From Play to Language: Infants' Actions on Objects Cascade to Word Learning

Catalina Suarez-Rivera[1,2], Emily Linn[3], & Catherine S. Tamis-LeMonda[1]

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

Revisions (round 3): February 25th, 2022

Author Note
Catalina Suarez-Rivera https://orcid.org/0000-0002-5900-9670
Catherine Tamis-LeMonda https://orcid.org/0000-0002-9071-4028

We have no conflicts of interest to disclose. Video examples, coding manuals, spreadsheets and scripts for data analysis are shared on Databrary. Correspondence concerning this article should be addressed to Catherine Tamis-LeMonda, catherine.tamis-lemonda@nyu.edu, 246 Greene ST, New York, NY 10003.

Acknowledgements. We thank participating families for letting us into their homes to learn about infant play and language. We are grateful to Katelyn Fletcher, Jacob Schatz, Orit Herzberg, Shimeng Weng, Aria Xiao, and Kelsey West for help with data collection, data coding, and transcription. This work was funded by National Institute of Child Health and Human Development Grant R01HD094830 and the LEGO Foundation.

Abstract

Infants build knowledge by acting on the world. We conducted an ecologically-grounded test of an embodied learning hypothesis—that infants' active engagement with objects in the home environment elicits caregiver naming and cascades to learning object names. Our home-based study extends lab-based theories to identify real-world processes that support infant word learning. Frame-by-frame coding of 2-hour-video recordings of 32 mothers and their 18- to 23-months-olds focused on infant manipulation and mother and infant naming of 245 unique objects. Objects manipulated by infants and/or named by mothers were more likely to appear in infants' vocabularies and spontaneous speech relative to non-manipulated objects and objects that were not named. Furthermore, vocabularies of 5,520 infants hosted on Wordbank revealed early age of acquisition for words of objects that mothers named and infants manipulated. Infants actively build object-word mappings from everyday engagements with objects in the context of everyday social interactions.

Key words: word learning, embodied learning, object play, parent responsiveness, naturalistic interaction.

Max word count: 150

Learning words is a multi-step process. Infants must identify the relevant phonemes in their language (Werker, 1989); figure out which syllables belong together as “words” (Saffran, Aslin & Newport, 1996); and ultimately map words to real-world referents (Smith & Yu, 2008). The mapping problem is not easy, because any given word could have numerous meanings (Quine, 1964). And so, researchers offer several possibilities for how infants disambiguate word meaning. Cognitive constraints may channel early word learning (e.g., Jones & Smith, 2002; Markman, 1987), equipping children with biases about the referents of words—for example, that words refer to whole objects rather than object parts (Markman, 1987) or to one object exclusively as opposed to a larger class (Markman & Wachtel, 1988; Markman, 1989).

However, appeals to cognitive constraints may be unnecessary (or at most only part of the story). From an embodiment perspective (Smith & Gasser, 2005), infants actively build knowledge through a bottom-up process that binds together their own actions with social and contextual cues to word meaning. Indeed, infants are active word-learners who develop through interactions with the people and things around them (Bruner, 1974; Gibson, 1988; Piaget, 1969; Vygotsky, 1978). Infants look to other people’s eyes and hands to identify the objects of talk (e.g., Baldwin, 1993); extract statistical regularities from speech streams (Saffran, Aslin & Newport, 1996); move their bodies through space and manipulate objects in ways that elicit verbs and nouns about what they are doing (Tamis-LeMonda et al., 2020); and track co-occurrences between nouns and objects to discover word-object mappings (e.g., Smith & Yu, 2008; Yu & Smith, 2011).

Here, we advance the embodied learning hypothesis. As infants actively engage with objects, their manual actions set in motion real-time cascades that make objects visually salient; elicit contingent verbal input; and ultimately simplify the learning task. If so, infants should learn and produce the words that appear in mothers' everyday speech that refer to the objects that infants manipulate. We test these real time processes in the ecologically valid home setting where infants have opportunities to interact with and learn the words for a wide range of common objects.

The Active Infant: Engagement with Objects Make Objects Salient

A first step in learning object names is to bring the "to-be-named" object to the perceptual and psychological foreground. Seminal work by Nelson (1973) revealed that the 50 first words produced by eighteen infants, according to caregiver diaries, referred to objects that infants likely manipulated and visually attended to at home, including balls and cups rather than also available objects such as walls and windows. Indeed, many objects are prevalent and frequent in infants' worlds—floors, ceilings, couches—but it takes several more months for infants to learn those words relative to words like spoon, ball, book, and so on. Why? We propose that infants learn words for manipulatable objects because learning occurs when infants' bodies (e.g., eyes and hands) are actively involved with meaningful and perceptually salient objects (Hirsh-Pasek & Golinkoff, 2012; Pruden et al., 2013).

At each moment in time, countless objects surround infants, but infants select and visually isolate from the environment the objects that they manipulate (Smith, Yu & Pereira, 2011; Yoshida & Smith, 2008; Yu et al., 2009). Indeed, infants visually attend longer to objects they touch relative to those that they don't (Ruff, 1986; Ruff & Lawson, 1990). Figure 1 illustrates that infants often look at the objects they hold and not at the objects around them. Objects of infant play are salient and present in infants' visual field. Infants' first-person views (from a mounted head-camera) are selective and consist of a non-uniform distribution of objects that is dominated by a small set of objects (Clerkin et al., 2017). Infants' visual plus manual engagement with objects, compared to viewing objects at a distance, generates salient, multi-sensory perceptual feedback about the attended-object and self (Lockman, 2000; Needham, 2000; Rochat, 1989; Soska, Adolph & Johnson, 2010; Yu & Smith, 2012; Yu et al., 2009).

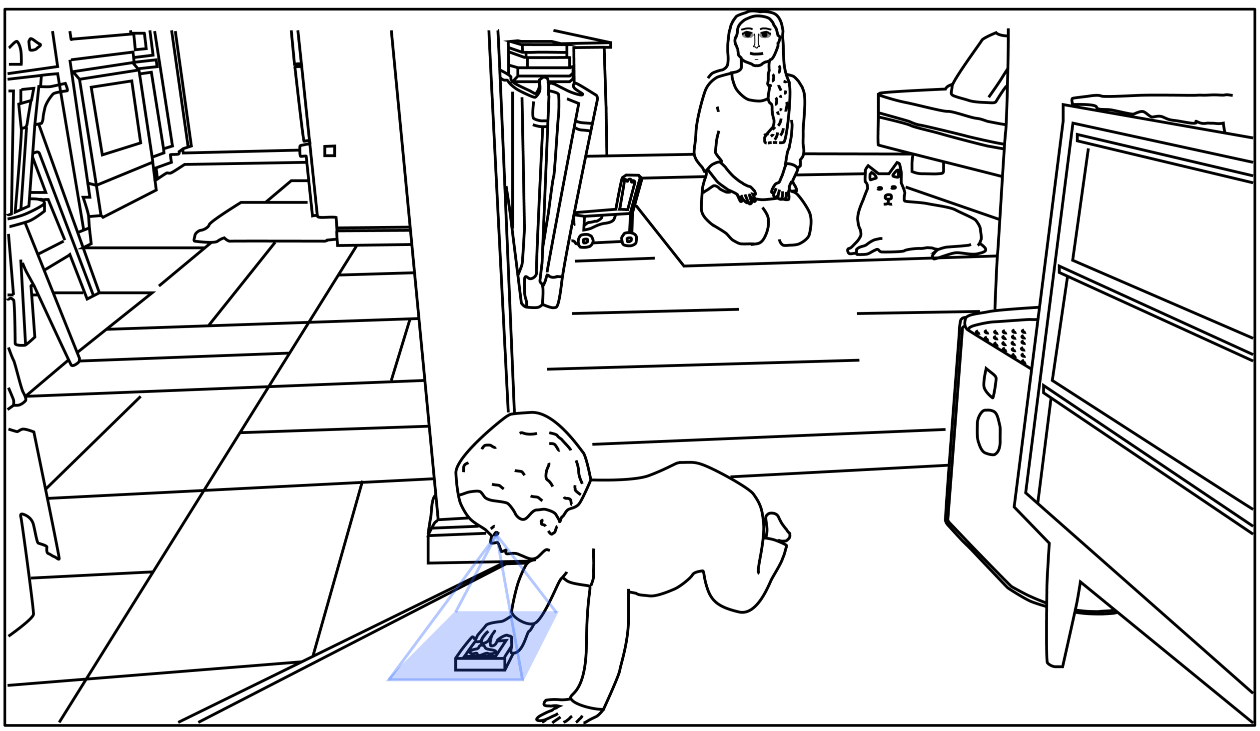

Figure 1. A sketch drawn from a video frame of a study dyad, illustrating in blue the object present in the infant's visual field.

In turn, infants should learn the names for the objects that they manipulate and see. Indeed, objects that dominate infants' everyday visual scenes at home—such as shirts, bowls, cups, and spoons—map to words that infants acquired earlier in development compared to objects available at home but that were infrequent in infants' visual scenes (Clerkin et al., 2017). And infants' manual actions facilitate the mapping of words to held objects (Pereira, Smith & Yu, 2014; Yu & Smith, 2012). Furthermore, infants may build robust object-representations through self-generated play that promotes mapping words to objects: Infants who generated variable object views at 15 months experienced large vocabulary growth over the next 6 months (Slone, Smith, & Yu, 2019). Thus, the embodied learning hypothesis suggests correspondence between the objects of infant manual engagement (and serendipitously, visual field dominance) and the words they know and produce.

The Active Infant: Engagement with Objects Drive Input

Infants' active engagement with objects alone is insufficient for word learning. Infants learn in a social context, in which object interactions cascade to hearing object names (Custode & Tamis-LeMonda, 2020). A key mechanism through which object interactions drive learning is by eliciting contingent, object-relevant language input. Infants learn features of language (for example the speech sounds of a foreign language) from a live interlocutor but not televised speaker (Kuhl, Tsao & Liu, 2003), and in interactive and responsive contexts (Roseberry, Hirsh-Pasek & Golinkoff, 2013). Specifically, infant actions are critical to establishing co-regulated, contingent social interactions that promote learning (Song, Spier & Tamis-LeMonda, 2014). Lab studies reveal that infants' visual and manual engagement lead social partners to jointly engage with the same object (Yu & Smith, 2013; Yu & Smith, 2017) and infants are likely to learn words during moments of shared engagement (Akhtar, Dunham & Dunham, 1991; Baldwin &

Markman, 1989; Tomasello & Todd, 1983). Fourteen-month-olds' manual engagement with objects is more likely to elicit maternal referential language and object touch than is off-task behavior (e.g., neither engaging with objects nor communicating with mother) (Tamis-LeMonda, Kuchirko & Tafuro, 2013). And caregivers are more likely to label objects during play that involves infants' eyes and hands, than during activities that do not involve object manipulation (West & Iverson, 2017). In turn, infants use caregivers' naming cues to discover correct word-object mappings (Frank, Goodman & Tenenbaum, 2009). Thus, by an embodied perspective, infants should be exposed to, and ultimately learn, the words for the objects they manipulate during home activities.

The Full Path: From Play to Naming to Learning

An embodied perspective of language learning posits that infant active engagement with objects serves to: (1) make objects salient, (2) elicit language input, and (3) thus predict word learning. Yet to our knowledge, only two lab-based studies tested a cascade from "play-to-input-to-learning". Infant-parent dyads sat at a table in the lab and played with three novel objects for 5-6 minutes. Caregivers were told the names of the objects and interacted spontaneously with infants without further direction. In both studies, infants learned the words for visually dominant objects in their hands (the active infant) during caregiver naming (the responsive adult) (Pereira, Smith & Yu, 2014; Yu & Smith, 2012).

Does the embodied learning hypothesis scale up to real-world infant object manipulation, mother naming, and word learning? A rigorous, ecological test of whether active object engagement ultimately cascades to language learning rests on studying everyday behaviors in the home setting. Indeed, the real-world tends to be more complex than the lab: Hundreds of objects are available for play; mothers attend to multiple activities; physical proximity between infant and mother is not guaranteed; and infants build unique vocabularies and produce words that may or may not correspond to the objects that they manipulate and mothers name.

Current Study

Our study provides a critical test of the embodied learning hypothesis by testing predictions based on 2-hour observations of infants and mothers during everyday activities. We identified the specific nouns in infants' vocabularies in two ways: through an infant vocabulary checklist (based on mothers' report) and transcription of the words that infants spontaneously produced during the observation. Additionally, we extended analyses to archival data on the age of acquisition for our list of words based on over 5,000 infants hosted on the Wordbank repository (Frank et al., 2017). Understanding the real-world processes that support infant word learning advances the science of development to infants' day-in day-out experiences in natural contexts (Thelen & Smith, 1998).

Three research questions guided analyses, each focusing on a unique dependent measure of language development:

(RQ1) To what extent do the words in infants' productive vocabularies (by mothers' report) map to the objects that infants manipulate and mothers name at home?

(RQ2) To what extent do infants produce the words for objects that they manipulate and mothers name?

(RQ3) Do infant manual engagement and mother naming in this sample predict the ages of acquisition of words based on the learning trajectories of over 5,000 infants?

For research questions 1 and 2, we expected associations between pairs of variables in the cascade model—between the objects of infant play and the words in infants' vocabularies (RQ1) and the words that infants produce (RQ2); between mother naming and the words in infants' vocabularies (RQ1) and the words that infants produce (RQ2); and between mother naming and the objects of infant play (RQ1 & 2). Considering the full model, we predicted that infants' vocabularies would be more likely to include words for the objects that: (1) infants manipulated, and (2) mothers named, relative to words for objects neither manipulated by infants and/or named by mothers (RQ1). Similarly, we expected to identify the same associations for infant word production (RQ2).

For research question 3, we predicted that the words for objects that infants frequently manipulated and mothers frequently named in this sample would map to words that infants generally acquire at young ages based on archival data from Wordbank. We estimated word age of acquisition from the vocabularies of thousands of infants and mapped those ages to the objects infants manipulated and mothers named in this sample.

Method

Participants

We video-recorded 32 infants (16 females) aged 18 to 23 months (M = 20.5, SD = 2.5, 95% CI [19.63 - 21.37]), a period of rapid language learning. Participants were recruited from a large urban city through hospitals, referrals, and brochures. Infants were full term without developmental delays, and from English-speaking families. Most mothers identified as White (81%); the remaining were Asian (7%) and Mixed (12%). Mothers ranged from 26 to 49 years of age (*M* = 35.08, *SD* = 5.23, 95% CI [33.27 – 36.89]); most (91%) had earned college or higher degrees, and 62% worked part- or full-time. Families received a $75 gift card for their participation. The present study was conducted with written informed consent obtained from a parent or guardian for each infant before 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.

Procedure

A female researcher video-recorded dyads during naturalistic activity for 2 hours between 8:30 a.m. and 5:30 p.m. on a weekday. The researcher used a handheld digital camera (30 fps) to record infant and mother behaviors with minimal interference. Mothers were asked to go about their day as if the researcher was not present, but to remain inside the home. The researcher recorded both infant and mother but kept the camera on the infant if they separated. The duration of the recording ranged from 114.7 to 122 min (*M* = 120.20, *SD* = 1.18, 95% CI [119.79 – 120.61]), with most recordings (94%) lasting the full 2-hours. At the end of recordings, mothers were interviewed about infants' productive vocabulary using the MacArthur-Bates Communicative Development Inventory (MB-CDI), a reliable (Cronbach's coefficient alpha = .96; test-retest correlation was .95, $p < .01$) and valid measure of vocabulary (Fenson et al., 1994).

Coding

Trained researchers coded each video for infant object manipulation and transcribed mother and infant language. Coders used Datavyu software (datavyu.org) which allows for frame-by-frame analysis and time-locks user-defined events to the video file. To examine inter-observer reliability, a primary coder scored each coding pass across each 2-hour visit, and a second coder independently scored each coding pass on 25% of each two-hour visit randomly selected from the beginning (i.e., first 40 minutes), middle (i.e., second 40 minutes), and end of the visit (i.e., third 40 minutes). Random distribution of reliability blocks ensured that coders agreed across different types of activities and time frames within and across dyads. Primary and secondary coder agreement at the frame level was calculated using Cohen's kappa for all 32 videos to ensure agreement was achieved in all videos. For all coding passes, reliability was high, and Cohen's Kappa averaged 0.84 (*Range*= 0.70 - 0.99). Coding manuals, transcription guidelines, video examples of the behaviors of interest, spreadsheets, and scripts for data analysis are shared via Databrary (Suarez-Rivera & Tamis-LeMonda, 2021, available at https://nyu.databrary.org/volume/1358). Repository of video-recordings can be accessed in Databrary (Tamis-LeMonda & Adolph, 2017, available at https://nyu.databrary.org/volume/563).

Selection of Concrete Words (i.e., MB-CDI subset)

We restricted coding to 271 nouns (drawn from a larger set of 294 nouns) from the MB-CDI that referred to objects that could be displaced in space within the confines of the home (e.g., sky was excluded because infants cannot manipulate the sky and sky is outside the home). We excluded objects that could *not* be moved inside the home, and therefore could not be tested in analyses of objects manipulated versus not manipulated (e.g., stairs). Furthermore, words (3) that were repeated twice on the MB-CDI for their different meanings (e.g., chicken the animal and the food) were excluded. Collapsing semantically equivalent words (e.g., dog and puppy) further reduced the nouns for analysis to a final set of 245. Appendix S1 includes all words used in analyses. Appendix S2 describes how we processed data for each grouped word (i.e., whether the infant manipulated the object, the mother named the object, the infant knew/named the object, and average age of acquisition of the object name in Wordbank).

Transcription of Mother and Infant language

Mothers' speech to infants and infants' vocalizations were transcribed at the utterance level by experienced coders following guidelines developed in our lab in consultation with language experts on a national project of child play (https://www.play-project.org/coding.html#Transcription). Transcription guidelines were adopted from conventions of the Codes for the Human Analysis of Transcripts (CHAT) (CHILDES; MacWhinney, 2000). Utterances are meaningful units of information distinguished by grammatical closure or a discernible pause. Speech from electronic toys and media was not transcribed.

MB-CDI Words in Mother and Infant language

The MB-CDI subset of 245 words were identified in language transcripts using an automated computer script and manual human coding. The computer script searched for the "root" of each MB-CDI subset word in utterances (e.g., "dog" in the utterance "*line up the doggies*"). Then, human coders read the utterances and corresponding MB-CDI subset of words, indicating if the script missed or incorrectly identified an MB-CDI subset word. Human coders flagged mistakes that needed correction including false positives (e.g., "can" in "*can you throw it?*" is not the noun) and false negatives (e.g., "rabbit" was not found in *"this is a bunny").* Mistakes were corrected to yield the coding of MB-CDI subset words in mother and infant language.

MB-CDI Word Referents in Infant's Manual Actions

Infant object interactions were defined as the manual displacement of an object in space. Object interactions referred to as *"object play"* were coded in 2 passes: Pass 1 identified bouts in which the infant manipulated an object or set of objects (with 3 seconds defining a break); pass 2 determined the identity of the whole object at the basic level (i.e., car instead of vehicle) using the specific noun form listed on the MB-CDI (e.g., instead of phone, the object was named "telephone"). Once all object interactions were coded, a computer script searched for the 245 MB-CDI subset of words (e.g., marking "cup" if infant touched a cup). Finally, human coders validated the script by correcting false positives and false negatives.

Available Sensory Information During Infant Naming

We conducted exploratory analyses on the available sensory information during infant naming by coding the 5 seconds before and after the naming event. If infants manipulated the object (as coded above), we classified the named referent as visually available in the infant's hands (as a 3D object or a 2D object representation, such as picture of dog in a book). When infants named objects not in their hands, we classified sensory information as visually available in the mother's hands, only visually available (i.e., present but neither manipulated by infant or mother), only auditorily available (e.g., present only in mothers' speech), or displaced (i.e., referent was not visible nor named). The resulting coding indicated a category of available sensory information for each object named by the infant. Approximately half of the objects named were produced by the infant only once. For the remaining objects, which individual infants produced multiple times, we chose the naming instance with more available sensory information about the object (e.g., if the infant named dog while holding a book with a picture of a dog, and while watching television, we coded the referent as being in the infant's hands). Our one-to-one mappings between objects and naming contexts weighted objects and infants equally in analyses.

Wordbank: Age of Acquisition for Words

Wordbank is an open database of infants' vocabulary growth (Frank et al., 2017). Analyses operated on Wordbank's instrument 7, which contains the productive vocabularies of 5,520 infants collected across multiple laboratories that administered the CDI-WS English form in the United States (i.e., Fenson et al., 2007; Fernald, Marchman & Weisleder, 2013; Thal, Marchman & Tomblin, 2013; Smith, Indiana University; Byers-Heinlein, Concordia University). Infants' age ranged from 16 to 30 months (M = 22.37, SD = 4.71, 95% CI [20.74 – 24.00]), generally aligning with the ages of 32 infants in the sample. Out of 4,094 infants with sex reported, 48% were females; out of 2,715 infants with ethnicity reported, 81% of infants were White, 8% Black, 5% Hispanic, 2% Asian and the remaining 4% Other ethnicities; and out of 2,776 infants with mothers' education reported, 58% had mothers with a college or higher degree. Age of acquisition for each word on the MB-CDI subset was estimated by submitting the Wordbank's percentage of infants who produced the word at 16 to 30 months to simple linear regression. The line of best fit for each word predicted the age in months at which 50% of infants produce each word. For example, "shoe" age of acquisition was 13.25 months and "basket" was 24.99 months.

Analysis Plan

We harnessed the resolution of data obtained from naturalistic recordings and MB-CDI vocabulary checklists to investigate associations from “play to language” at the level of individual words in the MB-CDI subset for individual infants (e.g., how often an infant’s productive vocabulary included words for manipulated objects relative to words for non-manipulated objects). In RQ1, we first examined associations between infant object play, mother naming, and infant language measures separately by calculating i) the odds ratio that an infant’s vocabulary included a word for an object the infant manipulated (versus not), ii) the odds ratio that an infant’s vocabulary included a word that the mother produced (versus not), iii) the odds ratio that a mother named an object that the infant manipulated (versus not). Then, a logistic mixed regression tested the full model—from infant play to mother naming to infant vocabulary. The model investigated whether infants’ object play (coded as 0/1 for no/yes) and mother naming (coded as 0/1 for no/yes) simultaneously predicted the odds ratio of infant’s vocabularies including words, while accounting for nesting of the 245 individual words in infants. The analysis plan for RQ2 was the same for RQ1 with words in infants’ spontaneous production serving as the outcome variable. We selected the best-fitting, most parsimonious model for RQ1 and RQ2. Specifically, we attempted to fit models with maximal random structure (Barr et al., 2013) and if the model did not converge, we performed likelihood ratio tests to reduce the model (Yu, 2015). Each term was retained in the final model if it improved model fit relative to a model without the term (as described in Results).

Odds ratios were used for RQ1 and RQ2 because their mathematical properties allow them to be easily estimated and modeled with logistic mixed regression.

Furthermore, they are appropriate and widely used measures of relative effects for binary outcomes (e.g., a word appearing or not in an infant's vocabulary). For example, odds ratios allowed us to compare in RQ1 the odds of an infant's vocabulary including words for objects that the infant manipulated and mother named relative to the odds of an infant's vocabulary including words for objects that the infant did not manipulate and mother did not name. Results show the odds as well as the corresponding probability. Notably, the odds (and probability) of any given word appearing in the vocabularies or in the spontaneous speech of 1-to-2-year-olds is very small (Frank et al., 2021). Therefore, the effects of infant object manipulation and mother naming is best interpreted relative to the odds of the same word appearing in infants' vocabularies in the absence of manipulation and naming. Notably, we were cautious to interpret findings in *relative* rather than absolute terms.

Findings in this sample were extended to word trajectories from Wordbank in RQ3. Pearson correlation coefficients quantified bi-variate associations between word age of acquisition from Wordbank and i) the number of infants who manipulated the word referent, and ii) the number of mothers who produced the word. We applied a square root transformation to the number of infants and mothers who manipulated and named objects, respectively, to meet assumptions of linearity for Pearson correlations, even though similar correlation coefficients were obtained with and without transformations. Then, we used a multiple linear regression to predict Wordbank age of acquisition for each word from the number of infants in the sample who manipulated the word referent (out of 32) and the number of mothers who produced the word during the visit (out of 32).

All analyses were conducted in R, version 4.1.0. Alpha was set to 0.05 for analyses and model assumptions (e.g., linearity, no outliers, no multicollinearity, random normally distributed residuals) were validated for all models.

Results

Prior to testing whether infants' object play cascades to mothers' object naming, words in infants' vocabularies, and infant real-time word production, we sought to confirm that the data contained sufficient instances of the variables of interest—namely that infants actually played with the objects on the MB-CDI subset of words; mothers named those objects; infants produced the names for target objects during the 2-hour-long visits; and infants had acquired those words (by mother's report).

Aggregating data across the sample showed that many of items on the MB-CDI subset of words were represented in infants' object play, mothers and infants' object naming, and infant vocabularies. Of the 245 words/objects, 180 objects (73%) were manipulated by at least one infant; 233 objects (95%) were named by least one mother; 125 objects (51%) were named by at least one infant during the visit; and 241 object names (98%) appeared in the vocabulary of at least one infant. In total, we identified 1,028 instances of infant play with the MB-CDI subset of objects; 2,236 instances of mother naming the MB-CDI subset of objects; 322 instances of infants producing the names for the MB-CDI subset of object words; and 2,542 word tokens in infant vocabularies.

Predicting Words in Infant Vocabularies

On average, an infant's vocabulary included the words for 85.72 out of 245 objects (*Mdn* = 80.00; *SD* = 69.15; *Range* = 0 – 241; 95% CI [61.76 – 109.68]). As a first step to testing the full model (from infant play to vocabulary through social input), we examined associations between key variables to ensure that i) infant object play and ii) mother real-time language *separately* increased the odds of infants having target words in their vocabulary; and iii) infant object play and mother naming were associated. All bivariate associations were significant.

Infant play and infant vocabulary

Figure 2A depicts a 2x2 contingency table that crosses words in an infant's vocabulary (yes-no) with objects manipulated by the infant during the visit (yes-no). Analyses revealed that the odds of an infant's vocabulary including words for manipulated objects (16.81/15.31) was 2.63 times the odds of an infant's vocabulary including words for non-manipulated objects (62.62/150.25), (OR = 2.63, 95% CI [1.24 - 5.59]).

Mother naming and infant vocabulary

We next tested the association between the words in an infant's vocabulary (yes-no) and the words that mothers produced during the visit (yes-no) again in a 2x2 contingency table (Figure 2B). The odds of an infant's vocabulary including words for the objects that mother named (36.12/33.75) was 3.26 times the odds of the infant's vocabulary including words for the objects that mother did not name (43.31/131.81), (OR = 3.26, 95% CI [1.82 - 5.83]).

Infant play and mother naming

Finally, we crossed the words mothers produced during the visit (yes-no) with the objects infants manipulated (yes-no) in a 2x2 contingency table (Figure 2C). The odds of a mother naming objects that the infant manipulated (21.94/10.18) was 7.41 times the odds of her naming a non-manipulated objects (47.94/164.94), (OR = 7.41, 95% CI [3.30 - 16.67]).

A

Infant vocabulary (by MCDI)	Infant object play: yes	Infant object play: no
yes	16.81	62.62
no	15.31	150.25
Odds	*1.10*	*0.42*

Odds ratio *1.10/0.42 = 2.63*

B

Infant vocabulary (by MCDI)	Mother naming: yes	Mother naming: no
yes	36.12	43.31
no	33.75	131.81
Odds	*1.07*	*0.33*

Odds ratio *1.07/0.33 = 3.26*

C

Mother naming	Infant object play: yes	Infant object play: no
yes	21.94	47.94
no	10.18	164.94
Odds	*2.15*	*0.29*

Odds ratio *2.15/0.29 = 7.41*

Figure 2. **(A)** Mean number of object words (out of 245) included (or not) in the infant's vocabularies whose referents were manipulated versus not manipulated by the infant. **(B)** Mean number of object words (out of 245) included (or not) in the infant's vocabularies that mothers did versus did not produce. **(C)** Mean number of object words (out of 245) produced (or not) by the mother whose referents were manipulated versus not manipulated by the infant.

Full model: from infant play and mother naming to infant vocabulary

Because analyses yielded significant associations among the variables of infant play, mother naming, and words in infants' vocabularies, we were able to formally test the full model. We submitted infants having individual object names in their vocabularies (based on MB-CDI) to a logistic mixed regression (see Analysis plan). The RQ1 model predicted the odds of an infant having the word for an object in their vocabulary from random and fixed effects. The fit of the final model for RQ1 (i.e., with two random intercepts, one random slope, and two fixed effects, Table 1) was significantly better than that of a null model, which only included random intercepts for infants, $\chi^2(5)$ = 1,781, p < .001 (medium marginal and conditional R-squared for RQ1 model was 2% and 77%, respectively). Random intercepts were specified for infants in the model to account for variation among infants in baseline vocabulary sizes, $\chi^2(3)$ = 3,247.7, *p* < .001. Random intercepts for words/objects accounted for variation in words appearing in infants' vocabularies, $\chi^2(1)$ = 1,083.2, *p* < .001. And random slopes for mother naming accounted for differences among dyads in the contributions of mother naming for infant vocabulary, $\chi^2(2)$ = 7.83, *p* = .020. Fixed effects for infant play and mother naming improved model fit (*p*'s < .001) and were included in the final model.

Table 1 shows RQ1 model results. Infant object play (exp(b) = 2.13, 95% CI [1.66 - 2.73], *p* < .001) and mother naming (exp(b) = 2.18, 95% CI [1.68 - 2.83], *p* < .001) increased the odds of words appearing in infants' vocabularies, as expected. Natural exponential functions can be applied to model coefficients (b's in Table 1) for ease of interpretation, namely by yielding odds ratios (exp(b) in Table 1). Notably, model fit improved when both fixed effects were considered jointly ($\chi^2(2)$ = 55.89, *p* < .001) but the interaction between infant object play and mother naming was not significant (exp(b)=0.96, *p* = .867) and it did not improve model fit relative to the RQ1 model ($\chi^2(1)$ = 0.03, *p* = .868). Therefore, infants are more like to have the words for objects they manipulate and mothers name (even if only one of these conditions occur).

Table 1 Logistic Mixed Regression Model for RQ1 to Predict Infant Vocabularies

Predictor	b	SE	Exp(b)	95% CI	*p*-value
Intercept	-2.11	0.49	0.12	0.04 – 0.32	<.001
Infant object play (0=no, 1=yes)	0.76	0.13	2.13	1.66 – 2.73	<.001

Mother naming (0=no, 1=yes)	0.78	0.13	2.18	1.68 – 2.83	<.001

Note. Model specification of fixed and *random* effects: log(in_vocabulary/1 – in_vocabulary) ~ obj_play + mother_naming + *(1+mother_naming|id)* + *(1|object)*. Beta (b) coefficients are in the logit scale and thus the natural exponential function was applied to yield odds ratio (Exp(b)), which are more interpretable.

Figure 3A depicts odds ratios for words appearing in infant vocabularies given that the infant manipulated the object and the mother named the object. As shown, the odds of infant vocabularies including words for manipulated objects (odds = 0.26, which correspond to a probability of 0.20) was 2.13 times the odds of including non-manipulated objects after adjusting for mother object naming (odds = 0.12, which correspond to a probability of 0.11). Similarly, the odds of infant vocabularies including words for objects that mothers named (odds = 0.26, which correspond to a probability of 0.21) was 2.18 times the odds of including words that the mother did not name after adjusting for infant object play (odds = 0.12, which correspond to a probability of 0.11). Notably, odds ratios for infants having specific words in their vocabularies were highest under the joint condition of infants manipulating the object *and* it being named by the mother (i.e., infant object play and mother naming equaled 1 in the model equation). Specifically, the odds that infants had words for the objects they manipulated that were also named by their mothers (odds = 0.56, which correspond to a probability of 0.36) was 4.64 times the odds of having words for non-manipulated and non-named objects (odds = 0.12, which correspond to a probability of 0.11).

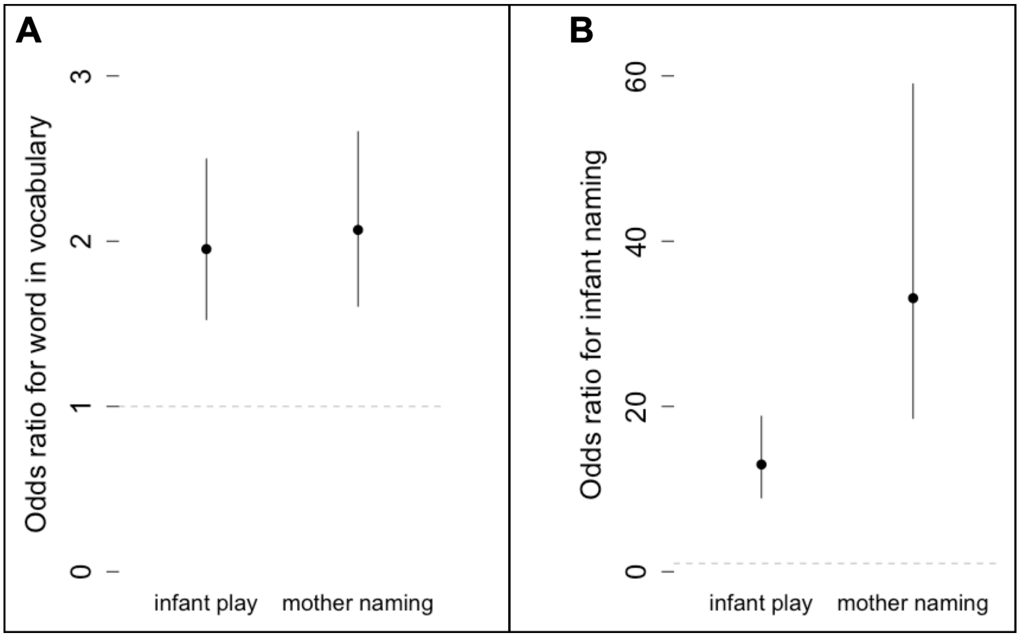

Figure 3. Multiplicative increase in the odds of words appearing in **(A)** infants' vocabularies and **(B)** in infant spontaneous speech for words referring to objects that infants manipulated and/or mothers named relative to the odds of words referring to objects that infants did not manipulate and mothers did not name. Error bars depict 95% confidence intervals derived from logistic mixed regression for infant play and mother naming only. Dashed lines mark no multiplicative increase (odds ratio = 1) and the y-axis changes for the dependent variables of infant vocabulary (A) vs. infant naming (B), with the odds ratio being strikingly high for the effects of infant play and mother naming on real-time infant word production.

Predicting Words in Infants' Spontaneous Speech

We next applied the cascade model to infants' spontaneous object naming at home. To what extent do infants produce the words for objects that they manipulate and/or mothers name? On average, infants named 10.06 objects (out of 245) during the visit (*Mdn* = 6; *SD* = 11.12; *Range* = 0-38, 95% CI [6.21 - 13.91]). As a first step to testing the full model (from infant play to mother naming to infant naming), we examined associations between key variables to ensure that i) infant object play, and ii) mother naming *separately* increased the odds of infant object naming (the association between infant play and mother naming was tested in RQ1). Bivariate associations were significant.

Infant play and infant naming

Figure 4A depicts a 2x2 contingency table that crosses objects named by the infant (yes-no) with objects manipulated by the infant during the visit (yes-no). Analyses revealed that the odds of an infant naming manipulated objects (6.88/25.24) was 17.92 times the odds of naming non-manipulated objects (3.19/209.69), (OR = 17.92, 95% CI [4.46 - 71.96]).

Mother naming and infant naming

We next tested the association between objects named by the infant (yes-no) and by the mother (yes-no) through a 2x2 contingency table (Figure 4B). The odds of an infant naming objects that mother named (9.63/60.25) was 63.45 times the odds of naming objects that mother did not name (0.44/174.69), (OR = 63.45, 95% CI [3.05 - 1,321]).

A

Infant naming	Infant object play: yes	Infant object play: no
yes	6.88	3.19
no	25.24	209.69
Odds	*0.27*	*0.015*
Odds ratio	*0.27/0.015 ≈ 17.92*	

B

Infant naming	Mother naming: yes	Mother naming: no
yes	9.63	0.44
no	60.25	174.69
Odds	*0.159*	*0.002*
Odds ratio	*0.159/0.002 ≈ 63.45*	

C

Mother naming	Infant object play: yes	Infant object play: no
yes	21.94	47.94
no	10.18	164.94
Odds	*2.15*	*0.29*
Odds ratio	*2.15/0.29 = 7.41*	

Figure 4. **(A)** Mean number of object words (out of 245) produced (or not) by the infant whose referents were manipulated versus not manipulated by the infant. **(B)** Mean number of object words (out of 245) produced (or not) by the infant that mothers did versus did not produce. **(C)** Mean number of object words (out of 245) produced (or not) by the mother whose referents were manipulated versus not manipulated by the infant.

Full model: from infant play and mother naming to infant naming

Because analyses yielded significant associations among the variables of infant play, mother naming, and infant naming, we were able to formally test the full model. We submitted infants' spontaneous naming of objects to a logistic mixed regression (see Analysis plan). The RQ2 model predicted the odds of an infant naming an object from random and fixed effects. The fit of the final model for RQ2 (i.e., with two random intercepts and two fixed effects, Table 2) was significantly better than that of a null model, which only included random intercepts for infants, $\chi^2(3) = 948.97$, p < .001 (medium marginal and conditional R-squared for RQ2 model was 37% and 71%, respectively). Random intercepts were specified for infants in the model to account for variation among infants in object naming, $\chi^2(1) = 246.1$, $p < .001$. And random intercepts for words/objects accounted for variation in words appearing in infants' spontaneous speech, $\chi^2(1) = 63.77$, $p < .001$. Random slopes were not included in the final model, they did not account for substantial variation in the outcome variable. Fixed effects for infant play and mother naming were included as they improved model fit (*p*'s < .001).

Table 2 shows RQ2 model results. Infant object play (exp(b) = 12.97, 95% CI [8.95 - 18.81], *p*<.001) and mother naming (exp(b) = 33.10, 95% CI [18.57 - 59.02], *p*<.001) increased the odds of words appearing in infants' spontaneous speech, as expected. Natural exponential functions can be applied to model coefficients (b's in Table 2) for ease of interpretation, namely by yielding odds ratios (exp(b) in Table 2). Notably, model fit improved when both fixed effects were considered jointly ($\chi^2(2) = 732.57$, $p < .001$), but the interaction between infant object play and mother naming was not significant (exp(b)=0.48, p = .216) and it did not improve model fit relative to the RQ2 model ($\chi^2(1) = 1.50$, p =.220). Therefore, both infant manipulation of objects and mother naming objects increased the odds of and infant producing the object name, even if the two did not occur jointly.

Table 2 Logistic Mixed Regression Model for RQ2 to Predict Infant Naming

Predictor	b	SE	Exp(b)	95% CI	*p*-value
Intercept	-7.78	0.46	0.0004	0.0001 – 0.001	<.001
Infant object play (0=no, 1=yes)	2.56	0.19	12.97	8.9466 – 18.809	<.001
Mother naming (0=no, 1=yes)	3.50	0.29	33.10	18.5690 – 59.020	<.001

Note. Model specification of fixed and *random* effects: log(in_infant_speech/1 – in_infant_speech) ~ obj_play + mother_naming + *(1|id)* + *(1|object)*. Beta (b) coefficients are in the logit scale and thus the natural exponential function was applied to yield odds ratio (Exp(b)), which are more interpretable.

Figure 3B depicts odds ratios for infants naming objects given that the infant manipulated and the mother named the object. As shown, the odds of infants naming manipulated objects (odds = 0.0054, which correspond to a probability of 0.0054) was 12.97 times the odds of naming non-manipulated objects after adjusting for mother naming (odds = 0.0004, which correspond to a probability of 0.0004). Similarly, the odds of infants naming objects that mothers named (odds = 0.0138, which correspond to a probability of 0.0136) was 33.10 times the odds of naming objects that the mother did not name after adjusting for infant object play (odds = 0.0004, which correspond to a probability of 0.0004). Notably, odds ratios for infants naming objects were highest under the joint condition of infants manipulating the object *and* it being named by the mother (i.e., infant object play and mother naming equaled 1 in the model equation). Specifically, the odds that infants named manipulated objects that mothers named (odds = 0.1793, which correspond to a probability of 0.1520) was 429.45 times (12.97*33.10) the odds that infants named objects that neither they manipulated nor the mother named (odds = 0.0004, which correspond to a probability of 0.0004). Note, the exceptionally high odds of infants producing an object name when infants both manipulated the object and mothers named the object was due to the near zero likelihood of infants producing words in the absence of mother naming and infant object touch.

Exploratory analyses: Does manual and visual engagement hold unique potential beyond only visual engagement?

Our cascade model highlights the importance of infants' multimodal, active engagement with the environment for the words they learn and produce. When infants manipulate objects, the object in hand dominates the visual field (Smith, Yu & Pereira, 2011; Yoshida & Smith, 2008; Yu et al., 2009- see Figure 1), and accordingly, infants should be more likely to name objects that they actively manipulate than those they do not. Alternatively, infants may be just as likely to name visibly present objects in the absence of manipulation. We thus sought to examine the sensory input commonly available during infant naming moments (although we recognize that our data do not allow us to directly disentangle the importance visual-only from visual-plus-manual engagement for infant word production).

Figure 5A shows the top 50 words that infants named, which largely map to manipulatable objects including stuffed animals (e.g., dogs), toys (e.g., balls), everyday objects (e.g., cups), and food (e.g., apples). Exploratory coding classified the available sensory information during naming instances by the infant. We tested for differences in the number of objects that infants named in five possible contexts: visually available in infants' hands, visually available in mothers' hands, visually available but neither manipulated by infant or mother, only auditorily-available, or displaced.

A one-way repeated-samples ANOVA and its non-parametric counterpart (i.e., Friedman's test), yielded similar results. Infants were more likely to produce the words for objects in contexts that involved visual-plus-manual sensory information than in any other context of sensory information (Figure 5B), $F(4,124) = 20.037$, $p < .001$, $\eta_p^2 = .393$ (large size effect), 95% CI [0.25 – 0.50] (all post-hoc tests against infants' hands had *p*'s with Bonferroni correction <.001). Indeed, infants tended to name objects that they manipulated (*M* = 6.88, *SD* = 1.43, 95% CI [6.38 – 7.38]) rather than name objects that were visually available but not manipulated by them (i.e., in mother hands + visual only, *M* = 1.50*, SD* = 2.11, 95% CI [0.77 – 2.23])*), t*(31) = 4.33, *p*<.001, Cohen's *d* = .766 (large effect size), 95% CI [0.37 – 1.18]). Post-hoc tests with Bonferroni corrections revealed no differences between the remaining categories of sensory input (all *p*'s > .177). However, when a named object was not in the infants' hands, it was more likely to be perceptually available in some form (i.e., in mother hands', visually available, or named by mother), *M* = 2.12, *SD* = 2.54, 95% CI [1.24 – 3.00]) than to be absent (*M* = 1.06, *SD* = 1.37, 95% CI [0.59 – 1.53]), *t*(31) = 3.14, *p* = .004, Cohen's *d* = .554 (medium effect size), 95% CI [0.18 – 0.94].

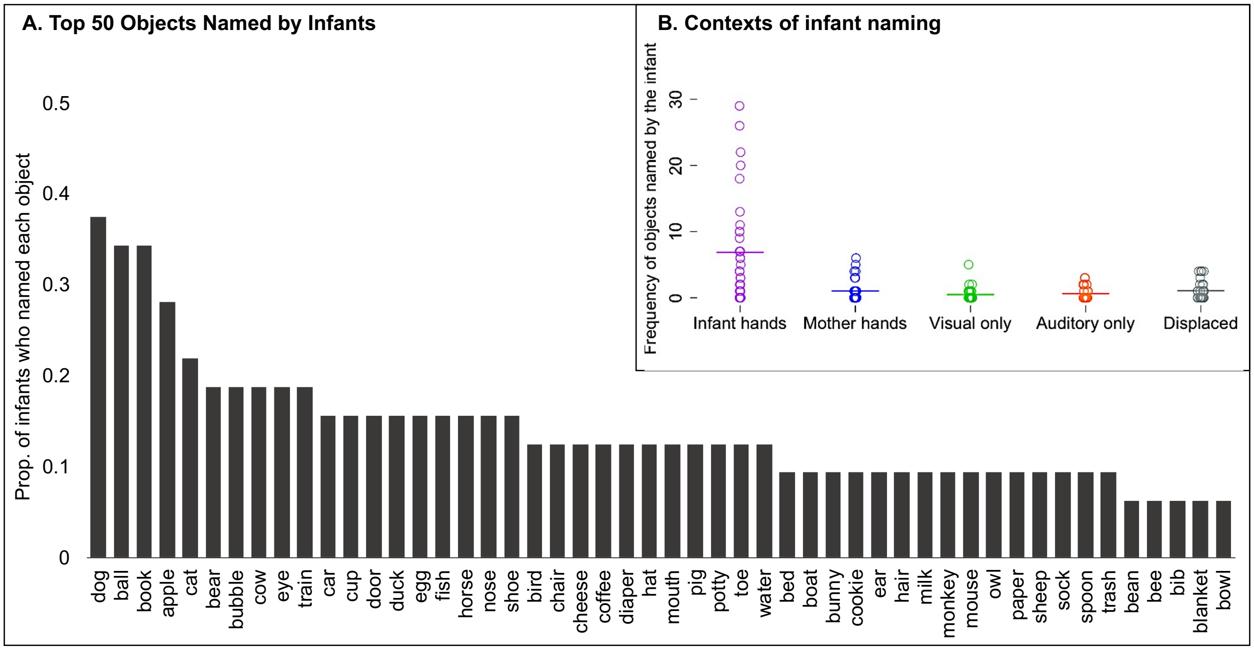

Figure 5. **(A)** Top 50 objects (out of 245) named by infants (ranked by the proportion of infants who named each object). **(B)** Number of objects named by infants with different types of available sensory information (horizontal lines represent the mean number of objects named per context).

Predicting the Age of Acquisition of Words

Results thus far support a cascade model from infant play to the words infants produce and have in their vocabularies through social input at home: Infants were more likely to know and produce the words for the objects they manipulated and that mothers named. Furthermore, exploratory analyses verified that visual plus manual engagement, rather than visual input alone, may be a critical component of early infant object naming. Thus, by extension, infants should acquire the words for objects they manipulate earlier in development than the words for objects that they do not manipulate (RQ3). To test this prediction, we related data on object play in this sample to the average ages of acquisition of the 245 MB-CDI subset words (*M* = 23.75 months, *Mdn* = 23.56, *SD* = 4.50, *Range* = 6.94 – 43.17, 95% CI [22.19 – 25.31]) drawn from over 5,000 infants represented in Wordbank (Frank et al., 2017).

Consistent with an embodied perspective of language development, objects that infants commonly manipulated in this sample mapped to words acquired at earlier ages by infants represented in Wordbank. As predicted, the number of infants who manipulated the word referent during the observation showed a strong and inverse correlation with the age of acquisition of specific words (*r* = -0.42, 95% CI [-0.52, -0.31], *p*<.001, *N* = 245, Figure 6A). Frequently manipulated objects by infants in the sample (e.g., book and cup) tended to map to words acquired early in development.

Likewise, the age of acquisition of words based on Wordbank was strongly associated with mother naming. Specifically, frequently named objects by mothers in the sample (e.g., book and dog) tended to be words that infants acquired early in development, as indicated in a negative association between age of acquisition and the number of mothers who produced the word during the observation (*r* = -0.72, 95% CI [-0.77, -0.65], *p*<.001, *N* = 245, Figure 6B).

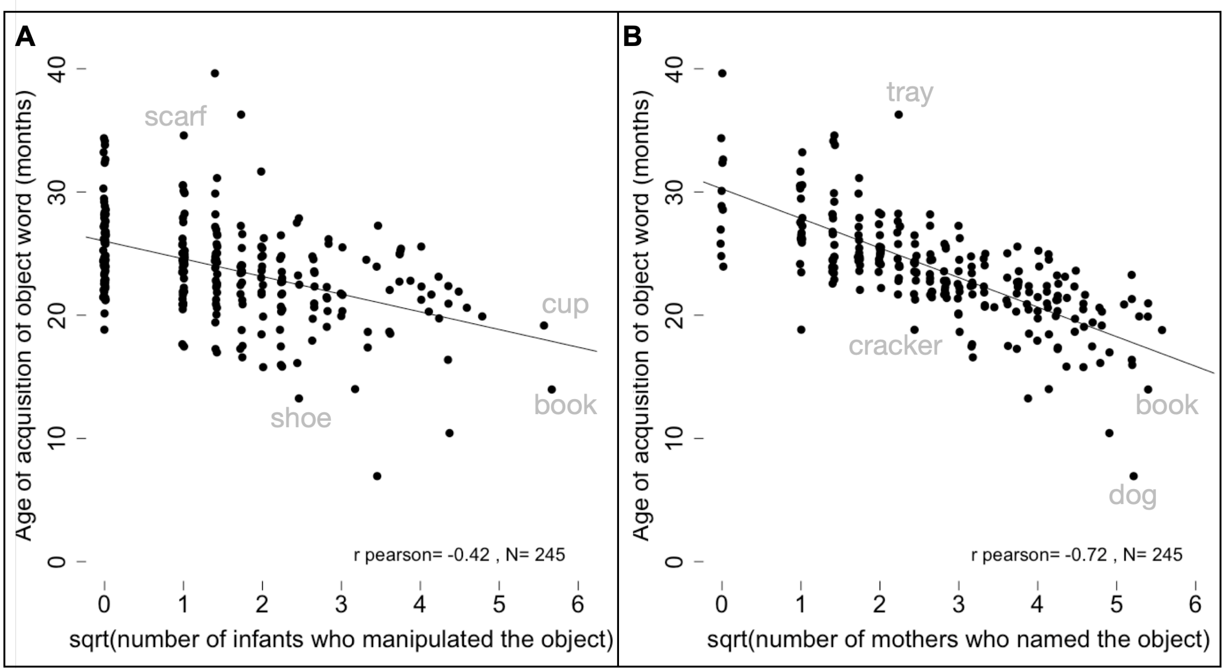

Figure 6. Each dot represents an object in the 245 MB-CDI subset; age of acquisition of each object word is plotted as a function of the **(A)** square root of the number of infants who manipulated the word's referent, and **(B)** square root of the number of mothers who produced the word.

Multiple linear regression examined the joint association of infant play and mother naming with the age of acquisition of words (see Analysis plan for RQ3). Table 3 presents results. Mother naming significantly predicted the age of acquisition of specific words, after adjusting for infant play ($F(2,242) = 103.2$, $p < .001$, Adj $R^2 = 45.6\%$), such that the more mothers who named the object, the younger the age of word acquisition. Alternative models that included infant object manipulation alone (b = -0.31, $p < .001$), and mother naming alone (b = -0.40, $p < .001$) revealed that both significantly predicted age of acquisition of words. Model fit improved when adding mother naming to infant manipulation ($\chi^2(1) = 1,552.2$, $p < .001$) but not when adding infant manipulation to mother naming ($\chi^2(1) = 9.07$, $p = .364$). The interaction between infant object play and mother naming was not significant (b = 0.003, $p = .487$) and did not improve model fit ($\chi^2(1) = 5.35$, $p = .486$).

Table 3 Multiple Linear Regression Model for RQ3 to Predict Words' Age of Acquisition

Predictor	b	SE	t-val	95% CI	*p*-value
Intercept	27.44	0.33	82.23	26.78 – 28.10	<.001
Infants who manipulated object	-0.04	0.04	-0.91	-0.13 – 0.05	.365
Mothers who named object	-0.39	0.03	-11.88	-0.45 – -0.32	<.001

Discussion

Infants' actions with objects are the glue that binds words to the objects of everyday life. Through frame-by-frame coding of extended naturalistic recordings, we built on lab-based studies by showing that everyday self-directed action corresponds with contingent caregiver input that cascades to the words that infants have in their vocabularies and produce in real time. Three key findings converge on this claim: (1) Words for objects manipulated by infants and objects named by mothers during natural recordings were more likely to appear in infants' vocabularies and spontaneous speech relative to words for non-manipulated objects and objects that mothers did not name; (2) infants were more likely to name the objects that they actively manipulated and not the objects that were visibly present in the absence of their manipulation; and (3) commonly manipulated and named objects in the sample mapped to words acquired earlier in development based on data in Wordbank. In fact, infant object play and caregiver naming during the 2-hour long observations accounted statistically (although not causally) for 45% of the variance in the age of acquisition of words (RQ3). Our approach complements traditional measures for the words in infant vocabularies (i.e., vocabulary checklists) by measuring infants' own production of object names in the home environment. We discuss implications of findings for the active infant; frequency effects; mechanisms of cascading action; and infant word production.

Active Object Selection by the Infant

The role of active learning and embodied cognition cuts across learning domains throughout development (Glenberg, Brown & Levin, 2007; Jant et al., 2014; Kim, Roth & Thom, 2011; Needham & Libertus, 2011; Needham, 2000; Soska, Adolph & Johnson, 2010). The present work supports an embodied perspective of language development, in which infants actively construct their experiences as they engage with objects and people in their environments. Of course, infants can only manipulate objects that are available in the home environment, and vocabulary growth and word production certainly depend on such availabilities. Nonetheless, infants selectively narrow in on a subset of objects from the much larger set of objects that are available to manipulate, and their active exploration appears to catalyze word learning.

Likewise, other word features may operate together with infant play to guide the timing of word learning–such as phonetics, iconicity, and contextual distinctiveness (Schneider, Yurovsky & Frank, 2015; Swingley & Humphrey 2018; Perry, Perlman & Lupyan, 2015; Roy et al., 2015). Presumably infants do not select objects based on whether their names are phonetically simple, highly iconic, and contextually distinctive. Instead, infants' interest and body mechanics jointly guided infant object selection from the environment in real-time across the two-hour window that we examined.

Beyond a Frequency Effect

Results align with research showing that word frequency predicts age of acquisition for a word (Braginsky et al., 2016; Goodman, Dale & Li, 2008; Schneider, Yurovsky & Frank, 2015; Swingley & Humphrey, 2018; Roy et al., 2015) and infants learn the words that caregivers produce (Hart & Risley, 2003). But our results uniquely implicate infants as active agents who engage with the referents of the words they learn. Indeed, words that map to manipulatable objects are acquired earlier in development (Nelson, 1973) compared to frequent words that do not have concrete referents such as articles, pronouns, and so on (Fenson et al., 1994; Frank et al., 2017). Such findings reinforce the claim that infants must actively engage with the world to learn the words for objects and actions.

As predicted, infant object play significantly predicted the specific words of infants' vocabularies and production after controlling for mother naming. Consistent with a cascade model, infants are more likely to learn the words for the objects they manipulate, and the likelihood of their vocabularies containing specific words increases further when mothers name manipulated objects. Object play helps infants to develop strong object representations, which may then easily map to words (Slone, Smith & Yu, 2019). Likewise, infants extend their conceptual abilities (e.g., learning to construct and deconstruct objects) through object exploration, which are foundational for language and cognitive development (Bornstein, Hahn & Suwalsky, 2013; Iverson, 2010).

Mechanism of Cascading Action

Under the embodied learning hypothesis—which proposes a real-time cascade from infant play to mother naming to word learning—why does infant object play predict infant word learning? Infant's object play filters incoming sensory input by centering the held object and expanding it in view (Smith & Yu, 2008). Even though we measured infants' manual engagement with objects, we posit that infants' active engagement through eyes and hands supports learning (rather than interpreting findings through a hands-only view). Indeed, infants look at the objects they touch (Ruff, 1986), and so infants' object interactions disambiguate word meaning by aligning the dominant object in the visual field with the timing of caregiver naming (Trueswell et al., 2016). Furthermore, rich and clear multimodal experiences (i.e., visual, tactile, auditory) allow infants to correctly map the words they hear to the objects they touch (Pereira, Smith & Yu, 2014; Yu & Smith, 2012). Infants who experience referential transparency during naming moments will grow their vocabularies rapidly (Cartmill et al, 2013). Additionally, infants' own engagement with objects leads to triadic engagement as caregiver jointly engage with the same object (Suarez-Rivera et al., 2022). Of course, sometimes infants may manipulate objects in response to caregiver naming; however, research on contingency and temporal features of mother-infant interactions suggest that mothers tend to follow-in on infant actions with labelling utterances (Tamis-LeMonda, Kuchirko & Song, 2014).

Several applications can further test the proposed cascading mechanism. Computational models of word learning that distinguish names for objects in the learner's active world may accurately simulate infant learning. Likewise, right-skewed, power-law distributions of infants' natural experiences (Clerkin et al., 2017; Smith et al., 2018) may build a curriculum for infant learning and could critically support learning from actions in artificial intelligent systems (Smith & Gasser, 2005). Finally, although we focused on common everyday nouns, acquiring verbs likely follow the same laws of embodiment: Infants are exposed to the words for "walk" and "jump" at the precise moments when they walk and jump just as they hear words such as "push" and "pull" during moments when they push and pull (West et al., in press).

Infants Produce the Words for the Objects of Play

Infants' manual actions corresponded with the words they produced. Of course, visual attention to the objects of touch may drive naming more than manipulation per se. However, the finding that infants were more likely to name what they manipulated than name visually available but not-manipulated objects suggests that infants' multimodal, active engagement with the environment is critical for infant object naming.

Findings also support the hypothesis that infant object naming may critically provide the *inputs* for word learning. Specifically, infants may initiate real-time cascades by naming the objects they manipulate, thereby strengthening word-object mappings in a self-socializing process (Halim et al., 2018). Moreover, while coding the contexts around infant naming, we observed (though did not formally test) that mothers frequently responded when infants produced object names themselves. Thus, infant naming gave mothers an opportunity to establish conversations and embellish on infant talk by saying “I like dogs too” after the infant produced the word dog.

Practical Implications

Findings have practical implications for educators, clinicians, and parents. Infants are active learners who acquire knowledge by interacting with the objects, spaces, and people in their world. Exposure to a wide variety of objects at home (toys and non-toys alike) allows infants to acquire the words for common everyday objects including boxes, pots, pans, and clothes. However, object manipulation is neither necessary not sufficient for language development and alternative developmental pathways that support word learning (Iverson, 2010) should be investigated. For example, infants with Type 2 spinal muscular atrophy experience motor difficulties but demonstrate typical communicative development (Sieratzi & Woll, 2002). Finally, our findings suggest that infants are most likely to learn the words that their caregivers produce. Caregivers can turn everyday routines, such as dressing and feeding, into optimal contexts for word learning by talking about the objects of everyday life (Tamis-LeMonda et al., 2019).

Limitations

Our approach to mapping infants' experiences to the words in their vocabularies from 2-hour long naturalistic recordings is far from perfect. Certainly, our observations do not capture all the objects that infants encounter in play or in mothers' speech. Ideally, longer video-recordings would sample a wider range of the objects that infants manipulate and words they hear daily. However, although technologies such as LENA (lena.org) allow for 12+ hours of continuous audio recordings, the absence of video prevents coding of infant object play and activities. Moreover, feasibility is a limiting factor, as each hour of video requires up to 10 hours of transcription time. Thus, we struck a balance between lengthy observations and our capacity for transcribing and coding. Certainly, 2-hour time windows improve on brief 5- to 15-minute observations typically used in infancy studies and allowed for testing associations among infant object play, mother naming, and infant naming with sufficient instances of each. Our decision to micro-code the behaviors of individual infants at home allowed us to test how infants' unique experiences played out in their own word learning trajectories. Furthermore, associations between the data we gathered during the 2-hour-long visits and the words that appear in the productive vocabularies of thousands of children from other households (Wordbank data) provide converging support for our embodied hypothesis.

Finally, although the odds of a word appearing in the infant's vocabulary doubles if the infant manipulated the word referent (or if mother named the word), the odds itself was small in absolute terms for the infants we observed (who were under 2 years). However, small vocabularies are expected for infants this age, as are small odds for words to appear in young infants' vocabularies (Frank et al., 2021). Research with older infants might better quantify the gains for word learning during object play and mother naming at other points in development.

Conclusion

We advanced lab-based research on infant word learning by venturing into the home environment to rigorously test an embodied learning hypothesis. Detailed behavioral coding, coupled with time-locked transcriptions, revealed that infants' spontaneous manipulation of objects corresponds to the words that mothers direct to their infants and maps to infants' vocabularies and own production of object labels. Findings spotlight cascading processes that operate in the real world to support early language development.

References

Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language*, *18*(1), 41-49. https://doi.org/10.1017/S0305000900013283

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, *29*(5), 832-843. doi:10.1037/0012-1649.29.5.832

Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. *Child Development*, 381-398. https://doi.org/1130984

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bornstein, M. H., Hahn, C. S., & Suwalsky, J. T. (2013). Physically developed and exploratory young infants contribute to their own long-term academic achievement. *Psychological Science*, 24(10), 1906-1917. https://doi.org/10.1177/0956797613479974

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. (2016). From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. In *CogSci*. https://www.cmu.edu/dietrich/psychology/callab/papers/bymf-cogsci-2016.pdf

Bruner, J. S. (1974). From communication to language—A psychological perspective. *Cognition*, *3*(3), 255-287. https://doi.org/10.1016/0010-0277(74)90012-2

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278-11283. https://doi.org/10.1073/pnas.1309518110

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy*, *25*(6), 809-826. https://doi.org/10.1111/infa.12361

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs for the Society for Research in Child Development*, 59(5, Serial No. 242). https://doi.org/1166093

Fenson, L., Marchman, V. A., Thal, D., Dale, P., Reznick, J. S. & Bates, E. (2007). MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual. 2nd Edition. Baltimore, MD: Brookes Publishing Co.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16, 234–248. http://doi.org/10.1111/desc.12019

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*(5), 578-585. https://doi.org/10.1111/j.1467-9280.2009.02335.x

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677. https://doi.org/10.1017/S0305000916000209

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project*. Cambridge, Massachusetts: MIT Press.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, *39*(1), 1-42. https://www.annualreviews.org/doi/pdf/10.1146/annurev.ps.39.020188.000245

Glenberg, A. M., Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology*, *32*(3), 389-399. https://doi.org/10.1016/j.cedpsych.2006.03.001

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515-531. https://doi.org/10.1017/S0305000907008641

Halim, M. L. D., Walsh, A. S., Tamis-LeMonda, C. S., Zosuls, K. M., & Ruble, D. N. (2018). The roles of self-Socialization and parent socialization in toddlers' gender-typed appearance. *Archives of Sexual Behavior*, *47*(8), 2277-2285. https://doi.org/10.1007/s10508-018-1263-y

Hart, B., & Risley, T. R. (2003). The 30 million word gap by age 3. *American Educator, Spring*.

Hirsh-Pasek, K., & Golinkoff, R. M. (2012). How babies talk: Six principles of early language development. In S. L. Odom, E. P. Pungello, & N. Gardner-Neblett (Eds.), *Infants, Toddlers, and Families in Poverty: Research Implications for Early Child Care* (pp. 77–100). New York, NY, US: The Guilford Press.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, 37(2), 229-261. https://doi.org/10.1017/S0305000909990432

Jant, E. A., Haden, C. A., Uttal, D. H., & Babcock, E. (2014). Conversation and object manipulation influence children's learning in a museum. *Child Development*, *85*(5), 2029-2045. https://doi.org/10.1111/cdev.12252

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, *5*(2), 219-232. https://doi.org/10.1111/1467-7687.00224

Kim, M., Roth, W. M., & Thom, J. (2011). Children's gestures and the embodied knowledge of geometry. *International Journal of Science and Mathematics Education*, *9*(1), 207-238. https://doi.org/10.1007/s10763-010-9240-5

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101. https://doi.org/10.1073/pnas.1532872100

Lockman, J. J. (2000). A perception–action perspective on tool use development. *Child Development*, *71*(1), 137-144. https://doi.org/10.1111/1467-8624.00127

MacWhinney, B. (2000). *The CHILDES project: The database (Vol. 2)*. Psychology Press.

Markman, E. M. (1987). How children constrain the possible meanings of words. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (p. 255–287). Cambridge, United Kingdom: Cambridge University Press.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, United Kingdom: MIT Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121-157. https://doi.org/10.1016/0010-0285(88)90017-5

Needham, A. (2000). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development*, *1*(2), 131-156. https://doi.org/10.1207/S15327647JCD010201

Needham, A., & Libertus, K. (2011). Embodiment in early development. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(1), 117-123. https://doi.org/10.1002/wcs.109

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, 1-135. https://doi.org/1165788

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178-185. https://doi.org/10.3758/s13423-013-0466-4

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PloS one*, *10*(9), e0137147. https://doi.org/10.1371/journal.pone.0137147

Piaget, J. (1969). *The Mechanisms of Perception*. Oxon, Great Britain: Routledge.

Pruden, S. M., Roseberry, S., Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Infant categorization of path relations during dynamic events. *Child*

Development, *84*(1), 331-345. https://doi.org/10.1111/j.1467-8624.2012.01843.x

Quine, W. V. O. (1960). *Word and Object. New Edition.* Cambridge: Massachusetts.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rochat, P. (1989). Object manipulation and exploration in 2-to 5-month-old infants. *Developmental Psychology*, *25*(6), 871-884. doi:10.1037/0012-1649.25.6.871

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*(3), 956-970. https://doi.org/10.1111/cdev.12166

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

Ruff, H. A. (1986). Components of attention during infants' manipulative exploration. *Child Development*, 105-114. https://doi.org/10.2307/1130642

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology*, *26*(1), 85-93. https://doi.org/10.1037/0012-1649.26.1.85

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926

Schneider, R., Yurovsky, D., & Frank, M. (2015). Large-scale investigations of variability in children's first words. In *CogSci* (pp. 2110-2115). https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.699.8481&rep=rep1&type=pdf

Sieratzki, J. S., & Woll, B. (2002). Toddling into language: precocious language development in motor-impaired children with spinal muscular atrophy. *Lingua*, 112(6), 423-433. https://doi.org/10.1016/S0024-3841(01)00054-7

Slone, L. K., Smith, L. B., & Yu, C. (2019). Self-generated variability in object images predicts vocabulary growth. *Developmental Science*, *22*(6), e12816. https://doi.org/10.1111/desc.12816

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, *22*(4), 325-336.

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9-17. https://doi.org/10.1111/j.1467-7687.2009.00947.x

Smith, L., & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life*, *11*(1-2), 13-29. doi: 10.1162/1064546053278973.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558-1568. https://doi.org/10.1016/j.cognition.2007.06.010

Song, L., Spier, E. T., & Tamis-Lemonda, C. S. (2014). Reciprocal influences between maternal language and children's language and cognitive development in lowincome families. *Journal of Child Language, 41*(2), 305-326. https://doi.org/10.1017/S0305000912000700

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*, *46*(1), 129-138. doi:10.1037/a0014618

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/infa.12446

Suarez Rivera, C. & Tamis-LeMonda, C. (2021). Data and Materials. From Play to Language: Infants' Actions on Objects Cascade to Word Learning. *Databrary*. Retrieved March 18, 2022 from http://doi.org/10.17910/b7.1358.

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific English words. *Child Development*, *89*(4), 1247-1267. https://doi.org/10.1111/cdev.12731

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*(3), 202-209. 10.1109/TAMD.2013.2269905

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*(2), 121-126. https://doi.org/10.1177/0963721414522813

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

Tamis-LeMonda, C. S., Suarez-Rivera, C., West, K. L., & Ossmy, O. (2020). The food for thought: Unpacking infants' everyday language experiences. Presented virtually at the 2020 Biennal International Congress on Infant Studies in Glasgow, Scotland.

Tamis-LeMonda, C. & Adolph, K. (2017). Videos. The Science of Everyday Play. *Databrary*. Retrieved March 18, 2022 from http://doi.org/10.17910/b7.563.

Thal, D. J., Marchman, V. A. & Tomblin, J. B. (2013). Late talking toddlers: Characterization and prediction of continued delay. In L. Rescorla & P. Dale (Eds.). Late Talkers: Language Development, Interventions, and Outcomes. Baltimore, MD.: Brookes Publishing.

Thelen, E., & Smith, L. B. (1998). Dynamic systems theories. In W. Damon & R. M. Lerner (Eds.), *Handbook of Child Psychology: Theoretical Models of Human Development* (pp. 563–634). Hoboken, NJ, US: John Wiley & Sons Inc.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, *4*(12), 197-211. https://doi.org/10.1177/014272378300401202

Trueswell, J. C., Lin, Y., Armstrong III, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, 148, 117-135. https://doi.org/10.1016/j.cognition.2015.11.002

Vygotsky, L. S. (1978). *Mind in society: The Development of Higher Mental Processes* (E. Rice, Ed. & Trans.). Cambridge, Massachusetts: Harvard University Press.

Werker, J. F. (1989). Becoming a native listener. *American Scientist*, *77*(1), 54-59. http://www.jstor.org/stable/27855552

West, K. L., & Iverson, J. M. (2017). Language learning is hands-on: Exploring links between infants' object manipulation and verbal input. *Cognitive Development*, *43*, 190-200. https://doi.org/10.1016/j.cogdev.2017.05.004

West, K.L., Fletcher, K.F., Adolph, K.E., & Tamis-LeMonda, C.S. (in press) Mothers talk about infant actions: How verbs correspond to infants' real-time behavior. *Developmental Psychology.*

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*(3), 229-248. https://doi.org/10.1080/15250000802004437

Yu, C., Smith, L. B., Shen, H., Pereira, A. F., & Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE transactions on autonomous mental development*, *1*(2), 141-151. doi: 10.1109/TAMD.2009.2031513.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165-180. https://doi.org/10.1111/j.1467-7687.2010.00958.x

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

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS One*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

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

Yu, H. T. (2015) Applying linear mixed effects models with crossed random effects to psycholinguistic data: Multilevel specification and model selection. *The Quantitative Methods for Psychology*, *11*(2), 78-88. doi: 10.20982/tqmp.11.2.p078

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Selected Nouns from the MB-CDI That Were Included in the Study

Appendix S2. Data Processing Guidelines for Grouping Words