each dimension of IDC to their infants' scores on the MCDI at two-time points: The time at which the play session took place, and 10 months later. For each time point, we compared two linear regressions (using the `anova()` function in R): (1) a "speech-only" model predicting infants' vocabulary size from the percentage of the 10-min interaction during which caregivers were speaking, and (2) a "full-idc" model predicting infants' vocabulary size from the percentage of the 10-min interaction during which caregivers were using speech, action, gesture, emotion, and touch (each entered separately). In all models, infant vocabulary size was standardized ( $z$ -scored) within each time point, with 0 indicating the average vocabulary size across participants. This model comparison approach allowed us to ask whether a full model that included all IDC dimensions (i.e., speech, action, gesture, emotion, and touch) improved model fit over a model predicting vocabulary size from caregivers' use of speech alone while, at the same time, enabling us to look at the influence of each dimension separately. These models were constructed as follows:

$$\text{speech-only model: } \text{vocab\_size} \sim \text{speech}$$

$$\text{full-idc model: } \text{vocab\_size} \sim \text{speech} + \text{action} + \text{gesture} + \text{emotion} + \text{touch}$$

Speech alone was not a significant predictor of infants' vocabulary size, neither when vocabulary was assessed at the time of the play session,  $\beta = -0.85$  ( $SE = 1.06$ ),  $p = 0.427$ ,  $\eta^2_p = 0.02$ , 95% CI: [0.00–1.00], nor 10 months later,  $\beta = -1.48$  ( $SE = 1.11$ ),  $p = 0.188$ ,  $\eta^2_p = 0.05$ , 95% CI: [0.00–1.00]. However, the full-idc model (including all five dimensions) significantly improved model fit over the speech-

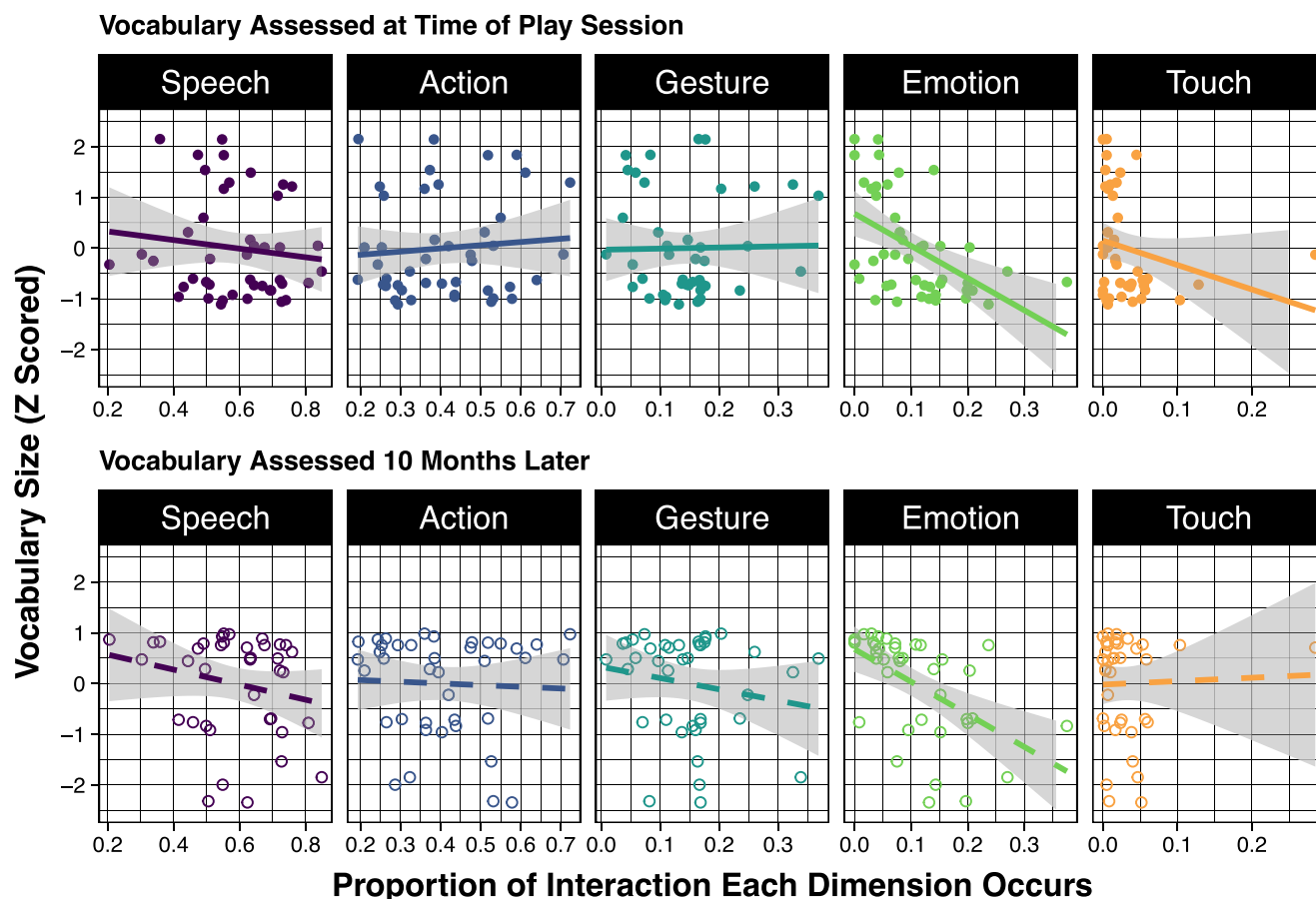
**TABLE 1** Coefficient estimates from the "speech-only" model predicting infants' vocabulary size at the time of the 10-min play session from speech only and the "full-idc" model predicting infants' vocabulary size at the time of the 10-min play session from speech, action, gesture, emotion, and touch.

Estimate	SE	t	p	
<i>Speech-Only model</i>				
Intercept	0.50	0.64	0.78	0.440
Speech	–0.85	1.06	–0.80	0.427
<i>Full-IDC model</i>				
Intercept	0.30	0.72	0.42	0.054
Speech	0.29	1.15	0.25	0.804
Action	0.67	1.01	0.66	0.512
Gesture	1.40	2.16	0.65	0.519
Emotion	–7.10	1.74	–4.08	<0.001
Touch	–6.04	2.80	–2.16	0.037

**TABLE 2** Coefficient estimates from the "speech-only" model predicting infants' vocabulary size 10 months after the play session from speech only and the "full-idc" model predicting infants' vocabulary size 10 months after the play session from speech, action, gesture, emotion, and touch.

Estimate	SE	t	p	
<i>Speech-Only model</i>				
Intercept	0.87	0.67	1.30	0.201
Speech	–1.48	1.11	–1.34	0.188
<i>Full-IDC model</i>				
Intercept	1.24	0.79	1.57	0.125
Speech	0.22	1.33	0.17	0.869
Action	–1.01	1.09	–0.93	0.361
Gesture	–1.82	2.35	–0.78	0.443
Emotion	–6.38	1.84	–3.47	0.001
Touch	–0.53	3.19	–0.17	0.868

only model both for infants' vocabulary at the time of the play session,  $F(4, 38) = 5.10$ ,  $p = 0.002$ , and 10 months later,  $F(4, 33) = 3.26$ ,  $p = 0.023$  (See Tables 1 and 2). This effect was primarily driven by caregivers' use of emotion during the play session, both when vocabulary was measured concurrently ( $\beta = -7.10$  ( $SE = 1.74$ ),  $p < 0.001$ ,  $\eta^2_p = 0.30$ , 95% CI: [0.12–1.00]) as well as 10 months later ( $\beta = -6.38$  ( $SE = 1.84$ ),  $p = 0.001$ ,  $\eta^2_p = 0.27$ , 95% CI: [0.08–1.00]). This effect was also driven by caregivers' use of touch, which was related to vocabulary at the time of the play session, ( $\beta = -6.04$  ( $SE = 2.80$ ),  $p = 0.037$ ,  $\eta^2_p = 0.11$ , 95% CI: [0.00–1.00]), but not 10 months later ( $\beta = -0.53$  ( $SE = 3.19$ ),  $p = 0.868$ ,  $\eta^2_p = 0.001$ , 95% CI: [0.00–1.00]). That is, caregivers generally expressed more emotion and used more touch with infants who had lower vocabulary scores (see Tables 1 and 2 and Figure 6), perhaps in a way that scaffolds their attention and learning (Nencheva & Lew-Williams, 2022).<sup>2</sup>



**FIGURE 6** Relation between the proportion of the interaction during which each dimension occurs and infants' vocabulary size (as measured by the MCDI) at the time of the 10-min interaction (top row, solid lines, and circles) and 10 months later (bottom row, dashed lines, and open circles).

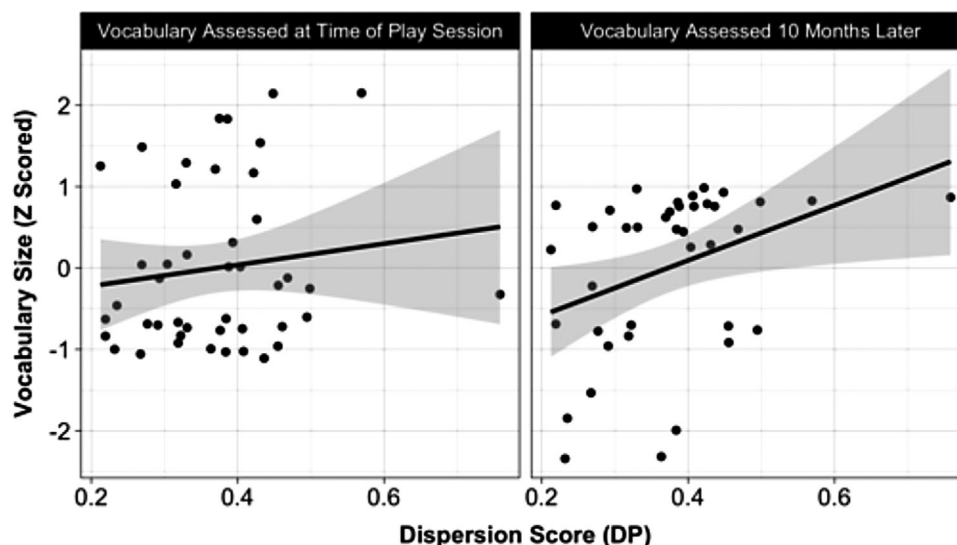
Finally, we asked if the temporal dynamics of caregivers' use of IDC (specifically, the extent to which overlapping communicative cues were clustered vs. spread out over time) was predictive of infants' vocabulary at either time point. We ran a linear regression predicting infants' vocabulary size from the *DP* value characterizing the dispersion of overlapping cues during the interaction with their caregiver. We found that dispersion was not predictive of vocabulary size assessed at the time of the play session,  $\beta = 0.014$  ( $SE = 0.016$ ),  $p = 0.383$ ,  $\eta^2_p = 0.02$ , 95% CI: [0.00–1.00]. However, infants who experienced more clustered input (i.e., had higher *DP* values) had larger vocabulary sizes 10 months after the play session,  $\beta = 0.039$  ( $SE = 0.017$ ),  $p = 0.022$ ,  $\eta^2_p = 0.13$ , 95% CI: [0.01–1.00] (see Figure 7).

#### 4 | DISCUSSION

By examining the dynamics of caregiver-infant interactions, we highlight the importance of characterizing the natural complexity of infants' everyday learning environments, involving the coordination of communicative cues across numerous modalities (e.g., Holler & Levinson, 2019; Perniss, 2018). That is, communication with infants is multidimensional, and focusing on only one or two dimensions of IDC

undercuts the full richness of infants' real-world learning input. Our findings, scrutinizing caregiver-infant play, highlight that the majority of speech that infants hear is accompanied by one or more multimodal cues and, even in the absence of speech, caregivers frequently communicate in parallel using non-verbal means. Furthermore, we found that these cues are used systematically in response to infants' own behaviors to their own level of language proficiency, suggesting that the structure and timing of infant-directed communication may support early learning. These findings call attention to the stark differences between infants' real, everyday learning environments and the ways that learning is typically tested in controlled, often lab-based, studies. While controlled studies provide insights that are crucial for elucidating potential mechanisms of early learning, research characterizing infants' real-world, everyday input is necessary for understanding how these mechanisms are actually instantiated in the environments in which infants are actively engaged.

We also found that IDC varied in multiple ways: across dimensions, between caregiver-infant dyads, and from moment to moment. Across dimensions, we found that caregivers use speech and action during the majority of the interaction, gesture and emotion throughout approximately one-quarter of the interaction, and touch more rarely. However, within individual dimensions, there was remarkable variability across



**FIGURE 7** Relation between dispersion and infants' vocabulary size at the time of the play session (left panel) and 10 months later (right panel).

caregivers and infants. For example, while on average caregivers gestured throughout 15% of the interaction, the frequency of gestures across individual dyads ranged from less than 1% to nearly 37%. This degree of variation is especially notable given that our sample consisted primarily of white, English-speaking, higher-income families in the United States. Thus, rather than focusing on invariance across individuals, we highlight the importance of centering *variation* (as opposed to averages) in the interpretation of our results and in considerations about how to extend the current research.

One prominent source of variation in IDC likely comes from caregivers' tailoring of communicative cues to their infant. Across a variety of domains, it has been demonstrated that caregivers vary their input in response to infants' behaviors and abilities (e.g., Elmlinger et al., 2019; Fukuyama et al., 2015; Gogate et al., 2000; Roy et al., 2009, 2015; Smith & Trainor, 2008). We also observed links between caregivers' use of IDC and infants' vocabulary (as measured by the MCDI). Specifically, caregivers used more facial expressions conveying emotion with infants who had lower vocabulary sizes, and the infants of caregivers whose use of multidimensional cues was more clustered in time (perhaps suggesting a more systematic or tailored use of multimodal behaviors) had higher vocabulary sizes 10 months later. These analyses were limited, though, in that they focused only on the total amount of IDC that parents used during the interaction. More sensitive analyses would consider links between caregivers' use of multiple dimensions of IDC from moment-to-moment as infants encounter familiar and novel objects in their world and as infants' language skills develop across the first years of their lives. For example, when infants are on the cusp of learning a new word, caregivers tend to reduce the length of utterances containing that word (Roy et al., 2009, 2015) and caregivers increase the exaggeration of action sequences that their infants have the fine-motor skills to perform but are not yet performing independently (Fukuyama et al., 2015). Just as we found that caregivers increase the number of communicative cues they are using in response

to infants' gestures and vocalizations, we might additionally predict that caregivers will use IDC systematically during periods in interaction in which infants are interacting with novel objects or encounter a new word. This tailoring of IDC may help create optimal moments for learning. Research is currently underway to better understand how caregivers' use of IDC is fine-tuned to their infants' unique needs and abilities.

However, there are many more reasons why we might expect infants' everyday input to vary in ways that are beyond the scope of the current study. First, this interaction represented only 10 minutes of each infant's life during one specific context (i.e., play with objects). Because context influences infants' input at a variety of levels—from *micro* contexts like the identity of interactions partners or the immediate situational context (e.g., daily routines) to *macro* contexts including political and economic systems in the area in which the child resides (e.g., Rowe & Weisleder, 2020)—it will be important to extend beyond this short snapshot of communicative experience. Additionally, much of the investigation of variation in infants' everyday input across families and communities has focused on speech (e.g., Casillas & Cristia, 2019; Cristia, 2023; Tamis-LeMonda et al., 2019), and minimal work with diverse populations has examined non-speech dimensions of variation in infants' everyday communicative input (but see Abu-Zhaya et al., 2019; Depowski et al., 2015; Gabouer et al., 2020; Tamis-LeMonda et al., 2012). Examining the sources and correlates of variation in infants' everyday experience—across communities and cultures, in a variety of everyday activity contexts, and from moment to moment—can provide important insight into how infants learn in the natural, everyday environments in which they are actually doing this learning (Adolph, 2020; Casillas, 2023; de Barbaro, 2019; de Barbaro & Fausey, 2022; Sullivan et al., 2021).

When making inferences about infants' natural experience, it is also important to consider the strengths and limitations that arise from recording brief 10-min interactions over Zoom. While there is



existing evidence that short play sessions like these mirror peak interactions in longer, more natural contexts (e.g., Tamis-LeMonda et al., 2017) and that there are similarities between in-lab and at-home samples of behavior (e.g., Manning et al., 2020), it is unlikely that these brief interactions are fully representative of infants' natural experience. For example, our analysis of brief periods during which caregivers are aware of being recorded likely overestimates the amount of IDC that occurs as caregivers and infants go about their day. Though coding these behaviors requires substantial time and resources even in 10-min recordings, more densely sampled, longer recordings are needed to assess the true nature of IDC in infants' everyday experience and how it might vary over time.

Another limitation of remote data collection is that we were only able to record from one, immovable camera angle, resulting in moments of the video that were not codable and restricting the interaction features that we were able to reliably detect (e.g., because infants moved around much more than caregivers, their facial expressions would have been extremely challenging to code from these videos). Furthermore, while we have previously highlighted the homogenous demographics of the participants included here, the requirements for participation in remote studies (i.e., reliable internet connection) likely narrowed the characteristics of our sample even more. On the other hand, our use of at-home Zoom recordings enabled us to see families interacting in their everyday environment and with objects that are familiar to them, which is a major strength of this method. Additionally, Zoom-based methods enable asynchronous recording of caregiver-infant interactions (which we have employed in ongoing research) and thus may both reduce the effects of having an experimenter present during the recording session while also enabling families to participate at times during which they would naturally engage in behaviors of interest to researchers (e.g., playing, feeding, or reading). In sum, any method of data collection has both strengths and challenges, and it is important to consider how to best measure behaviors of interest while acknowledging methodological limitations.

Finally, to enable further study of variation in caregivers' use of IDC, it is likely that our coding scheme will need to be continuously examined and revised. For example, as mentioned previously, one decision we made was to consider all behaviors "infant-directed," given that they were performed in a one-on-one interaction between a caregiver and their infant. An alternative approach may be to only consider behaviors that are "stereotypically" infant-directed (e.g., exaggerated, repetitive, etc.), as these may be the cues that are most likely to engage infants' attention and promote learning. However, caregivers vary in how they use infant-directed behaviors and what meets the threshold for an exaggerated action or gesture. Additionally, infants appear to adapt to their caregivers' use of infant-directedness (at least in the domain of speech; Outters et al., 2020), and this may be yet another way that caregivers and infants tune in to each other and adapt their own behaviors over time. A second decision in the current coding scheme was to code speech, action, gesture, emotion, and touch as separate categories. However, it is unlikely that these dimensions are truly separable; a caregiver may say "Wow!" while making an exaggerated expression of surprise and pointing at an object. While our

coding scheme would consider this caregiver to be using three overlapping communicative cues, it is more likely that these cues work together to produce meaning in context. Furthermore, it may be that the absence of a communicative cue is meaningful as well. For example, the meaning of "Wow" can vary depending on whether it is said with an exaggerated smile or with a neutral facial expression. Thus, future work should attempt to reconcile these issues to better represent the cross-caregiver, cross-community dynamics of infants' communicative experience.

Our characterization of infants' natural communicative environment as highly dynamic and multidimensional calls into question whether the learning mechanisms that we see evidence for in tightly controlled experiments truly extend to the everyday interactions in which infants are actually learning and how these learning mechanisms may vary across contexts. Tightly controlled experiments intentionally remove the richness of true caregiver-infant interactions. For example, multimodality is often treated as "noise" and caregivers are usually removed from the equation to the extent possible (e.g., by wearing headphones and/or being prevented from seeing visual stimuli). This is problematic when the results of controlled experiments are taken as evidence for mechanisms of early learning that generalize beyond a single restricted task. Controlled experiments provide important insight into how infants *can* learn, but until we have improved characterizations of "learning in vivo" (Casillas, 2023) and better understand how learning mechanisms are instantiated in the real world, it may be problematic to generalize the results of controlled experiments to how infants actually *do* learn in their real, everyday learning environments.

## 5 | CONCLUSION

Infants' everyday communicative environments are rich and multidimensional. Our investigation of IDC during natural, at-home interactions revealed that caregivers frequently use multiple communicative cues at the same time and that they increase the dimensionality of IDC in response to infants' gestures and vocalizations. This increase in dimensionality may be one way that caregivers implicitly adapt their behaviors to reinforce infants' engagement with their environment and generate optimal moments for learning. Attempts to isolate learning mechanisms by removing the complexity of the environment may be counter-productive to understanding how infants learn in the everyday interactions in which the majority of this learning occurs. Harnessing variability in multidimensional caregiver input provides important new insights into the true richness of infants' everyday social interactions and allows for the development of comprehensive, ecologically valid theories of early learning.

## ACKNOWLEDGMENTS

We thank the participating families, Ella Whitfield, Kennedy Casey, and Charlotte Brace for supporting data collection and "SAGET" coding, and members of the Princeton Baby Lab for valuable discussions and feedback. This work was supported by an NSF SBE Postdoctoral Research Fellowship (2004983) and F32 Postdoctoral Research



Fellowship from the National Institute of Child Health and Human Development (F32HD103439) to Jessica E. Kosie and by a grant from the National Institute of Child Health and Human Development to Casey Lew-Williams (R01HD095912).

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

### DATA AVAILABILITY STATEMENT

Coding materials, data, and analysis code are available on the Open Science Framework (<https://osf.io/u5avz/>) and linked GitHub repository. Select session videos are available, with caregiver permission, to authorized Databrary users (<https://nyu.databrary.org/volume/1589>).

### ORCID

Jessica E. Kosie  <https://orcid.org/0000-0002-2390-0963>

### ENDNOTES

<sup>1</sup>These results did not differ substantially when calculated using 5s, 10s, and 20s parts.

<sup>2</sup>We additionally planned to examine caregivers' use of IDC in relation to infants' sustained attention. However, we were unable to establish high reliability in coding of sustained attention. We thus opted not to include these analyses in the manuscript, though additional information is available in Section S5.

### REFERENCES

- Abney, D. H., Dale, R., Louwerse, M. M., & Kello, C. T. (2018). The bursts and lulls of multimodal interaction: Temporal distributions of behavior reveal differences between verbal and non-verbal communication. *Cognitive Science*, 42(4), 1297–1316. <https://doi.org/10.1111/cogs.12612>
- Abu-Zhaya, R., Kondaurova, M. V., Houston, D., & Seidl, A. (2019). Vocal and tactile input to children who are deaf or hard of hearing. *Journal of Speech, Language, and Hearing Research*, 62(7), 2372–2385. [https://doi.org/10.1044/2019\\_JSLHR-L-18-0185](https://doi.org/10.1044/2019_JSLHR-L-18-0185)
- Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues. *Journal of Child Language*, 44(5), 1088–1116. <https://doi.org/10.1017/S0305000916000416>
- Adolph, K. E. (2020). Ecological validity: Mistaking the lab for real life. In *My biggest research mistake: Adventures and misadventures in psychological research*. SAGE Publications, Inc. <https://doi.org/10.4135/9781071802601>
- Albert, R. R., Schwade, J. A., & Goldstein, M. H. (2018). The social functions of babbling: Acoustic and contextual characteristics that facilitate maternal responsiveness. *Developmental Science*, 21(5), e12641. <https://doi.org/10.1111/desc.12641>
- Anisfeld, E., Casper, V., Nozyce, M., & Cunningham, N. (1990). Does infant carrying promote attachment? An experimental study of the effects of increased physical contact on the development of attachment. *Child Development*, 61(5), 1617–1627.
- Bahrnick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology*, 41(4), 352–363. <https://doi.org/10.1002/dev.10049>
- Bahrnick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36(2), 190–201. <https://doi.org/10.1037/0012-1649.36.2.190>
- Bahrnick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, 13(3), 99–102. <https://doi.org/10.1111/j.0963-7214.2004.00283.x>
- Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press.
- de Barbaro, K. (2019). Automated sensing of daily activity: A new lens into development. *Developmental Psychobiology*, 61(3), 444–464. <https://doi.org/10.1002/dev.21831>
- de Barbaro, K., & Fausey, C. M. (2022). Ten lessons about infants' everyday experiences. *Current Directions in Psychological Science*, 31(1), 28–33. <https://doi.org/10.1177/09637214211059536>
- Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, 21(3), 205–226.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, 12(1), 2.
- Baumwell, L., Tamis-LeMonda, C. S., & Bornstein, M. H. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior and Development*, 20(2), 247–258. [https://doi.org/10.1016/S0163-6383\(97\)90026-6](https://doi.org/10.1016/S0163-6383(97)90026-6)
- Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function: Infant pointing serves an interrogative function. *Developmental Science*, 15(5), 611–617. <https://doi.org/10.1111/j.1467-7687.2012.01160.x>
- Ben-Shachar, M. S., Lüdtke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56), 2815.
- Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, 22(1), e12715. <https://doi.org/10.1111/desc.12715>
- Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning and Development*, 4(3), 179–202. <https://doi.org/10.1080/15475440802143091>
- Bornstein, M. H., Tamis-LeMonda, C. S., Tal, J., Ludemann, P., Toda, S., Rahn, C. W., Pêcheux, M. G., Azuma, H., & Vardi, D. (1992). Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child Development*, 63(4), 808–821.
- Bornstein, M. H., Putnick, D. L., Cote, L. R., Haynes, O. M., & Suwalsky, J. T. D. (2015). Mother-infant contingent vocalizations in 11 countries. *Psychological Science*, 26(8), 1272–1284. <https://doi.org/10.1177/0956797615586796>
- Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for "motionese": Modifications in mothers' infant-directed action: 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>
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), 1–20.
- Casillas, M. (2023). Learning language in vivo. *Child Development Perspectives*, 17(1), 10–17. <https://doi.org/10.1111/cdep.12469>
- Casillas, M., & Cristia, A. (2019). A step-by-step guide to collecting and analyzing long-format speech environment (LFSE) recordings. *Collabra: Psychology*, 5(1), 24. <https://doi.org/10.1525/collabra.209>
- Chong, S. C. F., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development*, 12(3), 211–232. <https://doi.org/10.1002/icd.286>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cooper, R., & Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, 61(5), 1584–1595.

- Cristia, A. (2023). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, 26(1), e13265. <https://doi.org/10.1111/desc.13265>
- Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy*, 23(1), 4–28.
- Depowski, N., Abaya, H., Oghalai, J., & Bortfeld, H. (2015). Modality use in joint attention between hearing parents and deaf children. *Frontiers in Psychology*, 6, 1–8. <https://doi.org/10.3389/fpsyg.2015.01556>
- DeMayo, B., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C. F., Frank, M. C., & Marchman, V. (2021). Web-CDI: A system for online administration of the MacArthur-Bates Communicative Development Inventories. *Language Development Research*, 1(1), 55–98.
- Dimitrova, N., & Moro, C. (2013). Common ground on object use associates with caregivers' gestures. *Infant Behavior and Development*, 36(4), 618–626. <https://doi.org/10.1016/j.infbeh.2013.06.006>
- Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of Child Language*, 46(5), 998–1011. <https://doi.org/10.1017/S0305000919000291>
- Feldman, R., Singer, M., & Zagoory, O. (2010). Touch attenuates infants' physiological reactivity to stress. *Developmental Science*, 13(2), 271–278. <https://doi.org/10.1111/j.1467-7687.2009.00890.x>
- Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-Bates communicative development inventories*. Brookes Publishing Company.
- Ferber, S. G., Feldman, R., & Makhoul, I. R. (2008). The development of maternal touch across the first year of life. *Early Human Development*, 84(6), 363–370. <https://doi.org/10.1016/j.earlhumdev.2007.09.019>
- Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, 66(6), 103–114. <https://doi.org/10.1525/aa.1964.66.suppl.3.02a00060>
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8(2), 181–195. [https://doi.org/10.1016/S0163-6383\(85\)80005-9](https://doi.org/10.1016/S0163-6383(85)80005-9)
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501. <https://doi.org/10.1017/S0305000900010679>
- Fleiss, J. (1981). *Statistical measures for rates and proportions*. Wiley.
- Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, 43(1), 238–252. <https://doi.org/10.1037/0012-1649.43.1.238>
- Franco, F., Fogel, A., Messinger, D. S., & Frazier, C. A. (1996). Cultural differences in physical contact between Hispanic and Anglo mother-infant dyads living in the United States. *Early Development and Parenting*, 5(3), 119–127. [https://doi.org/10.1002/\(SICI\)1099-0917\(199609\)5:3<119::AID-EDP123>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-0917(199609)5:3<119::AID-EDP123>3.0.CO;2-Y)
- Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Developmental Science*, 12(4), 504–509. <https://doi.org/10.1111/j.1467-7687.2008.00794.x>
- Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, 9(1), 1–24. <https://doi.org/10.1080/15475441.2012.707101>
- Fukuyama, H., Qin, S., Kanakogi, Y., Nagai, Y., Asada, M., & Myowa-Yamakoshi, M. (2015). Infant's action skill dynamically modulates parental action demonstration in the dyadic interaction. *Developmental Science*, 18(6), 1006–1013. <https://doi.org/10.1111/desc.12270>
- Gabouer, A., Oghalai, J., & Bortfeld, H. (2020). Parental use of multimodal cues in the initiation of joint attention as a function of child hearing status. *Discourse Processes*, 57(5–6), 491–506. <https://doi.org/10.1080/0163853X.2020.1759022>
- Gabouer, A., Oghalai, J., & Bortfeld, H. (2018). Hearing parents' use of auditory, visual, and tactile cues as a function of child hearing status. *International Journal of Comparative Psychology*, 31, <https://doi.org/10.46867/ijcp.2018.31.02.03>
- Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, 69(2), 133–149. <https://doi.org/10.1006/jecp.1998.2438>
- Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, 71(4), 878–894. <https://doi.org/10.1111/1467-8624.00197>
- Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think*. Harvard University Press.
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, 100(13), 8030–8035. <https://doi.org/10.1073/pnas.1332441100>
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19(5), 515–523. <https://doi.org/10.1111/j.1467-9280.2008.02117.x>
- Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, 15(4), 362–391. <https://doi.org/10.1111/j.1532-7078.2009.00020.x>
- Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, 24(5), 339–344. <https://doi.org/10.1177/0963721415595345>
- 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/inf.12006>
- Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, 6, 1–10. <https://doi.org/10.3389/fpsyg.2015.01167>
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403–437. <https://doi.org/10.1075/ijcl.13.4.02gri>
- Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, 30(6), 509–516. <https://doi.org/10.1177/0165025406071914>
- Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*, 19(4), 385–408. <https://doi.org/10.1111/inf.12054>
- Hertenstein, M. J. (2002). Touch: Its communicative functions in infancy. *Human Development*, 45(2), 70–94. <https://doi.org/10.1159/000048154>
- Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Alemu, E., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., ... Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, 6(11), 1545–1556. <https://doi.org/10.1038/s41562-022-01410-x>
- Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (2015). Editorial: Turn-taking in human communicative interaction. *Frontiers in Psychology*, 6, 1–4. <https://doi.org/10.3389/fpsyg.2015.01919>
- 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>
- Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development*, 14, 57–75.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development*, 66(2), i–149.



- Jean, A. D. L., Stack, D. M., & Fogel, A. (2009). A longitudinal investigation of maternal touching across the first 6 months of life: Age and context effects. *Infant Behavior and Development*, 32(3), 344–349. <https://doi.org/10.1016/j.infbeh.2009.04.005>
- Karmazyn-Raz, H., & Smith, L. B. (2023). Discourse with few words: Coherence statistics, parent-infant actions on objects, and object names. *Language Acquisition*, 30(3–4), 211–229. <https://doi.org/10.1080/10489223.2022.2054342>
- Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, 42(6), 703–735. <https://doi.org/10.1177/01427237211066405>
- Kosie, J. E., & Baldwin, D. (2023). *Pupillometry offers a new lens on how caregivers scaffold infants' learning*. [Preprint]. PsyArXiv. <https://osf.io/zx4ek/>
- Koterba, E. A., & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. *Infant Behavior and Development*, 32(4), 437–444. <https://doi.org/10.1016/j.infbeh.2009.07.003>
- Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy*, 19(6), 543–557. <https://doi.org/10.1111/infa.12060>
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326), 684–686. <https://doi.org/10.1126/science.277.5326.684>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26.
- Lenth, R., & Lenth, M. R. (2018). Package 'lsmmeans'. *The American Statistician*, 34(4), 216–221.
- Levinson, S. C. (2006). On the human "interaction engine". In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality* (1st ed., pp. 39–69). Routledge. <https://www.taylorfrancis.com/books/9781000325423/chapters/10.4324/97810003135517-3>
- Lew-Williams, C., Ferguson, B., Abu-Zhaya, R., & Seidl, A. (2019). Social touch interacts with infants' learning of auditory patterns. *Developmental Cognitive Neuroscience*, 35, 66–74. <https://doi.org/10.1016/j.dcn.2017.09.006>
- Lucca, K., & Wilbourn, M. P. (2019). The what and the how: Information-seeking pointing gestures facilitate learning labels and functions. *Journal of Experimental Child Psychology*, 178, 417–436. <https://doi.org/10.1016/j.jecp.2018.08.003>
- 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>
- MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.
- Manning, B. L., Harpole, A., Harriott, E. M., Postolowicz, K., & Norton, E. S. (2020). Taking language samples home: Feasibility, reliability, and validity of child language samples conducted remotely with video chat versus in-person. *Journal of Speech, Language, and Hearing Research*, 63(12), 3982–3990. [https://doi.org/10.1044/2020\\_JSLHR-20-00202](https://doi.org/10.1044/2020_JSLHR-20-00202)
- ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, 3(1), 24–52. <https://doi.org/10.1177/2515245919900809>
- Messer, D. J. (1978). The integration of mothers' referential speech with joint play. *Child Development*, 49(3), 781–787.
- Meyer, M., Hard, B., Brand, R. J., McGarvey, M., & Baldwin, D. A. (2011). Acoustic packaging: Maternal speech and action synchrony. *IEEE Transactions on Autonomous Mental Development*, 3(2), 154–162. <https://doi.org/10.1109/TAMD.2010.2103941>
- Meyer, M., van Schaik, J. E., Poli, F., & Hunnius, S. (2022). How infant-directed actions enhance infants' attention, learning, and exploration: Evidence from EEG and computational modeling. *Developmental Science*, 26(1), e13259. <https://doi.org/10.1111/desc.13259>
- Morganstern, A. (2024). Gesture and first language development: The multimodal child. In A. Cienki (Ed.), *The Cambridge handbook of gesture studies*. Cambridge University Press.
- Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, 72(3), 718–735. <https://doi.org/10.1111/1467-8624.00311>
- Nencheva, M. L., & Lew-Williams, C. (2022). Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Developmental Review*, 66, 101047. <https://doi.org/10.1016/j.dr.2022.101047>
- Nicely, P., Tamis-LeMonda, C. S., & Bornstein, M. H. (1999). Mothers' attuned responses to infant affect expressivity promote earlier achievement of language milestones. *Infant Behavior and Development*, 22(4), 557–568. [https://doi.org/10.1016/S0163-6383\(00\)00023-0](https://doi.org/10.1016/S0163-6383(00)00023-0)
- O'Neill, M., Bard, K. A., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science*, 8(4), 352–359. <https://doi.org/10.1111/j.1467-7687.2005.00423.x>
- Outters, V., Schreiner, M. S., Behne, T., & Mani, N. (2020). Maternal input and infants' response to infant-directed speech. *Infancy*, 25(4), 478–499. <https://doi.org/10.1111/infa.12334>
- Perniss, P. (2018). Why we should study multimodal language. *Frontiers in Psychology*, 9, 1109. <https://doi.org/10.3389/fpsyg.2018.01109>
- Piazza, E. A., Iordan, M. C., & Lew-Williams, C. (2017). Mothers consistently alter their unique vocal fingerprints when communicating with infants. *Current Biology*, 27(20), 3162–3167.e3. <https://doi.org/10.1016/j.cub.2017.08.074>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, 24(2), 221–234. <https://doi.org/10.1017/S0142716403000110>
- Rowe, M. L. (2000). Pointing and talk by low-income mothers and their 14-month-old children. *First Language*, 20(60), 305–330. <https://doi.org/10.1177/014272370002006005>
- Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, 323(5916), 951–953. <https://doi.org/10.1126/science.1167025>
- 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>
- Rowe, M. L., & Weisleder, A. (2020). Language development in context. *Annual Review of Developmental Psychology*, 2, 201–223.
- Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, 112(41), 12663–12668. <https://doi.org/10.1073/pnas.1419773112>
- Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society*.
- RStudio Team. (2020). RStudio: Integrated development for R. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7–55). Academic Press.
- Schatz, J. L., Suarez-Rivera, C., Kaplan, B. E., & Tamis-LeMonda, C. S. (2022). Infants' object interactions are long and complex during everyday joint engagement. *Developmental Science*, 25(4), e13239. <https://doi.org/10.1111/desc.13239>
- Schmidt, C. L. (1996). Scrutinizing reference: How gesture and speech are coordinated in mother-child interaction. *Journal of Child Language*, 23(2), 279–305.
- Seidl, A., Tincoff, R., Baker, C., & Cristia, A. (2015). Why the body comes first: Effects of experimenter touch on infants' word finding.



- Developmental Science*, 18(1), 155–164. <https://doi.org/10.1111/desc.12182>
- Shatz, M. (1982). On mechanisms of language acquisition: Can features of the communicative environment account for development? In *Language acquisition: The state of the art* (pp. 102–127). Cambridge University Press.
- Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development*, 38(5), 1. <https://doi.org/10.2307/1165783>
- Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy*, 28, 708–737. <https://doi.org/10.1111/inf.12545>
- Slone, L. K., Abney, D. H., Smith, L. B., & Yu, C. (2023). The temporal structure of parent talk to toddlers about objects. *Cognition*, 230, 105266. <https://doi.org/10.1016/j.cognition.2022.105266>
- Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, 13(4), 410–420. <https://doi.org/10.1080/15250000802188719>
- Snow, C. E., & Ferguson, C. A. (1977). *Talking to children*. Cambridge University Press.
- 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>
- Soderstrom, M., Casillas, M., Bergelson, E., Rosemberg, C., Alam, F., Warlaumont, A. S., & Bunce, J. (2021). Developing a cross-cultural annotation system and metacorpus for studying infants' real world language experience. *Collabra: Psychology*, 7(1), 23445. <https://doi.org/10.1525/collabra.23445>
- Stack, D. M., & Arnold, S. L. (1998). Changes in mothers' touch and hand gestures influence infant behavior during face-to-face interchanges. *Infant Behavior and Development*, 21(3), 451–468. [https://doi.org/10.1016/S0163-6383\(98\)90019-4](https://doi.org/10.1016/S0163-6383(98)90019-4)
- Stack, D., & Muir, D. W. (1990). Tactile stimulation as a component of social interchange: New interpretations for the still-face effect. *British Journal of Developmental Psychology*, 8(2), 131–145. <https://doi.org/10.1111/j.2044-835X.1990.tb00828.x>
- Stack, D., & Muir, D. W. (1992). Adult tactile stimulation during face-to-face interactions modulates five-month-olds' affect and attention. *Child Development*, 63(6), 1509–1525.
- Stern, D. N. (1985). *The interpersonal world of the infant*. Basic Books.
- Sullivan, J., Mei, M., Perfors, A., Wojcik, E., & Frank, M. C. (2021). SAYCam: A large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open mind*, 5, 20–29. [https://doi.org/10.1162/opmi\\_a\\_00039](https://doi.org/10.1162/opmi_a_00039)
- Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent-toddler interactions. *Developmental Neuropsychology*, 41(5–8), 324–341. <https://doi.org/10.1080/87565641.2016.1256403>
- Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, 55(1), 96–109. <https://doi.org/10.1037/dev0000628>
- Suarez-Rivera, C., Schatz, J. L., Herzberg, O., & Tamis-LeMonda, C. S. (2022). Joint engagement in the home environment is frequent, multimodal, timely, and structured. *Infancy*, 27(2), 232–254. <https://doi.org/10.1111/inf.12446>
- Tamis-LeMonda, C. S., Adolph, K. E., Lobo, S. A., Karasik, L. B., Ishak, S., & Dimitropoulou, K. A. (2008). When infants take mothers' advice: 18-month-olds integrate perceptual and social information to guide motor action. *Developmental Psychology*, 44(3), 734–746. <https://doi.org/10.1037/0012-1649.44.3.734>
- Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, 72(3), 748–767. <https://doi.org/10.1111/1467-8624.00313>
- Tamis-LeMonda, C. S., Song, L., Leavell, A. S., Kahana-Kalman, R., & Yoshikawa, H. (2012). Ethnic differences in mother–infant language and gestural communications are associated with specific skills in infants. *Developmental Science*, 15(3), 384–397.
- Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, 20(6), e12456. <https://doi.org/10.1111/desc.12456>
- Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development*, 90(6), 2135–2152. <https://doi.org/10.1111/cdev.13089>
- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53–71. [https://doi.org/10.1207/s15327078in0701\\_5](https://doi.org/10.1207/s15327078in0701_5)
- Tincoff, R., Seidl, A., Buckley, L., Wojcik, C., & Cristia, A. (2019). Feeling the way to words: Parents' speech and touch cues highlight word-to-world mappings of body parts. *Language Learning and Development*, 15(2), 103–125. <https://doi.org/10.1080/15475441.2018.1533472>
- Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, 9(2), 335–340. <https://doi.org/10.3758/BF03196290>
- Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99. <https://doi.org/10.2307/3001913>
- Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C., Milán-Maillo, I., & Perniss, P. (2019). *Onomatopoeia, gestures, actions and words: How do caregivers use multimodal cues in their communication to children?* [Preprint]. PsyArXiv. <https://osf.io/v263k>
- 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>
- Williamson, R. A., & Brand, R. J. (2014). Child-directed action promotes 2-year-olds' imitation. *Journal of Experimental Child Psychology*, 118, 119–126. <https://doi.org/10.1016/j.jecp.2013.08.005>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *5th international conference on language resources and evaluation (LREC 2006)* (pp. 1556–1559).
- Wu, Y., Schulz, L. E., Frank, M. C., & Gweon, H. (2021). Emotion as information in early social learning. *Current Directions in Psychological Science*, 30(6), 468–475. <https://doi.org/10.1177/09637214211040779>
- Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244–262. <https://doi.org/10.1016/j.cognition.2012.06.016>
- Zukow, P. (1991). A socio-perceptual/ecological approach to lexical development: Affordances of the communicative context. *Anales de Psicología*, 7(2), 151–162.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kosie, J. E., & Lew-Williams, C. (2024).

Infant-directed communication: Examining the many dimensions of everyday caregiver–infant interactions.

*Developmental Science*, 27, e13515.

<https://doi.org/10.1111/desc.13515>