Joint Engagement in the Home Environment is Frequent, Multi-modal, Timely, and Structured

Catalina Suarez-Rivera[1,2], Jacob L. Schatz[1], Orit Herzberg[1], & Catherine S. Tamis-LeMonda[1]

[1]New York University: Department of Applied Psychology
[2]University College London: Population, Policy and Practice Research and Teaching Department

Author Note
Catalina Suarez-Rivera https://orcid.org/0000-0002-5900-9670
Jacob Schatz https://orcid.org/0000-0002-2192-3619
Orit Herzberg https://orcid.org/0000-0003-3706-2531
Catherine Tamis-LeMonda https://orcid.org/0000-0002-9071-4028

Corresponding Author. Correspondence concerning this article should be addressed to Catherine S. Tamis-LeMonda, catherine.tamis-lemonda@nyu.edu, 246 Greene Street, New York, NY 10003, +1 (212) 998-5399 or Catalina Suarez-Rivera, catalina.suarez-rivera@nyu.edu

Data availability statement. Video examples, coding manuals, raw data, and scripts are shared on Databrary: ***https://nyu.databrary.org/volume/1178***

Acknowledgements. We thank Katelyn Fletcher, Emily Linn, and Melody Xu for collection of data; Shimeng Weng and Emily Linn for behavioral coding; and Karen Adolph, Sandy Gonzalez, Daniel Suh, and Kelsey West for fruitful discussions. This work was funded by National Institute of Child Health and Human Development R01HD094830 and LEGO Foundation grants to Catherine S. Tamis-LeMonda and Karen Adolph. The authors declare no conflicts of interest with regard to the funding source for this study.

Keywords. *Joint engagement; joint attention; multimodal behavior; infant-directed language; parent responsiveness; object play.*

Abstract

Infants develop in a social context, surrounded by knowledgeable caregivers who scaffold learning through shared engagement with objects. However, researchers have typically examined joint engagement in structured tasks, where caregivers sit near infants and display frequent, prompt, and multimodal behaviors around the objects of infant action. Which features of joint engagement generalize to the real-world? Despite the importance of joint engagement for infant learning, critical assumptions around joint engagement in everyday interaction remain unexamined. We investigated behavioral and temporal features of joint engagement in the home environment, where objects for play abound and dyad proximity fluctuates. Infant manual actions, mother manual and verbal behaviors, and dyad proximity were coded frame-by-frame from 2-hour naturalistic recordings of 13-to-23-month-old infants and their mothers (N=38). Infants experienced rich, highly structured, multimodal mother input around the objects of their actions. Specifically, joint engagement occurred within seconds of infant action and was amplified in the context of interpersonal proximity. Findings validate lab-based research on characteristics of joint engagement while highlighting unique properties around the role of mother-infant proximity and temporal structuring of caregiver input over extended time frames. Implications for the social contexts that support infant learning and development are discussed.

Keywords: Joint engagement; joint attention; multimodal behavior; parent responsiveness; object play.

Humans are social beings. Even before infants walk or talk, they share their interest in objects and events with caregivers—who look at, touch, and often comment on the objects of infant attention. And in the second year, object sharing increases in frequency and complexity, allowing infants to experience the world in new ways. As such, caregivers are key contributors to infant learning and development. Infants learn skills and competencies through everyday interactions with experienced adults, gradually internalizing initially external knowledge (Vygotsky, 1978). Thus, infants are not solitary scientists: They learn through interactions with others.

Joint engagement, sometimes referred to as "triadic engagement" (Little, Carver, & Legare, 2016), is the quintessential context or "hotspot" for infant learning about the world—in which caregivers look at, touch, and talk about the objects of infant action (Tomasello, 1995). The coordination of attention with a social partner forms the basis for intersubjectivity, infant understanding of intentions and goals (Carpendale & Lewis, 2004; Moore & Dunham, 2014), and word learning (Bakeman & Adamson, 1984; Tomasello & Farrar, 1986). Infants' shared engagement with mature partners during early social interactions (Rossmanith et al., 2014) has real-time consequences. Joint engagement extends infant attention to objects of shared focus (Yu & Smith, 2016), which in turn supports learning (Richards & Casey, 1992) and vocabulary growth over time (Yu et al., 2019).

Yet despite the importance of joint engagement for learning in-the-moment and over development, the characteristics of joint engagement in ecologically valid settings remain unexamined. Joint engagement is typically observed during structured tasks in laboratory settings, with a caregiver seated next to or across from the infant, interacting with novel toys and objects. Findings from structured interactions confirm the importance of joint engagement for learning by revealing how caregiver touch and speech elicit infant attention and scaffold in-the-moment behaviors (Suarez-Rivera et al., 2019; Tomasello & Farrar, 1986; Yu & Smith, 2016). However, tabletop settings create dyad proximity by design, may heighten social desirability, and just about guarantee that caregivers will frequently look at, talk about, and touch the objects of infant attention.

Here we offer a unique lens on the temporal and behavioral features of joint engagement during everyday activities in the home environment. We extend research to the real-world to test critical assumptions about everyday joint engagement outside lab-based contexts. Through detailed micro-behavioral coding of mother-infant behaviors, we illuminate the natural ebb-and-flow of joint engagement over extended time frames, asking how infants' manual actions and proximity to caregiver function to establish opportunities ripe for manual and verbal input.

Joint Engagement in Structured Tasks

Structured tasks have led to a deep understanding of how infants and caregivers establish bouts of joint engagement and share interests through gaze, touch, gesture, and vocalizations. Indeed, research converges on three take-home messages: joint engagement is multimodal, frequent, and promptly occurs within seconds of infant behavior. Collectively, these features of joint engagement form an important context for in-the-moment learning.

Lab-based Joint Engagement is Multimodal

Initial studies of joint engagement focused on moments when caregivers and infants jointly looked at the same object. However, joint engagement involves behaviors beyond gaze. When infants visually attend to objects, caregivers signal their interest through multiple modalities. Manual behaviors, such as touch and gesture, act as salient cues that guide social partners' eyes to a focal point, often the shared object or activity (Deák et al., 2014; 2018; Yu & Smith, 2013, 2017b). Caregivers' language refers to the object of infant action (Tamis-LeMonda et al., 2013) and serves to maintain infants' visual engagement with objects (Baldwin & Markman, 1989; Belsky et al., 1980). And together, caregivers' manual and verbal behaviors present infants with rich, temporally coordinated, multimodal inputs. In one study, half (51%) of joint visual attention moments involved multimodal input (i.e., manual contact and vocalizations by parent) (Suarez-Rivera et al., 2019).

Lab-based Joint Engagement is Frequent and Timely

Years of fruitful lab-based research suggest that infant-caregiver dyads jointly engage around objects relatively frequently. Caregivers and infants coordinate eye-gaze toward objects approximately 40% of the time (42% in Yu & Smith, 2013; 34% in Yu & Smith, 2017a) at a rate of 7-8 coordinated looks per minute (Yu & Smith, 2013, 2017a). Approximately 65% of infant object looks involve parent eye-gaze; 42% involve parent touch; and 46% involve parent talk (Suarez-Rivera et al., 2019). These rates of joint engagement in lab contexts (i.e., 40-65%) can be interpreted as "frequent" because caregiver engagement for 100% of infant behavior would be unrealistic and not necessarily desirable for infants' social and emotional development (Isabella & Belsky, 1991; Ispa et al., 2004).

Findings also suggest that parent and infant behaviors are tightly coupled in time, following one another within fractions of a second (Yu & Smith, 2013), with 76% of joint visual attention moments having latencies within 1 second (s) (Suarez-Rivera et al., 2019). Furthermore, caregivers' talk is more rapid than caregivers' manual action, which tends to precede infant attention for up to several seconds. Thus, whereas 67% of utterances occur within 1 s of the infant look onset, 48% of manual actions fall within 1 s. Importantly, timeliness matters. The contingency of caregiver behavior following infant behavior supports infants' connecting words to the objects of their actions (Tamis-LeMonda et al., 2014).

Joint Engagement in the Home Environment

What does joint engagement look like in everyday life? Virtually nothing is known. A handful of researchers have observed mother-infant interactions in the home environment (e.g., Deák et al., 2018; Tamis-LeMonda et al., 2017; de Barbaro et al., 2016; Rossmanith, et al., 2014) during relatively brief unstructured tasks in samples of 10-40 dyads. However, these studies did not specify the features of mother-infant shared object interactions over extended periods. One study compared 30-minute-long interactions in the home and laboratory using behavioral coding every 15 s. Mothers engaged more frequently with their infants in the lab than at home, whereas infant behavior remained the same (Belsky, 1980). Naturalistic research, through the application of micro-behavioral coding, is needed to quantify the frequency, latencies, forms, and structure of joint engagement in ecologically valid contexts for interactions that last more than a few minutes.

Moreover, investigation of joint engagement in ecologically valid natural settings is fundamental to science—offering a test of critical theories around the social context of learning and development. Lab-created phenomena could potentially cascade into theories that do not translate to the real-world (Rothwell, 2005; Schmuckler, 2001). For example, infants' perseverating errors in the A-not-B task led to the conclusion that young infants believe objects cease to exist when out of sight. Yet infants' errors are purely task-dependent rather than due to a lack of object permanence (Smith & Thelen, 2003). Likewise, structured lab-based interactions do not always mirror everyday interactions (Dahl, 2017; Gardner, 2000). Mothers direct fewer utterances to their infants at home than in structured-tasks as mothers' language input peaks and drops throughout the day (Bergelson et al., 2019; Tamis-LeMonda et al., 2017). Thus, if joint engagement indeed plays a pivotal role in infant learning and development, its features must be examined in environments natural to infants.

Do features of joint engagement identified in structured tasks generalize to home settings? Frequent and timely bouts of joint engagement may be unlikely at home where infants transition from object to object quite quickly (Herzberg et al., 2021). Furthermore, dyad proximity likely fluctuates as mothers and infants move in and out of reach, creating challenges to bouts of joint touch. Perhaps then, distal modes of engagement (i.e., language and gesture) are primary means of joint engagement at home. Or, the changing proximity of bodies over time may determine how and when caregivers join in with infants (Yamamoto et al., 2019). The study of joint engagement in the "real" world has implications for theories around scaffolding of learning (Bruner, 1974), by illuminating the characteristics of shared interactions around objects that infants regularly experience.

Current study

We build on previous research on joint engagement by investigating mothers' spontaneous inputs around shared objects in the home environment. We defined joint engagement as mothers' manual and/or verbal behaviors directed toward the object(s) of infant manual action. We videorecorded dyads at home for 2 hours, a time period that yielded between 8 and 24 times the amount of data common in 5- to 15-minute structured tasks. Infants' spontaneous manual interactions with objects (i.e., object bouts) provided a starting point for examining mothers' behaviors—manual touch, gestures, vocalizations, and whole-body proximity—relative to the objects in infants' hands. We assessed which object-related behaviors mothers displayed, the frequency of those behaviors, mothers' latency to respond, and the ebb-and-flow of joint engagement over time (i.e., time between bouts of joint engagement).

Specifically, we asked: (1) As infants touch objects, how often and how quickly do mothers touch or gesture toward those objects? (2) How often and how quickly do mothers talk about the objects of infant manual action? (3) How often do mothers coordinate their manual actions and object-related language (i.e., multimodal input) across the two hours? (4) How does proximity relate to the rates and latencies of mothers' manual action and verbal engagement? We also tested for age differences in infant object bouts, joint engagement (rates and latencies), and proximity.

Because lab and home environments differ considerably, we expected infants and mothers to jointly engage around objects less often at home than typically seen in laboratory studies. In particular, we expected mother touch—which depends on proximity to infant—to be a relatively infrequent modality of joining infant object interactions compared to talk and gesture. Therefore, we expected multimodal joint engagement to occur sporadically at rates lower than those documented in structured tasks. Relatedly, mother latencies at home were expected to be longer than the 2-second rapid responses seen in the lab. Furthermore, we expected proximity to align with greater rates and faster latencies of joint engagement relative to when mothers were out of reach. Perhaps, during moments of proximity, mothers' behaviors begin to converge on characteristics of joint engagement in structured tasks. Finally, we did not expect age-related changes in the duration of infant object bouts (Herzberg et al, 2021) and in the rates, forms, and latencies of joint engagement (Yu & Smith, 2017a). However, we expected base rates of proximity to change with infant age, with older infants being near their mothers less than younger infants.

Method

Participants

Thirty-eight mother-infant dyads (20 girls) were video-recorded at home for 2 hours. We recruited locomoting infants at different points in the second year, a critical period for infants' sharing of objects and attention with social partners (Karasik et al., 2011; Yu & Smith, 2017a). The convenience-based sample included 13 13-month-olds (*M* = 13.01, *SD* = 0.18), 13 18-month-olds (*M* = 18.01, *SD* = 0.19), and 12 23-month-olds (*M* = 22.95, *SD* = 0.14). Families were recruited in a large metropolitan city through hospitals, referrals, and brochures. Mothers ranged from 28 to 49 years of age (*M* = 35.2*, SD* = 4.8)*;* most (97%) had earned college or higher degrees; 68% worked part- or full-time; 90% were White. All infants were healthy, born at term, and firstborn. Participants were compensated with a $75-dollar gift card for the visit. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each infant before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board (Protocol IRB-FY2016-825) at New York University and titled Motor Development in Infants, Children, and Adults.

The law of large numbers (Dekking et al., 2005) indicates that relatively small samples are appropriate if analyses operate on estimates calculated from dense sampling of each participant in which data are hierarchical (e.g., object interactions nested within infants). Our infant-level estimates (N=38) were obtained from dense sampling of infant-mother dyads (at a rate of 30 frames per s for 2 hours per dyad, thus 8,208,000 frames across dyads); and bout-level estimates were computed by aggregating 7,454 object bouts across all infant-mother dyads. Thus, dense sampling of thirty-eight families for 2 hours each and the robustness in estimates provided the required power to describe characteristics of joint engagement. Moreover, the sample size of thirty-eight dyads surpassed the projected minimum sample size of 24 determined by statistical power analysis to achieve a large effect size (GPower 3.1).

Procedure

A female researcher visited infants and mothers on a weekday and recorded each dyad's typical activities for 2 hours (*M* = 118.83 min, *SD* = 3.97). Mothers were asked to go about their day as if the researcher was not present and to remain inside the home. The researcher used a handheld digital video camera (30 fps) with external microphone to record infant and mother behaviors with minimal interference. When infant and mother separated, the experimenter followed the infant and captured the mother when she came back into view.

All visits were conducted during daytime hours (between 8:30 a.m. and 5:30 p.m.), but the priority to scheduling depended on mothers' and infants' routines (with mothers estimating the time that babies would be most alert and up from a nap). However, analyses confirmed that time of day in which the recording started did not relate to variables of interest and did not improve model fit.

Coding

Experienced researchers (unaware of the specific research questions) coded mother and infant behaviors in Datavyu (datavyu.org), a computerized coding tool that time locks behaviors to video frames. Reliabilities were assessed by comparing a primary coder's data to a secondary coder's data on a randomly selected 25% of each video. Cohen Kappa's averaged 0.77 to 0.96 across variables.

Infant Object Bouts

Coders marked the onset and offset of infant object interactions. An object interaction occurred when the infant manually moved the object(s) or a part of an object (e.g., door to cabinet) in space. The onset of each bout corresponded to the frame when the infant began moving an object(s) and the offset was marked by at least 3 s off object(s). An object bout could involve multiple objects simultaneously (e.g., the infant held a sippy cup with one hand and played with a car with the other) or multiple objects in sequence if the transition between objects lasted fewer than 3 s (e.g., infant touched a crayon and transitioned to another crayon within 3 s). Bouts that involved interactions with food were excluded.

Mother Manual Actions during Infant Object Bouts

The onsets of mother's first touch and gesture were coded for each infant object bout (i.e., if and when either occurred). Mother touch was coded if the mother's hand made contact with the same or related object touched by the infant (e.g., mother handed a crayon to her infant or picked up another crayon to color with the infant). Mother gesture to the object(s) of infant touch was coded if mother pointed with her index finger, requested the object(s) with an open palm, tapped on the object(s), or displayed an iconic gesture related to the object(s) (e.g., extending arms to mimic a plane as the infant played with a toy airplane). Notably, gestures could not be coded when the mother was out of camera view.

Mother Object-related Language during Infant Object Bouts

Mothers' utterances that were related to the object(s) of infant touch were coded by viewing video-recordings in conjunction with transcriptions created in Datavyu (following conventions of the Codes for the Human Analysis of Transcripts, CHAT). Object-related utterances could describe or encourage infants' actions with the object(s) (e.g., "you are building a tower" or "good job!"), name or describe the object(s) of infant touch (e.g., "yellow spoon"), or suggest an action with the object(s) (e.g., "turn the page").

Mother-infant Physical Proximity

Physical proximity was coded continuously across the 2 hours, defined as infant and mother being arms distance from each other or the object(s) of infant touch. For example, proximity was coded if an infant played with blocks, and mother could touch the infant or a block of the set that the infant could also reach. Proximity was not coded when mother was standing next to her infant who was sitting on the floor because the infant was not within arms' reach unless the mother bent down. Mothers did not need to be interacting to be proximal (e.g., a mother working on her laptop on the couch next to her playing infant was coded proximal); similarly, a mother could be jointly engaged around an object of infant touch when not proximal, as when she spoke about the object from across the room.

Latencies

Latencies between mothers' behavior onset (i.e., touch/gesture/utterance) relative to infant touch were calculated by subtracting the onset of infant touch from the onset of the mother behavior. Mothers' behaviors could temporally *follow* an infant touch onset when the mother engaged with the object(s) of an ongoing infant action, or temporally *precede* an infant touch when the infant touched the object after mother's ongoing action (such as if mothers "initiated" an infant object bout by handing an object to the infant). Thus, latencies of 0.1 s or greater represented times when infants' touch preceded mothers' behavior (mother follow) and latencies of 0 s indicated times in which infants' touch followed mothers' behavior (mother initiate)[1].

Results

We report the frequencies and timing of mothers' manual actions (touch and gesture); language input; and multimodal joint engagement relative to infant touch, followed by examination of mothers' physical proximity in relation to mother manual and verbal behaviors. Analyses were not formally pre-registered but all raw data and materials (i.e., coding manuals, video excerpt examples, Datavyu, SPSS, and R scripts) are available on Databrary, https://nyu.databrary.org/volume/1178.

For analyses, infant-level means were compared using appropriate linear methods (t-tests or ANOVAs) that met normality assumptions. Generalized linear mixed models tested bout-level associations such as whether the duration of the infant bout predicted the likelihood of mother joining the bout or the latency to join the bout. Mixed effects models accounted for nesting of object bouts within infants with random intercepts for infants and object bouts. All mixed models met assumptions, and included random intercepts for participants and object bouts, and infant age and sex as fixed-effects predictors. Consistent with previous research (Yu & Smith, 2017a), infant age and sex did not relate to any measures across analyses (all p's >.155). Likewise, no variables differed by infant age (all p's >.179).

To preview results, even when observing dyads in the natural home setting, joint engagement was frequent, timely, multimodal, and structured. Mothers jointly engaged on nearly half of infant object bouts (47% with manual input, 46% with verbal input; Figure 1A-B) and did so quickly (Figure 3A-C), particularly when they were proximal to their infants (Figure 5). Infants were more likely to experience multimodal than unimodal input (Figure 4B). Finally, joint engagement was structured over time *within bouts* (i.e., mother touch tended to precede infant touch, and mothers then followed infant touch with talk and gestures; Figure 2) and *between bouts* (i.e., joint engagement was characterized by a "bursty" pattern in which multimodal bouts were distributed over time and separated by brief periods of no input, Figure 4A and 4C).

Infant Object Bouts

Infant touch was our starting point for understanding how and when mothers engage with their infants around objects. Infants spent 58% of the visit interacting with objects (*SD* = 12), distributed across 196.24 object bouts (*SD* = 55.81, *Range* = 104-316), yielding 99.16 bouts per hour for each infant (*SD* = 28.32, *Range* = 51.92-157.95). The duration of infant object bouts varied greatly, ranging from fractions of seconds to as long as nineteen minutes (*Mdn* = 7.24 s, *M* = 20.96 s, *Range* = 0.03 s – 19 min).

Mothers Frequently Touch the Objects that Infants Touch

How often do mothers coordinate their manual actions with infant touch at home? Counter to expectations, mothers often manually interacted with the same object(s) as infants—on nearly half of infant object interactions (*M* = 47%, *SD* = 12%, *Range* = 19-85%) by touching (36%), gesturing (2%), or combining the two (9%). As a result, infants experienced 45.62 bouts of *joint action* (with touch and/or gesture) per hour (*SD* = 18.49).

Perhaps the high co-occurrence of joint touch/gesture occurred for infant bouts that lasted longer and gave mothers more time to engage. Therefore, we expected mothers to be unlikely to show joint manual engagement during short infant object bouts (e.g., under six seconds). Figure 1A shows the percentage of infant bouts of varying durations that involved mother joint touch/gesture. Surprisingly, mothers still touched/gestured on approximately 36% of infant bouts that lasted fewer than 6 s. In fact, the percentage of infant bouts with mother touch in the home was surprisingly similar to the 42% of bouts seen in lab-based studies (Suarez-Rivera et al., 2019; Yu & Smith, 2013).

Although mothers displayed joint touch on roughly 1/3 of brief infant object bouts, the likelihood of joint engagement with touch or gesture increased for longer bouts. Results indicated that the duration of the object bout predicted the likelihood of mother engaging with the object through manual action. Every additional 10 s in the duration of infant object bouts represented a 14% increase in the odds of mother manually engaging with the same or related objects ($e^{0.133}$ = 1.14*odds, p<.001). Thus, infants' sustained interactions with objects offered mothers greater opportunity (i.e., time) to join in if they were not already participating.

Mothers' Touch Precedes Infant Touch

Mothers tended to initiate bouts of joint touch (i.e., having latencies of 0 s as infant touch followed mother touch). Figure 2 shows that mother-initiated joint touch bouts (white fill) were more frequent than joint touch bouts in which mothers followed infant touch (with 0.1 s latencies or greater, pattern fill). Specifically, 60% of object bouts (*M* = 52.61 bouts, *SD* = 32.31) with joint touch were mother-initiated, whereas mothers followed in on the remaining 40% of joint touch object bouts (*M* = 35.39 bouts, *SD* = 13.38), $t(37)$ = 3.21, p=.003, 95% CI [6.34, 28.07], *d* = .52.

Mothers quickly followed infant touch with touch. The frequency histogram of mother's touch latencies that followed infant touch (i.e., latencies 0.1 s or greater) is shown in Figure 3A. The median latency to mother touch following infant touch was 5.54 s (*SD* = 24.65), with 11% of mothers' follow-in touches falling within 1 s, and 66% falling within 10 s. In fact, mothers joined quickly even when infant object bouts were long (i.e., over 30 s, with 22% of touches occurring within 3 s). Of course, because by definition mothers have more time to follow-in with longer latencies for longer infant bouts, an increase of 10 s in the duration of the object bout decreased the odds that the mother touched the object(s) with quick latencies (i.e., under 3 s) by 5% ($e^{-0.049}$ = 0.95*odds, p<.001) as indicated in a logistic mixed regression. Although mothers jointly touched the object(s) of infant interaction within a few seconds, the median latency to mother's touch in the home environment exceeded the latency seen in lab studies (where mothers touch within 0-2 s of infant touch—Suarez-Rivera et al., 2019; Yu & Smith, 2013), as hypothesized.

Mothers Follow in with Gestures

Whereas mother touch tended to precede infant object interactions, mother gesture tended to follow infant touch. When mothers gestured to the objects of infant touch, they followed in for 93% of those bouts (M = 19.32 bouts, SD = 8.83) and initiated the remaining 7% of bouts with gesture (M = 1.37 bouts, SD = 1.50), $t(37) = -13.33$, $p<.001$; 95% CI [-20.67, -15.22]; d = 2.16. Figure 2 shows the mean number of infant object bouts with mother gesture that were mother-initiated (0 s latency- white fill) or mother-follow (0.1 s latency and greater- pattern fill).

Mothers' gestures tended to follow infants' touch likely because they largely co-occurred with (and temporally followed) mothers' touch. Mothers were over 4 times more likely to gesture and touch than to solely gesture during an infant object bout. As a result, latencies to mother's follow-in gestures were greater than latencies to mothers' follow-in touches. The frequency histogram of mother gesture latencies to follow infant touch (with 0.1 s latencies and greater) are shown in Figure 3B. The median latency for mothers to follow with gesture was 9.84 s (SD = 51.04); 7% fell within 1 s, and 50% fell within 10 s. Counter to our hypothesis, mother touch was quicker than mother gesture in the home environment.

Mothers Frequently Talk about the Objects of Infant Touch

Mothers often spoke about the objects of infant touch. Mothers produced *object-related language* for nearly half (M = 46%, SD = 12%) of infant object interactions. Thus, infants experienced language input that aligned in real-time with the object(s) of their manual actions on 44.82 object bouts per hour (SD = 16.49). Furthermore, infant object bouts typically elicited multiple utterances. On average, mothers directed 2.54 object relevant utterances (SD = 1.44, *Range* = 0.62-6.25) per infant object bout. Object-related language that *named* the object(s) of infant action (i.e., the word "cup" when the infant played with a cup) occurred on half of infant object bouts that elicited object-related language, yielding 24.86 instances of named objects for the targets of infant touch per hour (SD = 11.04).

However, bout duration again related to the likelihood of mothers' talking about the object(s) of infant manipulation, as shown by the high percentages of long infant bouts with mother relevant language (Figure 1B). Logistic mixed regression confirmed that mothers were more likely to produce object-related utterances during long infant object bouts. Specifically, every additional 10 s in the duration of infant object bouts represented a 33% increase in the odds that the mother produced object-related language ($e^{0.283}$ = 1.33*odds, $p<.001$). However, longer object bouts did not elicit greater *density* in mother utterances. That is, the normalized number of utterances per second of object interaction with mother talk *decreased* as object bouts were longer. Every additional 10 s to the duration of bouts with mother utterances resulted in a 5.73% decrease in the rate of mother utterance according to a linear mixed model (b=-0.059, SE=0.002, $p<.001$). Thus, infant object play tends to elicit mother object-related language; long bouts enhance likelihood of eliciting maternal speech. However, long infant bouts involved lower rates of language per second, perhaps because once mothers referenced the objects of infant touch a few times, they stopped commenting. As was the case for mothers' manual actions, percentages of bouts with mother language in the home were surprisingly similar to the ~46% of bouts seen in lab-based studies (Suarez-Rivera et al., 2019).

Mothers Follow in with Object-related Language, and do so Quickly

Mothers were more likely to talk about the objects of infant touch after infant touch (0.1 s latencies and greater, pattern fill in Figure 2) than to talk prior to the infant object bout (0 s latencies, white fill). Specifically, mother utterances followed in on 92% of object bouts with object-related language (M = 82.08, SD = 32.19) and initiated the remaining 8% (M = 7.55, SD = 7.14), $t(37) = -13.94$, $p<.001$; 95% CI [-85.36, -63.69]; d = 2.26. Unlike the finding that mother touch was likely to precede an infant touch, mother object-related language was likely to *follow* infant touch.

Mothers' object-relevant language that followed infant touch occurred in less than 3 s (Mdn = 2.45, SD = 14.64), with 24% of latencies being under 1 s, and 84% under 10 s. Figure 3C presents frequencies of mothers' language latencies to follow infant touch (those with 0.1 s latencies and greater). Thus, language could be classified as a rapid, follow-in behavior in line with studies on the contingently responsive nature of language input (e.g., Tamis-LeMonda et al., 2014). As hypothesized, mother language was much more rapid than was mother touch and in fact, latencies of mother language closely approximated latencies in the laboratory (Suarez-Rivera et al., 2019). However, because by definition mothers have more time to follow-in with longer latencies for longer infant bouts, an increase of 10 s in the duration of the bout represented a 5.82% decrease in the odds that the mother produced object-related language quickly (i.e., under 3 s) ($e^{-0.060}$ = 0.94*odds, $p<.001$) as indicated in a logistic mixed regression.

Multimodal Joint Engagement is Frequent and Structured in Time

How often did mothers coordinate manual and verbal inputs toward the object(s) of infant action? Counter to our hypothesis that the home environment would mostly yield joint engagement bouts that were based in mothers' verbal input only, multimodal behaviors were the most frequent form of engagement around infant object interactions—accounting for 53% of infant bouts that involved any mother input. Figure 4A shows timelines of each infant's object bouts (one infant per row) across the two hours. The color of each infant object bout denotes the mother behaviors that occurred during the infant object bout. As shown, multimodal (red) bouts predominated across the two hours: Mothers were more likely to engage multimodally—through manual action and object-related language combined (M = 62.79 bouts, SD = 24.62)—than to engage through only manual action (M = 28.45, SD = 16.66) or only object-related language (M = 26.84, SD = 13.60), as confirmed in a one-way repeated-measures ANOVA (N=38) (Wilks' Lambda = 0.127, $F(3,35) = 80.06$, $p<.001$, η^2=0.873; Figure 4B). Post-hoc comparisons indicated that the frequency of multimodal bouts differed from frequencies of bouts involving only manual action or only language input (p's<.001).

How was multimodal input—the most frequent form of joint engagement—temporally distributed? Bouts of multimodal engagement were distributed across time in bursts followed by short breaks (indicated by the spacing between red blocks in Figure 4A). Figure 4C and inset 4D are histograms of the time span between multimodal joint engagement bouts (i.e., time in minutes between offset of a multimodal bout and onset of the *next multimodal bout*). The median inter-bout time span was 0.26 min (M = 1.16, *Range* = 0.0002-43.59, *SD* = 2.56) or 15 s. "Long breaks" defined as inter-bout time spans lasting at least 5 min accounted for only 6% of all inter-bout time spans, and occurred 3.5 times across the two-hours on average for an infant (*Range* = 1-7 times). Infants' longest inter-bout time span averaged 13.95 min (*Mdn* = 11.90, *SD* = 7.52, *Range* = 5.86-43.59).

Does Proximity Frame Rates and Latencies of Touch, Gesture, and Language?

We next examined whether mothers jointly engaged at higher rates and joined infants more quickly when they were proximal to them.

Mother and infant dyads were in close proximity for over half of the visit; note that proximity does not imply joint engagement. Over the two-hour session, dyads were proximal 65% of time (*SD* = 11%, *Range* = 32-83%); across the two-hours, dyads entered proximity 76.53 times (*SD* = 24.35, *Range* = 36-127). Most bouts of proximity were short (*Mdn* = 14.38 s) but the longest bout lasted 34 min (mother was using her laptop while her infant played with a cell phone on the couch). Infants were as likely to manipulate objects in the context of proximity as they were when outside proximity after controlling for base rates of mothers' location and infants' object play (χ^2 = 0.119, *p* = .730).

Consistent with our hypothesis, rates of mother touch, gesture, and object-related language were greater when mothers were proximal than when they were not proximal (Table 1). Although mothers clearly cannot touch the object(s) of infant action when *not proximal*, they could talk or gesture; when mothers were *proximal* they could – but need not – engage with the objects of infant touch in any modality. Findings show that the rates of mother behaviors in the context of proximity at least tripled the rates outside proximity. Mothers engaged with the objects of infant touch when they were proximal at rates much higher than when they were not proximal. Figure 5 illustrates findings with timelines of an infant's bouts (colored by co-occurring mother behaviors), mother manual/verbal input, and proximity to mother. Panel A shows behaviors across the two hours, whereas Panel B zooms in on the first 30 minutes. Panel C spotlights examples of three combinations of maternal behavior during infant object bouts. As shown in panels A and B, mother language in the absence of proximity was very rare compared to mother language (with or without manual action) when proximal.

Did physical proximity to infant change mothers' latency of following in with object-related language? We focused on mothers' talk because mother touch and gesture did not occur from afar (Table 1). As hypothesized, when mothers were proximal to infants, they followed infant's touch with relevant language quicker (*Mdn* = 2.86 s) than they did when not proximal to infants (*Mdn* = 4.99 s), with latencies being nearly halved under conditions of proximity. A paired-samples t-test confirmed that normalized mean latencies (with a natural log transformation) to follow-in with language were quicker when mother was proximal versus not proximal, $t(36) = 3.06$, $p=.004$; 95% CI [1.10 s, 1.65 s]; $d = 0.50$. Table 2 shows the median latency of mothers' follow-in object-related language when proximal and not proximal.

Discussion

Joint engagement is a critical context for infant learning and development, as underscored by decades of rigorous, well-controlled studies of structured play interactions. Here we directly tested assumptions about features of joint engagement by documenting the modalities, rates, timing, and ebb-and-flow of spontaneous joint engagement in natural, home-based interactions. Infants experienced rich maternal input (predominantly multimodal input that combined touch and talk) in over half of their object interactions (a rate that can be interpreted as "often"), typically within 5 seconds. Findings converge on three take-home messages: joint engagement is not a lab-only phenomenon; everyday joint engagement is temporally structured and occurs in bursts; and physical proximity sets the stage for joint engagement. We discuss the implications of the nature and timing of everyday joint engagement for infant learning, research direction, and practice.

Joint Engagement is not a Lab-only Phenomenon

A key contribution of our study is that infant shared engagement with social partners is not a by-product of structured tasks. Despite the added challenges for infant-caregiver interactions in the home environment, joint engagement at home, particularly in the context of proximity, shared key properties with structured lab-based observations. Maternal input was most often multimodal, and latencies to follow infant behavior were quick. Mothers joined a similar percentage of infant object interactions at home as seen in lab settings (Suarez-Rivera et al., 2019; Yu & Smith, 2013), with features and forms of joint engagement generalizing across three infant ages. This work adds to the growing number of studies of infant experiences, learning, and development in ecologically valid contexts, in which researchers test the robustness of well-established lab-based phenomena (e.g., Adolph et al., 2012; Bergelson et al., 2019; Dahl, 2017; Fausey et al., 2016; Karasik et al., 2011; Roy et al., 2015; Tamis-LeMonda et al., 2017; VanDam et al., 2016).

Joint Engagement is Structured in Real-time and it Occurs in Bursts

Mothers' spontaneous manual and verbal inputs were highly structured—hands on objects consistently "led" and language consistently "followed" infant object interactions. Latencies of mother behaviors distinguished between "initiating" versus "following" infant touch. Such findings spotlight the role of hands for shared attention (see also Deák et al., 2018; Yu & Smith, 2013), and indicate that the temporal dynamics of manual actions differ from the dynamics of maternal speech. As mothers and infants go about their days, mothers' talk may support learning when it *follows* infant action in contingently responsive ways (Tamis-LeMonda et al., 2001; Tamis-LeMonda et al., 2014), whereas mothers' manual contact with an object may support learning by *leading* infants to objects. The structuring of caregiver input across different modalities may be a critical, understudied phenomenon that illuminates how social interactions guide attention and learning (e.g., caregiver's hands guide infant's eyes and hands to objects; then caregiver's language offers information about what infants are doing). A full understanding of the functions and temporal dynamics of caregiver input requires consideration of multiple behavioral modalities as they unfold over time. Indeed, future work might delve deeper into the temporal dynamics of verbal and manual behaviors using sliding time windows surrounding infant behavior.

The temporal structure of multimodal bouts resembled other "bursty" human social interactions (Abney et al., 2018; Xu et al., 2020), including child-directed speech (Tamis-LeMonda et al., 2017). We focused on the timing of multimodal bouts of joint engagement because infant learning happens in real time, and a fuller understanding of learning processes requires delving into the temporal structure of mature partners' input. Notably, a pattern of bursts separated by lulls in joint engagement may promote infant learning and development in critical ways. Perhaps bursts of repeated input about the objects of infant action scaffold learning (e.g., object properties, labels, functions, and affordances) through short-term integration mechanisms like massing (Schwab & Lew-Williams, 2016). Lulls may promote long-term integration mechanisms, as seen in infant memory consolidation (Vlach, 2019) and abstraction (Gomez et al., 2006). Quantitative measures of burstiness continue to be developed and may help expand the lens into the natural time-course of joint engagement (Abney et al., 2018). Experimental manipulation of the temporal structure or burstiness of inputs may be a valuable direction for future work.

Proximity Sets the Stage for Joint Engagement

Table-top structured tasks position mothers proximal to infants by design—to facilitate interactions. However, such settings prevent systematic inquiry into the role of interpersonal proximity for everyday interactions. Findings revealed that proximity critically guides multimodal, rich bouts of joint engagement, and that mothers displayed quicker language latencies when they were proximal. Thus, proximity to infants yielded benefits for joint engagement.

The role of interpersonal proximity in child development warrants further study. What are the moment-to-moment behaviors that bring mothers and infants together? Infant vocalizations may be salient signals for mothers to approach (Albert et al., 2018); infants may move in space to find mothers (Karasik et al., 2011); or other cues in the environment may alert mothers to check on their infants. Likewise, do individual differences in proximity map to individual differences in other aspects of child development? Though not the main question of this study, large individual differences characterized dyads' time in proximity. Questions on the cascading real-time processes that instigate proximal interactions, rich and timely bouts of joint engagement, and in turn infant learning, are ripe for investigation.

Costs and Benefits to Micro-Behavioral Coding

Of course, detailed micro-behavioral coding (particularly over long observations) requires substantial time investment. We spent 1,520 hours coding the onsets, offsets, and types of behaviors displayed by the 38 dyads in our sample (i.e., infant object bouts, language transcription, mother manual engagement, mother verbal engagement, and proximity). More precisely, coding took about 168 days of a full-time researcher's work. However, benefits outweighed costs. In return, we quantified multiple dimensions of joint engagement in natural settings, and we validated and extended prior research.

In light of such investments, how can micro-behavioral coding continue to be a viable pathway for understanding infant learning and development? Collaborative initiatives (e.g., Play and Learning Across a Year, play-project.org; CHILDES, childes.talkbank.org; ManyBabies, manybabies.github.io) present models on how to collect and code data communally across multiple laboratories, while openly sharing the products. By leveraging the power of video, multiple researchers can address unique questions using the same corpus. For example, researchers could use the videos we have shared here, along with our coding, to ask how mothers socialize gender norms, document strategies of infant emotional regulation, and so on. Research teams can design theoretically relevant "foundational coding passes" that others can build on by coding nested behaviors. Indeed, coding of theoretically relevant behaviors for one paper may yield insights for scientific discovery in another (Gilmore & Adolph, 2019). Strategies that leverage the power of micro-behavioral coding promise to unpack the complexities of real-world learning processes.

Moreover, although our approach to documenting real time processes is relatively rare, it can be implemented in cost-efficient ways within a single lab. For example, researchers can micro-code behaviors on a subset of infants or dyads (rather than on an entire sample) to illustrate timelines and key phenomena through case studies. Such visualizations reveal temporal processes that move beyond the traditional reporting of group averages. Open-access behavioral coding software such as Datavyu provides flexible tools for coding moment-to-moment behaviors, includes downloadable instructions, and offers support for researchers who are new to the approach.

From Joint Engagement to Learning

Our work elucidates how joint engagement might support infant learning and development through the moment-by-moment scaffolding of infant behavior. Indeed, lab-based research shows that caregiver multimodal input guides infants' eyes and hands toward objects, extends infants' visual engagement, and ultimately facilitates word-to-object mappings (Suarez-Rivera et al., 2019; Yu & Smith, 2012; Yu & Smith, 2016). However, the robustness of the phenomenon requires testing whether features of joint engagement in the home environment mirror behaviors identified in controlled lab studies. The current work provided such test and confirmed meaningful parallels.

Of course, the generalizability of joint engagement to infant learning in diverse samples remains to be tested. Participants comprised highly educated, white families of firstborn infants. And consistent with studies that have drawn on similar samples, mothers were highly engaged and responsive, with joint engagement occurring in 43-46% of infant play bouts and within brief time windows. Are the features of joint engagement documented here characteristic of families beyond this narrow sample? Research on mother-infant everyday interactions in unrepresented cultures has shed light on both similar and unique ways that mothers and infants engage around objects (e.g., Kuchirko & Tamis-LeMonda, 2019; Little et al., 2016). Yet such studies are rare, leaving open questions about how often and how mothers from different cultural backgrounds spontaneously engage with their infants around objects. What behaviors do mothers display when sharing attention with infants, and to what extent is joint engagement a common context of infant learning across cultures?

Implications

Findings have implications for science and practice. Most centrally, developmental phenomena hypothesized to be key for infant learning should be investigated in lab and everyday contexts alike to understand real-world inputs that propel learning, and to model those inputs to further elucidate potential mechanisms of change. Specifically, mature partners scaffold infant word learning by providing timely and structured input. Computational models of infant word learning may be trained more effectively with data ordered in time that combines massed and spaced input. Furthermore, other developing systems, such as robots, may benefit from learning schedules during difficult tasks (such as object handling; Ito et al., 2006) in which mature partners contingently respond to learners' actions and provide repeated feedback interspersed with independent activity by the learner. Finally, future work might implement more nuanced coding, and involve samples across a wider age window to test for developmental changes in learning processes. Mechanisms of change should only matter if they play out in everyday settings, where infants explore, move, learn, and grow.

From a clinical perspective, findings suggest that infant action spurs triadic engagement in ways that are tightly linked to physical proximity between an active infant and a responsive caregiver. Infants actively contribute to shared interactions by engaging with objects and approaching caregivers in space (Karasik et al., 2011). In turn, caregivers may join their infants in play with touch, talk, and gesture by maintaining proximity but also allowing independent infant play (Figure 4). However, what happens if dyadic opportunities to jointly engage with objects are hampered? Infants with motor and/or language delays, or developmental disorders such as Autism Spectrum Disorder may less often seek social partners to initiate exchanges around objects (Srinivasan & Bhat, 2020). Caregivers themselves may limit opportunities to jointly engage with objects depending on their stress (Ward & Lee, 2020), mental health (Lovejoy et al., 2000), sleep quality (McQuillan et al., 2019), and so on. Extending research to populations at risk for biological, psychological, and social problems promises to elucidate how caregivers and infants adapt to one another in real time, including how caregivers compensate for infants with motor or language delays.

Conclusion

Infants experienced frequent, timely, multimodal, and structured input from their mothers as they navigated the objects of their home environments. Findings confirm that joint engagement is a common context for learning, and that features of joint engagement at home align with features observed during structured tasks. At the same time, by venturing into the ecologically valid home setting, we shed light on undocumented characteristics of joint engagement, including the temporal unfolding of maternal behaviors from hands-to-talk and the role of proximity in dyadic interactions. Joint engagement is not a researcher-created phenomenon, but rather characterizes spontaneous exchanges of infants' everyday lives. Home-based observations extend the value of lab-based research by revealing how social inputs naturally organize to support infant learning and development.

Footnotes

[1] As a validity check, coders determined for five participants' joint touch bouts with latencies to mothers' manual action under 300 milliseconds (msec), whether the mother or the infant's touch initiated the joint touch bout. Latencies that were less than 100 msec occurred when mothers acted on the object first (mother-initiate). Latencies between 100 and 300 msec occurred when infants acted on the object first and mother's behaviors followed quickly (mother-follow).

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*, 1297-1316. https://doi.org/10.1111/cogs.12612

Adolph, K. E., Cole, W. G., Komati, M., Garciaguirre, J. S., Badaly, D., Lingeman, J. M., Chan, G. L. Y., & Sotsky, R. B. (2012). How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychological Science, 23*, 1387-1394. https://doi.org/10.1177/0956797612446346

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*, 1-11. https://doi.org/10.1111/desc.12641

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

Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. *Child Development, 60*, 381-398.

Belsky, J. (1980). Mother-infant interaction at home and in the laboratory: A comparative study. *The Journal of Genetic Psychology*, *137*(1), 37-47. https://doi.org/10.1080/00221325.1980.10532800

Belsky, J., Goode, M. K., & Most, R. K. (1980). Maternal stimulation and infant exploratory competence: Cross-sectional, correlational, and experimental analyses. *Child Development, 51*, 1168-1178.

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*, 1-10. https://doi.org/10.1111/desc.12715

Bruner, J. S. (1974). From communication to language-a psychological perspective. *Cognition, 3*, 255-287.

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*, 79-151.

Dahl, A. (2017). Ecological commitments: Why developmental science needs naturalistic methods. *Child Development Perspectives, 11*, 79-84. https://doi.org/10.1111/cdep.12217

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science, 17*, 270-281. https://doi.org/10.1111/desc.12122

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.

de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother–infant–object interactions. *Child Development, 87*(2), 494-512.

Dekking, F. M., Kraaikamp, C., Lopuhaa, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how.* Springer Science & Business.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition, 152*, 101-107. https://doi.org/10.1016/j.cognition.2016.03.005

Gardner, F. (2000). Methodological issues in the direct observation of parent-child interaction: Do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review, 3*, 185-198.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17*(8), 670-674. https://doi.org/10.1111/j.1467-9280.2006.01764.x

Gilmore, R. O., & Adolph, K. E. (2019). Open sharing of behavioral research datasets: Breaking down the boundaries of the research team. In Hall, R. Croyle, & A. Vogel (Eds), *Strategies for team science success: Handbook of evidence-based principles for cross-disciplinary science and practical lessons from health researchers.* Cham: Springer, pp 575-583.

Herzberg, O., Fletcher, K. K., Schatz, J. L., Adolph, K. E., & Tamis-LeMonda, C. S. (2021). Infant exuberant object play at home: Immense amounts of time-distributed, variable practice. *Child Development.* https://doi.org/10.1111/cdev.13669.

Isabella, R. A., & Belsky, J. (1991). Interactional synchrony and the origins of infant-mother attachment: A replication study. *Child Development, 62*(2), 373-384. https://doi.org/10.1111/j.1467-8624.1991.tb01538.x

Ispa, J. M., Fine, M. A., Halgunseth, L. C., Harper, S., Robinson, J., Boyce, L., Brooks-Gunn, J., & Brady-Smith, C. (2004). Maternal intrusiveness, maternal warmth, and mother–toddler relationship outcomes: Variations across low-income ethnic and acculturation groups. *Child Development, 75*(6), 1613-1631. https://doi.org/10.1111/j.1467-8624.2004.00806.x

Ito, M., Noda, K., Hoshino, Y., & Tani, J. (2006). Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model. *Neural Networks*, *19*, 323-337.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development, 82*, 1199-1209. https://doi.org/10.1111/j.1467-8624.2011.01595.x

Kuchirko, Y. A., & Tamis-LeMonda, C. S. (2019). The cultural context of infant development: Variability, specificity, and universality. In D. A. Daphne, E. Votruba-Drzal & P. Miller (Eds.), *Advances in child development and behavior* (pp. 27-63). JAI. https://doi.org/10.1016/bs.acdb.2019.04.004.

Little, E. E., Carver, L. J., & Legare, C. H. (2016). Cultural variation in triadic infant–caregiver object exploration. *Child Development*, *87*(4), 1130-1145.

Lovejoy, M. C., Graczyk, P. A., O'Hare, E., & Neuman, G. (2000). Maternal depression and parenting behavior: A meta-analytic review*. Clinical psychology review, 20*(5), 561-592. https://doi.org/10.1016/S0272-7358(98)00100-7

McQuillan, M. E., Bates, J. E., Staples, A. D., & Deater-Deckard, K. (2019). Maternal stress, sleep, and parenting. *Journal of Family Psychology, 33*(3), 349–359. https://doi.org/10.1037/fam0000516

Moore, C., & Dunham, P. J. (2014). *Joint attention: Its origins and role in development*. New York: Psychology Press.

Richards, J. E., & Casey, B. J. (1992). Development of sustained visual attention in the human infant. In B. A. Campbell, H. Hayne, & R. Richardson (Eds.), *Attention and information processing in infants and adults* (Vol. 1, pp. 30-60). New Jersey: Lawrence Erlbaum Associates, Inc.

Rossmanith, N., Costall, A., Reichelt, A. F., López, B., & Reddy, V. (2014). Jointly structuring triadic spaces of meaning and action: book sharing from 3 months on. *Frontiers in Psychology*, *5,* 1-22. https://doi.org/10.3389/fpsyg.2014.01390

Rothwell, P. M. (2005). External validity of randomised controlled trials: "To whom do the results of this trial apply?". *The Lancet, 365*, 82-93. https://doi.org/10.1016/S0140-6736(04)17670-8

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.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2*, 419-436.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*(8), 343-348. https://doi.org/10.1016/S1364-6613(03)00156-6

Srinivasan, S., & Bhat, A. (2020). Differences in caregiver behaviors of infants at-risk for autism and typically developing infants from 9 to 15 months of age. *Infant Behavior and Development*, *59*, 101445.

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology, 55*, 96-109. http://dx.doi.org/10.1037/dev0000628

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, *52*(6), 879. http://dx.doi.org/10.1037/dev0000125

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*, 748-767. https://doi.org/10.1111/1467-8624.00313

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., & Escobar, K. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science, 20*, e12456. https://doi.org/10.1111/desc.12456

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science, 23*, 121-126. https://doi.org/10.1177/0963721414522813

Tamis-LeMonda, C. S., Kuchirko, Y., & Tafuro, L. (2013). From action to interaction: Infant object exploration and mothers' contingent responsiveness. *IEEE Transactions on Autonomous Mental Development, 5*, 202-209. http://doi.org/10.1109/TAMD.2013.2269905

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

Tomasello, M., & Farrar, J. (1986). Joint attention and early language. *Child Development, 57*, 1454-1463.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech Language, 37*, 128-142. https://doi.org/10.1055/s-0036-1580745

Vlach, H. A. (2019). Learning to remember words: Memory constraints as double-edged sword mechanisms of language development. *Child Development Perspectives, 13*(3), 159-165. https://doi.org/10.1111/cdep.12337

Vygotsky, L. S. (1987). *Mind in Society: The Development of higher mental process.* Cambridge, MA, Harvard University Press.

Ward, K. P., & Lee, S. J. (2020). Mothers' and fathers' parenting stress, responsiveness, and child wellbeing among low-income families. *Children and Youth Services Review, 116*, 105218.

Xu, T. L., de Barbaro, K., Abney, D. H., & Cox, R. F. A. (2020). Finding structure in time: Visualizing and analyzing behavioral time series. *Frontiers in Psychology, 11*, 1-19. https://doi.org/10.3389/fpsyg.2020.01457

Yamamoto, H., Sato, A., & Itakura, S. (2019). Eye tracking in an everyday environment reveals the interpersonal distance that affords infant-parent gaze communication. *Scientific reports, 9*, 1-9. https://doi.org/10.1038/s41598-019-46650-6

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244-262.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE, 8*, e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology, 26*, 1235-1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, C., & Smith, L. B. (2017a). Hand-eye coordination predicts joint attention. *Child Development, 88*, 2060-2078. https:/doi.org/10.1111/cdev.12730

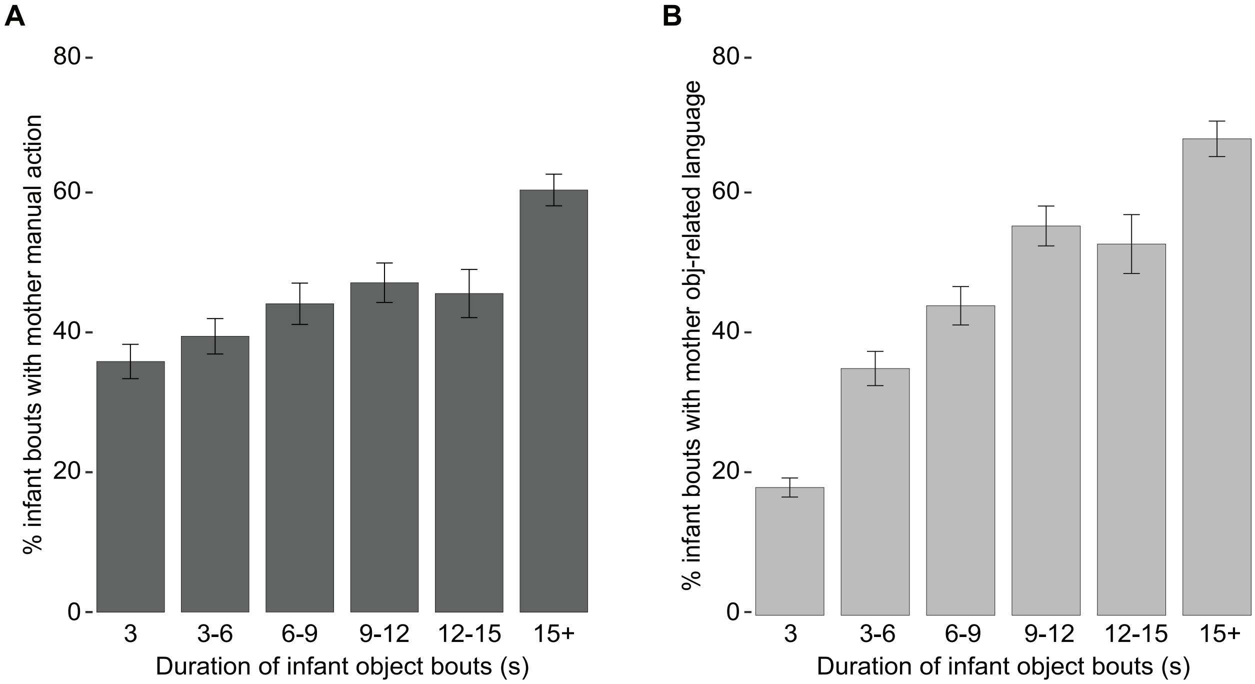
Yu, C., & Smith, L. B. (2017b). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science, 41*, 5-31. https://doi.org/10.1111/cogs.12366

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

Figures

Figure 1

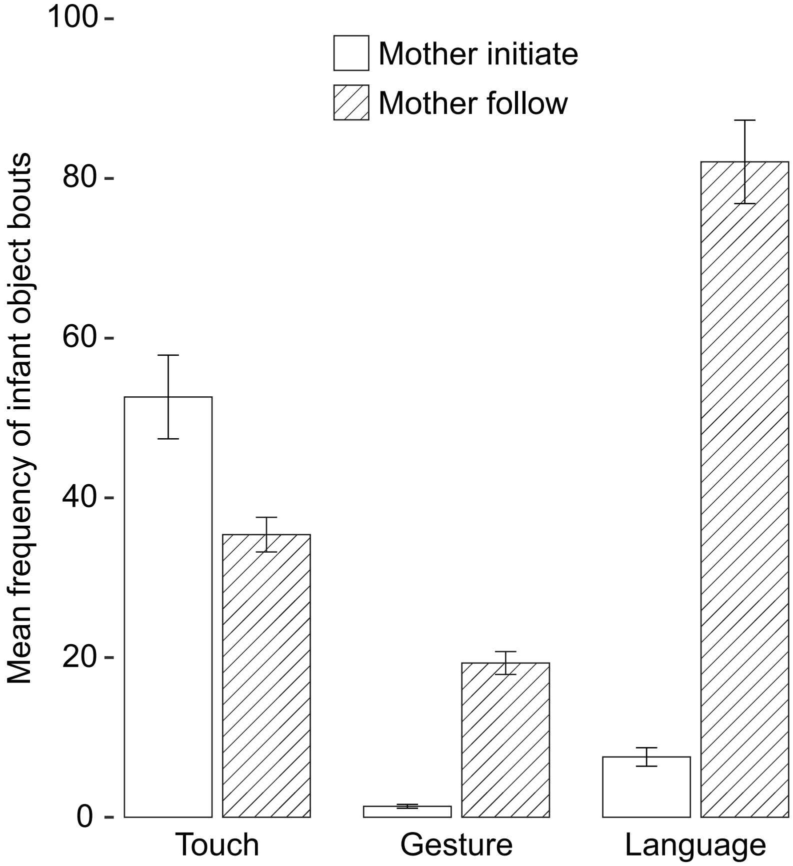
Infant Object Bouts with Mother Manual Action and Language



Note. Percentage of infant object bouts of varying durations in which mothers jointly engaged through manual action (A) and object-related language (B).

Figure 2

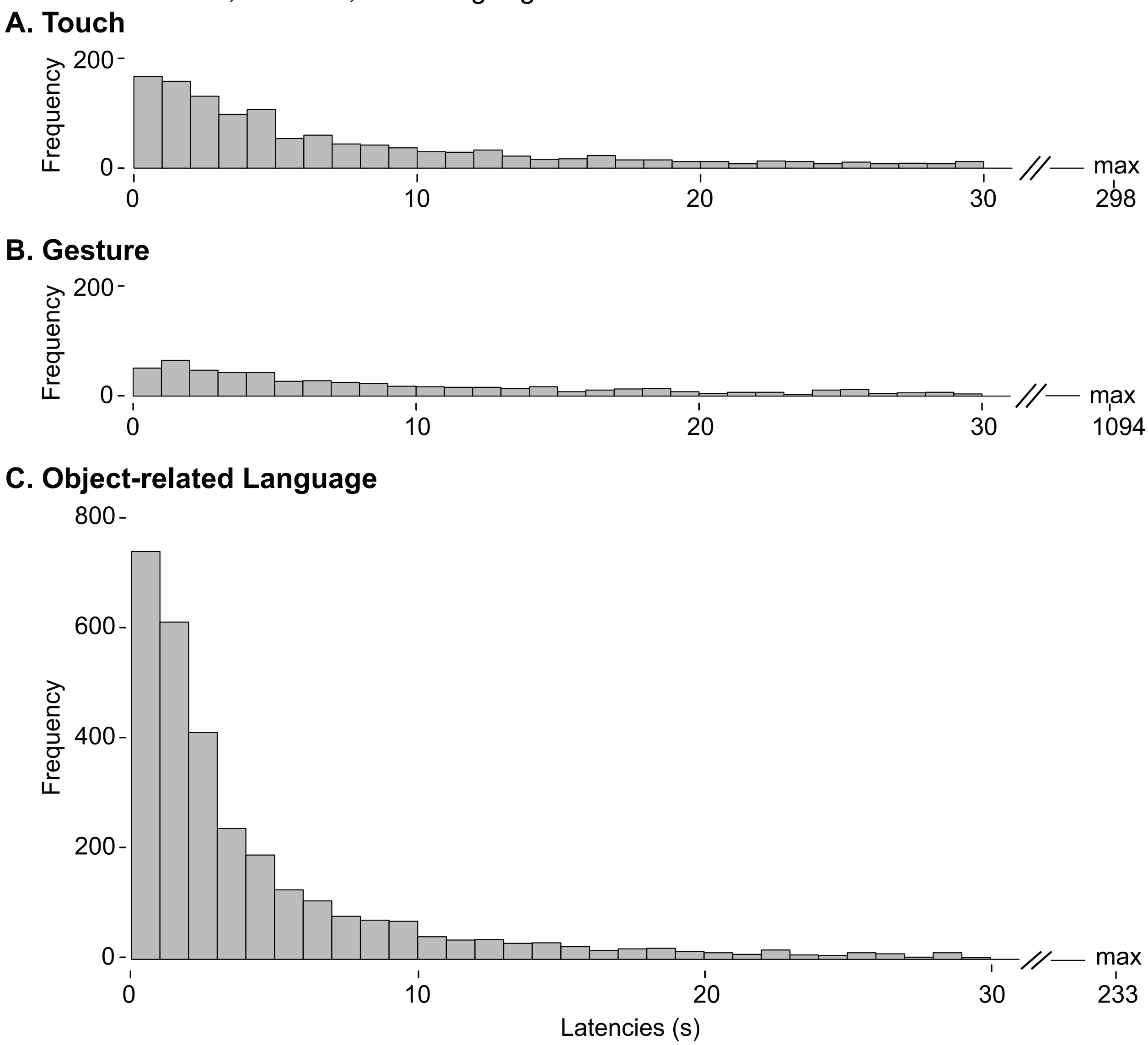
Infant Object Bouts with Mother Behaviors in which Mother Initiated or Followed-in



Note. Mean frequency of infant object bouts initiated or followed by mother with mother touch, gesture, and object-related language.

Figure 3

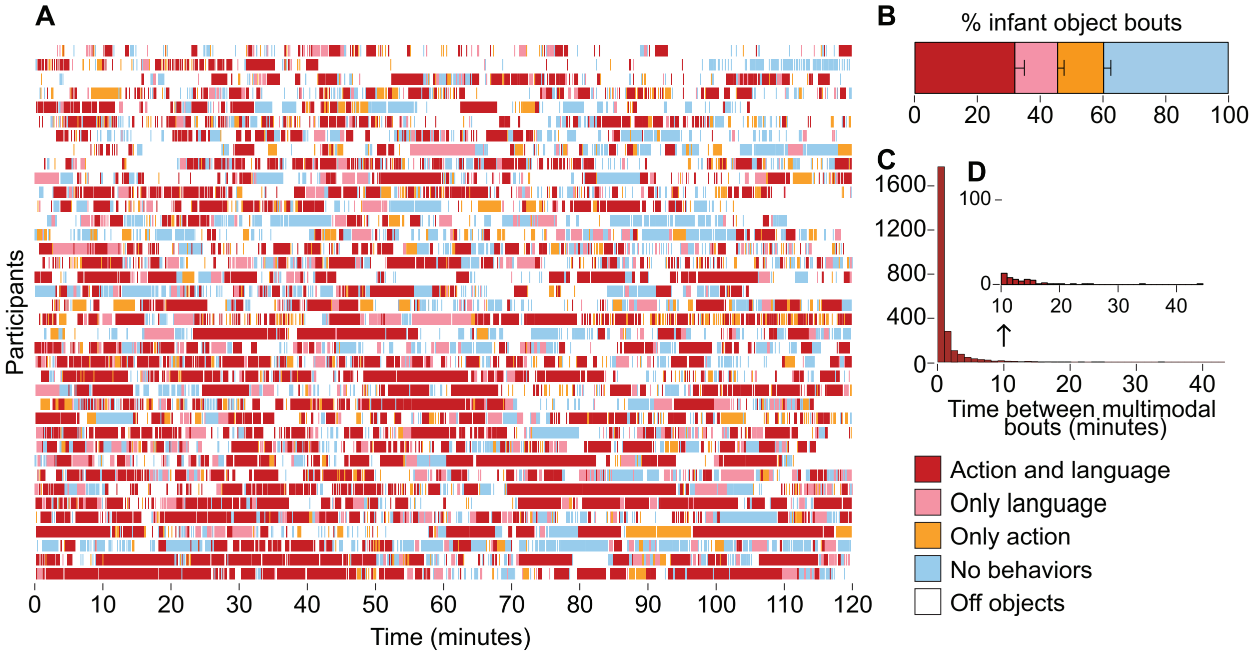
Latencies for Mother to Follow-in (i.e., 0.1 s latencies and greater) on Infant Object Bouts with Touch, Gesture, and Language



Note. Frequency distributions of follow-in latencies (that is, latencies equal or greater than 0.1 s) from onset of infant bouts to onset of mother behaviors that occurred after the infant bout started (A: Touch, B: Gesture, C: Object-related language). Tail of the distributions extends to the Maximum number shown in each panel. Note y-axis ends at 200 for panels A and B.

Figure 4

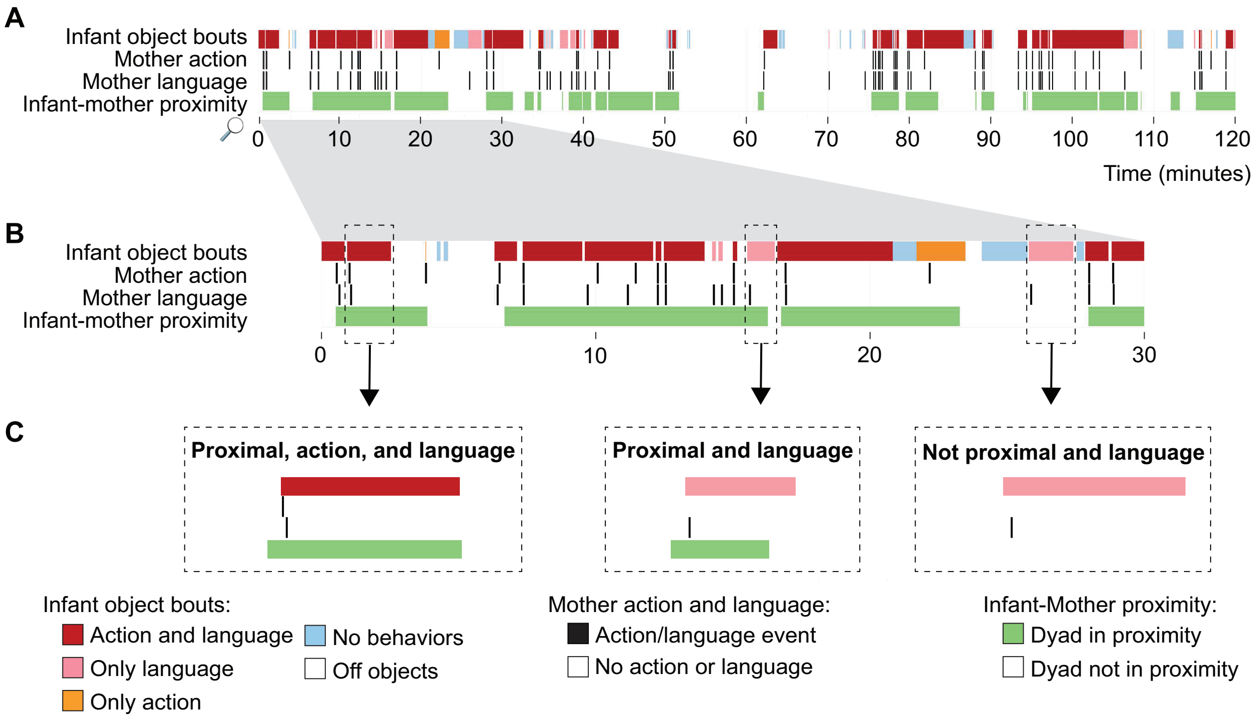
Temporal Distribution of Infant Object Bouts with Multimodal Mother Engagement



Note. (A) Timelines for each infant's object bouts across the 2 hours. Red blocks denote multimodal bouts with mother manual action and object-related language; pink and orange blocks denote bouts with only object-related language and with only manual action, respectively; blue blocks denote infant bouts without mother object-related language and action; white blocks denote time off objects. Timelines are ordered from least to most time in multimodal joint engagement. (B) Mean percentage of infant object bouts that involved different combination of mother behaviors (red, pink, orange or blue). (C) Frequency histogram of time spans in minutes between multimodal infant object bouts. (D) Frequency of time spans >10 min.

Figure 5

Timeline of Infant Object Bouts, Mother Behaviors, and Interpersonal Proximity



Note. Panel A timelines of behaviors for a representative dyad across the 2 hours; behaviors are represented by blocks that extend horizontally for the duration of infant object touch and proximity, and black vertical lines for events of mother touch, gesture, and language. The first row shows onsets and offsets of infant object bouts with colors corresponding to co-occurring mother behaviors. The second row shows onsets of co-occurring mother manual actions (i.e., first touch and/or first gesture). The third row shows onsets of mother first co-occurring object-related utterances. The fourth row shows onsets and offsets of bouts of infant-mother proximity. Panel B timelines zoom in on 30 minutes. Panel C zooms in on example combinations of mother manual actions, language, and proximity: mother multimodal input in the context of proximity; mother language in the context of proximity; and mother language outside proximity. Note that mother language in the absence of proximity was very rare relative to language with or without manual action in the context of proximity in the 30- and 120-minutes timelines.

Tables

Table 1

Mean Rates and Number of Infant bouts with Mother Behaviors by Proximity

	Proximal	Not Proximal	t-test
Rate of Touch	**0.59** (87.21/148.32)	**0.02** (0.79/47.92)	***t*(37)=27.99, *p*<.001**
Bouts with touch	87.21 (*SD*=36.65, *SE*=5.94)	0.79 (*SD*=1.32, *SE*=0.21)	*t*(37)=14.58, *p*<.001
Bouts without touch	61.11 (*SD*=25.57, *SE*=4.15)	47.13 (*SD*=34.05, *SE*=5.52)	
Total	148.32	47.92	
Rate of Gesture	**0.14** (20.21/148.32)	**0.01** (0.47/47.92)	***t*(37)=15.37, *p*<.001**
Bouts with gesture	20.21 (*SD*=9.39, *SE*=1.52)	0.47 (*SD*=0.83, *SE*=0.13)	*t*(37)=13.10, *p*<.001
Bouts without gesture	128.11 (*SD*=43.50, *SE*=7.06)	47.45 (*SD*=34.13, *SE*=5.54)	
Total	148.32	47.92	
Rate of Language	**0.55** (81.76/148.32)	**0.16** (7.87/47.92)	***t*(37)=16.71, *p*<.001**
Bouts with language	81.76 (*SD*=30.88, *SE*=5.01)	7.87 (*SD*=5.55, *SE*=0.90)	*t*(37)=15.35, *p*<.001
Bouts without language	66.56 (*SD*=24.93, *SE*=4.04)	40.05 (*SD*=32.84, *SE*=5.33)	
Total	148.32	47.92	

Table 2
Latencies to Mother Language by Mother-Infant Proximity

	Bouts with Proximity	**Bouts without Proximity**	**t-test**
Number of **mother-initiate** bouts	7.16 (*SD*=6.75)	0.40 (*SD*=0.82)	
Number of **mother-follow** bouts	74.60 (*SD*=29.97)	7.47 (*SD*=5.49)	
Mother-follow latency (mean of participants' medians)	2.86 s *SD*=1.15 *Range*= 1.33 – 6.14	4.99 s *SD*=3.89 *Range*= 1.22 – 18.20	$t(36)=3.06$, $p=0.004$
	Total proximal bouts with Language = 81.76	Total not proximal bouts with Language = 7.87	

Note: t-test was computed between means of log(mothers' 0.1 s latencies and greater to follow-in with language).