

Big data about small people: The Play & Learning Across a Year (PLAY) Project

Rick O. Gilmore^{1,3}, Karen E. Adolph^{2,3}, Catherine L. Tamis-LeMonda², Kasey Soska³, & Joy
L. Kennedy^{2,3}

¹ The Pennsylvania State University

² New York University

³ Databrary.org

Author Note

Rick O. Gilmore is in the Department of Psychology, The Pennsylvania State University, University Park, PA 16802. Karen E. Adolph is in the Department of Psychology, New York University, 4 Washington Place, New York, NY 10003. Catherine L. Tamis-LeMonda is in the Department of Applied Psychology, New York University. Kasey Soska is is Scientific Project Director at Databrary. Joy L. Kennedy is Scientific Support Specialist at Databrary. We acknowledge support from the National Science Foundation (BCS-1238595), the Eunice Kennedy Shriver National Institute for Child Health and Human Development (U01-HD-076595), the Society for Research in Child Development, the Alfred P. Sloan Foundation, and the LEGO Foundation.

Correspondence concerning this article should be addressed to Rick O. Gilmore, Department of Psychology, University Park, PA 16802. E-mail: rogilmore@psu.edu

Abstract

Piaget, Montessori, and Bruner observed that play is the work of infants. The PLAY (Play & Learning Across a Year) project seeks to catalyze discovery about the form and dynamics of this essential work across a critical period from 12 to 24 months of age when infants show remarkable advances in language, object interaction, locomotion, and emotion regulation. PLAY will leverage the joint expertise of 65 “launch group” researchers and capitalize on the Databrary video-sharing library and Datavyu video-coding tool to exploit the power of video to reveal the richness and complexity of behavior. The PLAY researchers will collect, transcribe, code, share, and exploit a video corpus of infant and mother naturalistic activity in the home to test hypotheses about behavioral, developmental, and environmental cascades. In turn, the project will demonstrate the value and feasibility of a cross-domain synergistic approach to infant research while advancing new ways to use video as data and documentation to facilitate discovery and ensure transparency.

Keywords: keywords

Word count: X

Big data about small people: The Play & Learning Across a Year (PLAY) Project

Behavior lies at the core of developmental science (Gibson, 1994). It is a uniquely powerful tool for capturing the richness and complexity of behavior (K. E. Adolph, n.d.; Rick O Gilmore & Adolph, 2017), documenting its microstructure in real time and global patterns of change over development (Gesell, 1946, 1991). Video chronicles who did what, and how, when, and where they did it [Rick O Gilmore, Adolph, Millman, and Gordon (2016); R O Gilmore and Adolph (n.d.); Adolph2017-ac]. Most infancy researchers collect video as primary data or as a backup to online coding procedures, but until recently, few have openly shared video because of ethical and technical challenges. As those challenges recede and the culture of developmental science begins to embrace more open, transparent, and reproducible practices (???), the time is ripe to capitalize on the unique power of video to catalyze discovery and transform knowledge about behavioral development in infancy.

The Play & Learning Across a Year (PLAY) project (???; K. E. Adolph, Tamis-LeMonda, & Gilmore, 2018) builds on the NICHD/NSF-funded Databrary video-sharing library (“Databrary,” n.d.; R O Gilmore, Adolph, & Millman, 2016) and the Datavyu video-coding tool (K. E. Adolph, 2015; “Welcome! || datavyu,” n.d.) developed and supported by Adolph and Gilmore, and it unites the joint expertise of 65 PLAY “launch group” researchers in the United States and Canada. PLAY will create the first-of-its-kind, large-scale, openly shared, readily reusable, transcribed, coded, and curated video corpus of human behavior. In addition, PLAY seeks to advance new video-based means of research documentation that hold promise to increase transparency and bolster reproducibility (K. E. Adolph, Gilmore, & Kennedy, 2017, Rick O Gilmore and Adolph (2017)) across the behavioral sciences. In this paper, we describe the process of planning PLAY, our preliminary results from a small pilot study, and plans for a larger scale implementation we expect to launch in late 2018.

Project Planning

We named the project “PLAY” and use the terms “unstructured play” and “everyday play” to broadly refer to infants’ natural activities while awake. To paraphrase Piaget (Piaget, 1967), Montessori (Montessori, 1984), and Bruner (Jerome S Bruner, 1975; Jerome Seymour Bruner, 1976), play is the work of infants. It is an approach to action, not a particular form of activity. Some of infants’ play involves toys and some play is joyful and goal directed, but all of their spontaneous vocalizations, interactions with objects and people, and locomotor bouts involve exploration and opportunities for learning and growth, regardless of affect or intent.

Planning began in late 2015. Adolph, Tamis-LeMonda and Gilmore (PLAY PIs) invited researchers to join the launch group based on their interest in open science and infant-mother natural activity in the home, willingness to collaborate on data collection and coding, lab location, and domains of expertise (language, gesture, play, object exploration, tool use, locomotion, posture, physical activity, emotion, temperament, parent responsiveness, gender, home environment, media use, spatial demography, and sampling). Nearly every invitee agreed. The current launch group consists of 65 from 49 institutions and 24 states. 34% are new investigators, 68% women, 18 non-white, from varied institutions (public and private universities and colleges, hospitals, agencies) with varied resources (38 public universities, 18% R15-eligible institutions) across the United States and Canada.

Figure @ref{fig:play-site-map}

To distribute the burden of video coding across researchers, we recruited 10+ experts to shape the development of coding passes in four fundamental domains—communication, object interaction, locomotion, and emotion. These domains represent key areas of infant development and provide foundational information for future discovery when time-locked to video. Compared with other behaviors (e.g., visual attention), these are quick, easy, and reliable to code. Experts in language and communication transcribe all mother speech and infant vocalizations in formats exportable to CHILDES (MacWhinney, 2000a). These codes

should be informative even to non-experts and are designed to facilitate further discovery through subsequent coding passes that build on the prior work. Temporally aligned transcriptions and codes should enable researchers with expertise in any domain to analyze cascades within and among these behaviors.

Through a yearlong series of telephone conversations with each launch group member and 12 group webinars (K. E. Adolph, Tamis-LeMonda, & Gilmore, n.d.), we jointly developed a common sampling method and protocol (including materials, technical specifications, questionnaires, and non-video measures), designed common video codes, and established an infrastructure to divide responsibilities between PLAY staff and launch group. We achieved consensus, with input from NICHD program staff, about all aspects of PLAY at a daylong workshop at NIH in December 2016, materials from which are shared on Databrary (K. E. Adolph, Gilmore, & Tamis-LeMonda, 2016).

The launch group jointly decided that the centerpiece of PLAY would be 900 one-hour videos of infant-mother dyads during natural play in the home. Home videos are widely believed to be representative of natural activity, and provide a stark contrast to the 2- to 20-minute “snapshots” typical of standard structured lab tasks. Based on their extensive experience with naturalistic home observations (Barbaro, Johnson, Forster, & Deák, 2015; Fausey, Jayaraman, & Smith, 2016; Iverson & Wozniak, 2007; Karasik, Adolph, Tamis-LeMonda, & Zuckerman, 2012; Karasik, Tamis-LeMonda, & Adolph, 2011; Karasik, Tamis-LeMonda, & Adolph, 2014; Karasik, Tamis-LeMonda, Adolph, & Bornstein, 2015; Rowe, 2012; Soderstrom & Wittebolle, 2013), launch group members determined that one hour is sufficiently long to capture an ecologically valid window into infant and mother natural behaviors. Longer recording times produce diminishing returns, risk infants becoming excessively tired or hungry, and increase the cost and burden to families and researchers.

Given the cost of going to families’ homes, the launch group also determined that we should augment recordings of natural play with a set of additional video, questionnaire, and non-video measures, that together would add only ~45 minutes to the home visit. This

would enable researchers to test whether variations in natural play, or in characteristics of distal and proximal environments, predict infant and mother behaviors when materials and conditions are held constant. Thus, the solitary and dyadic play tasks are of interest in their own right, and might also serve as correlates in tests of experiential and environmental influences.

To obtain objective data on stable home conditions (cracks in walls, broken windows, ceiling stains, safety issues, etc.), physical layout (furniture, clutter, space to move, etc.), educational and electronic media (writing/drawing materials, TVs, computers, etc.), and gendered characteristics of infants' room, toys, and clothes, we decided to conduct video home tours. Launch group experts also expressed interest in understanding whether clothing and footwear affect infants' locomotion and physical activity, and each takes only moments to video record. Clothing/footgear videos can reveal gendered features (bows/frills, superhero emblems, patent leather shoes, army boots)⁶⁶ and influences on spontaneous activity and locomotion (Cole, Lingeman, & Adolph, 2012). The launch group also deemed important a variety of questionnaire measures of infant skills, experiences, and home environment: mothers' report of infants' vocabulary, locomotor milestones and falls, temperament, and use of gender labels; mother's report of family demographics, media use, health, and home chaos; and a researcher-completed survey on physical characteristics of the home. The PIs developed a custom tablet-based app to collect these questionnaire data efficiently, limit data input errors, use a stylus for flexible data entry, and allow automatic transfer to permanent storage on Databrary. The launch group devised methods to measure room size with a commercial laser device and ambient noise level with a commercial decibel meter. Finally, with input from the launch group, we decided to transcribe all mother speech and infant vocalizations in the CHAT format exportable to CHILDES (MacWhinney, 2000a). These codes should be informative even to non-experts and are designed to facilitate further discovery through subsequent coding passes that build on the prior work. Temporally aligned transcriptions and codes should enable researchers with expertise in any domain to

analyze cascades within and among these behaviors.

Pilot Study

Based on the launch group's recommendations, we carried out a pilot study in the New York City area to test the feasibility of the approach.

Methods

Participants. A total of $n = 20$ infants were tested, $n = 4$ 12-month-olds (3 female), $n = 12$ 18-month-olds (5 female), and $n = 4$ (0 female) 24-month-olds.

We chose these ages because 12- to 24-months represents a period of important, rapid growth when children begin talking, using objects in symbolic play, walking, and regulating emotions. For example, by 12 months, about half of infants can walk and half still crawl. By 18 months, infants are proficient walkers, and by 24 months, they can run, walk backwards, and walk up stairs (Onis, 2006; Robinson, 2015). Around 12 months, infants produce their first words. By 18 months, most display a vocabulary spurt, and by 24 months infants combine words into simple sentences (Bloom, 1995; Hoff, 2013). But these ages represent only group averages; individual infants show tremendous variability in these behaviors at each age.

All infants recruited were from the New York City area. Fifteen infants were White, one was Asian, two reported more than one race, and two did not report a race. Six were of Hispanic or Latino ethnicity.

Procedure. A publicly-accessible wiki (Soska, Adolph, Tamis-LeMonda, & Gilmore, 2016) was used to document all procedures and code definitions. The wiki links descriptions of every aspect of the protocol with exemplar third-person video clips (e.g., researcher scheduling home visit) to demonstrate typical procedures. The wiki links text-based descriptions of each code to video clips illustrating the types of behaviors that do and do not satisfy the coding criteria.

Data analysis. For this report, we used R (Version 3.4.4; R Core Team, 2017) and the R-packages *acs* (Version 2.1.3; Glenn, 2018), *bindrcpp* (Version 0.2; Müller, 2016), *choroplethr* (Lamstein, 2017; Version 3.6.1; Lamstein & Johnson, 2017), *choroplethrMaps* (Version 1.0.1; Lamstein, 2017), *databraryapi* (Version 0.1.0; ???), *dplyr* (Version 0.7.4; Wickham & Francois, 2016), *forcats* (Version 0.3.0; Wickham, 2018a), *ggplot2* (Version 2.2.1; Wickham, 2009), *httr* (Version 1.3.1; Wickham, 2017a), *jsonlite* (Version 1.5; Ooms, 2014), *papaja* (Version 0.1.0.9709; Aust & Barth, 2017), *purrr* (Version 0.2.4; Henry & Wickham, 2017), *readr* (Version 1.1.1; Wickham, Hester, & Francois, 2017), *stringr* (Version 1.3.0; Wickham, 2018b), *tibble* (Version 1.4.2; Wickham, Francois, & Müller, 2017), *tidyr* (Version 0.8.0; Wickham, 2017b), *tidyverse* (Version 1.2.1; Wickham, 2017c), and *XML* (Version 3.98.1.10; Lang & CRAN Team, 2018) for all of our statistical analyses. Datavyu (“Welcome! || datavyu,” n.d.) was used for video coding. All code used in data analysis and for this manuscript may be found in the GitHub repository associated with the paper (“PLAY project protocol manuscript,” n.d.).

To compute inter-observer reliability, we ran scripts in Datavyu to select a random segment of video (25% or 5 mins) from each 20- minute segment of natural play. Inter-observer reliability for the pilot videos was 96.7%-99.3% exact frame agreement ($\kappa_s > .93$, $ps < .001$) for duration codes (object interactions, locomotion, emotion) and 94.1%-99.5% code agreement ($\kappa_s > .89$, $ps < .001$) for categorical events (communicative acts, gesture).

Results

Informal inspection of the videos verified that dyads were unaffected by the presence of the researcher: After a few minutes of acclimation prior to natural play, mothers and infants ignore the researcher. Mothers appeared comfortable with all procedures, including the video home tour. The PIs found wide variation in home “disarray,” suggesting parents present homes as they would for any other casual visitor. Pilot testing verified the feasibility of

collecting all data in < 2 hours. So, with 1-2 hours of round-trip travel time, the entire data collection could be completed in < 4 hours.

All families were asked permission to share data with authorized researchers on Databrary using language adapted from the Databrary release template (???). Nineteen of the 20 families agreed to share; 3 families agreed to share data only with authorized Databrary researchers and 16 agreed to allow authorized researchers to show clips in public settings for informational or educational purposes.

In carrying out coding on X of the participants' sessions, we established that all four foundational coding passes are easy to learn (by undergraduate coders on Datavyu) and time efficient (< 3 hours for both infant and mother for each coding pass per hour of video). Adaptations to Datavyu enable transcription in English or Spanish to take 7-9 hours per hour of video, while infant and mother communicative acts and gestures together take < 3 hours to code.

Demographic and self-report measures. Due to some technical difficulties with pilot versions of the tablet app, we are only able to report here demographic and self-report data from $n=18$ families. Reconstructions of the missing data from video are ongoing, and updates will appear on the project's Databrary volume (???) once those are completed.

Infants had largely uncomplicated, healthy births. Birth weights (in g) ranged from 2616 to 4495 (mean = 3,443.80 g); 4 had birth complications (e.g., C-section, preeclampsia). Figure @ref{fig:plot-sound-avg-db} shows the parent-reported ages for the onset of hands and knees crawling and walking using procedures Adolph has used previously.

Mothers reported having between 17 and 23 years of education, ($M = 20.50$) and 15 were working at least part time. Sixteen of the eighteen respondents reported a partner cohabitating with the family. The partner's education level ranged from 17 and 23 years of education, ($M = 20.60$) and 15 were working at least part time.

All of the infants were exposed to and spoken to in English in the home. Many were spoken to or had exposure to languages other than English. At home 4 infants were spoken

to in Spanish; 3 infants were exposed to Spanish in a childcare setting. Other children were exposed to Armenian, French, Hindi, Polish, and Russian in the childcare setting. Six children in this sample were exposed only to English.

Ambient sound levels. Figure @ref{fig:plot-sound-avg-db} shows the average ambient sound levels recorded from the sites during an hour of recording, and Figure @ref{fig:plot-sound-peak-db} shows the peak ambient sound levels.

Video coding.

Discussion

The pilot study verified that the protocol met the launch group’s criteria on scientific and practical grounds. Four of the full sessions (13, 18, 29, and 20) met all internal criteria benchmarks. So we proceeded to design and plan a full-scale implementation.

Planned study

The proposed study builds upon and extends the pilot study.

Methods

A publicly-accessible wiki similar to the one used in the pilot (Soska et al., 2016) will be used to document all procedures and code definitions.

Participants. We will collect data from $n = 900$ infant-mother dyads from at least 30 different communities in 17 states located around the U.S. Each site will collect data from 30 infants, 10 each at 12-, 18-, and 24-months of age (+/- 1 week), with equal numbers of females and males. Figure ?? shows the proposed data collection sites and the non-collecting, data coding and analysis sites.

While not designed to be nationally representative, the data collection sites are diverse in aggregate based on Census data. To gather Census data reproducibly, we used the `choroplethr` package (Lamstein, 2017; Lamstein & Johnson, 2017) to download data from

the Census Bureau’s public API. This workflow allows us to easily gather and analyze other Census Bureau data about the communities targeted for sampling.

Figure @ref(fig:PLAY_race) shows the proportion of African American, Hispanic/Latino, and Asian residents in the counties surrounding the collection sites from which participating researchers will recruit. Figures ?? and ?? show economic and educational attainment indicators, respectively. Figure ?? shows the proportion of households speaking only English versus those where Spanish or other languages are spoken, exclusively or in addition to English. Data collection sites will have soft, advisory recruiting targets based on these sorts of measures for their individual communities.

Families will be two-parent, English and/or Spanish speaking households with resident fathers, with both parents greater than 18 years of age. Infants will be term firstborns, without birth complications or disabilities, and 12, 18, or 24 months of age (± 1 week); half of infants at each age and site will be boys.

The launch group deliberated over several sampling strategies [J. and P. Bornstein Marc H and Jager (2013); @#Jager2017-wv; @{DavisKean2017-sr}] Ultimately, we decided on homogeneous sampling to contain costs and understand behavioral variation. Homogenous sampling maintains some control over sample characteristics through a set of inclusionary criteria (here, firstborn status, English/Spanish home language, term pregnancy, etc.), while maximizing select aspects of diversity (e.g., geography, SES) and retaining sufficient power for group comparisons. We ruled against conventional convenience sampling, which leaves sampling decisions entirely to researchers’ discretion. Although convenience sampling is easy and cost efficient, it risks yielding a sample that varies on too many demographic dimensions to control. At the other extreme, population-based (probability) sampling is cost-prohibitive due to the required sample size. The sheer volume of data would prohibit transcription and video coding, and would require hiring and training special researchers for data collection, rather than relying on the existing expertise of the launch group.

Procedure. During an initial screening call, a researcher will determine eligibility for participation, and obtain demographic information. The researcher will schedule a 2-hour visit (weekday or weekend) when infants are between naps, meals, and baths, and would normally be home with their mothers. At the end of the visit, the researcher will ask mothers if the one-hour natural play session was representative of a typical day at home. If mothers report that infants' behavior or health was atypical, we will replace the dyad and document the replacement.

The visit will include, in order, parent consent, one-hour natural play, two structured play tasks, questionnaires, video home tour, and Databrary sharing permission. Although mothers will agree to share data in the initial screening call, we will request their signed permission at the end of the home visit when they are maximally informed. Consent, sharing permission, and questionnaire data will be entered on the custom app; paper forms and video camera on tripod will provide backup. Based on the pilot study, and our own experience in other studies, we anticipate that most families agree to share data. If families decline to share, their data can still be stored on Databrary and used by the collecting researcher for their own purposes.

Natural play: Like the pilot, we will record one hour of infant and mother activity in the home. Infant and mother will go about their daily routines without restrictions. They can move from room to room; mother can do chores; TV, music, or other media can be on. All of these behaviors were common in our pilot data. The researcher will hand-hold an HD video camera at the child's eye level, prioritizing view of infant over mother, keeping the infant's face, hands, and feet in view. If mother is visible, the researcher will capture as much of her face and body as possible without losing view of the baby. A cardioid microphone will amplify infant and mother speech and isolate background noise. Home visits avoid the artificiality of unfamiliar lab environments, materials, and tasks, and therefore come closest in fidelity to infant and mother natural behavior.

We will take several precautions to minimize effects of experimenter and camera

presence on dyads (McCune-Nicolich, Fenson, & Others, 1984; Stevenson, Leavitt, Roach, Chapman, & Miller, 1986). We will train researchers to remain unobtrusive. They will stay at a distance, resist talking to mother or infant, and watch the infant through the viewfinder to avoid eye contact. Filming will begin after several minutes of infant-mother acclimation to the camera and researcher.

Solitary play will entail infants playing with a set of 10 nesting cups (placed half up, half down) while sitting on a mat with mother nearby but not interacting (2 minutes). Nesting cups can reveal developmental differences in attention, manual action, spatial problem solving, and symbolic play. Younger pilot infants manually explored the cups but had difficulty nesting consecutive sizes; older pilots nested cups and used them symbolically (e.g., pretended to drink). In dyadic play infant and mother will play together on the mat with a standard set of toys (3 minutes)— truck, doll, baby bottle, small blanket, 2 tea cups, plates, and spoons—with mother instructed to “share the toys with her child.” The toys are conducive to non-symbolic (stacking plates, cups), symbolic (feeding doll, putting doll to sleep), and gendered play (with truck versus doll).

Following the play episodes, the researcher will walk through each room, filming walls, floors, ceilings, windows, room contents (including infants’ toys, books, media), and the contents of infants’ closets and drawers. The mother will name each room, describe infant’s access to rooms and spaces, and open closet doors and drawers for filming. Prior work and piloting ensured that mothers did not find this procedure intrusive. The researcher will narrate the video with comments about floor coverings (“throw rug,” “linoleum”) and anything not transparent to video.

During the visit, the researcher will record the clothes (front and back) and footgear (bottom, side, top) infants wore during the natural play session, and record the date infants began wearing the shoes and for how many days/week infants play indoors in shoes, socks, and barefoot.

After the naturalistic and structured play tasks, the researcher will interview mothers

on a range of infant and family measures that will yield information about language, locomotion, temperament, gender, home environment, and health (Table 2). The researcher will administer all questionnaires orally, and video records the interview (camera on tripod) for quality assurance, transparency, and possibly later coding.

Language (QL1): We will use the 12-month (words and gestures) and 18- to 24-month (words and sentences) versions of the MacArthur-Bates Communicative Development Inventory (MCDI). The MCDI is the most widely used instrument of infant language development, administered to over 60,000 children in 23 languages⁷⁵. The 12-month MCDI measures receptive and expressive vocabulary size and communicative gestures; the 18- to 24-month version contains a larger set of vocabulary items and simple sentence constructions. NL2: Mothers will report the language(s) spoken to infant by parents and childcare workers. Locomotion (QM1): Mothers will report the onset ages of hands-knees crawling and walking, using cell phone videos, photos, and diaries to jog their memories^{76,77}. QM2: Mothers will report on infants' fall-related injuries. Infant temperament (QE1) will be indexed with the Rothbart Early Childhood Behavior Questionnaire (ECBQ), very short form^{78,79}, which measures dimensions of surgency, negative affect, and effortful control. Gender (QG1): Mothers will report infants' use of gender labels (e.g., boy, girl) to refer to themselves or other people. QG2: Mothers will report their own and the father's attitudes to gender normative behavior (e.g., "I would be upset if my son wanted to dress like a girl"); and household division of labor (e.g., who does cooking). Environment (NH1): Ambient noise will be measured during natural play with a decibel meter, placed in the main room, to record peak and average dB every 100 ms. NH2: In the home video tour, the researcher will measure room dimensions with a laser distance measurer. QH3: At the end of the visit, the researcher will fill out a survey on the home environment (from launch group member Evans). QH4: Mothers will report use of electronic media (TV, computers, apps, etc.) by infant and family members (from launch group member Barr). Health (QF1-4): Mothers will report infant, parent, and family demographics, infants' health history (based on a subset of

questions from the ECLS-B 9-month and 2-year interviews), childcare experience, and parents' and family health history including SLI, ASD, and mental illnesses.

Video coding. Transcriptions and four core coding passes will be scored for both infant and mother. PLAY staff will transcribe speech at the utterance level, using standard criteria for segmenting speech (MacWhinney, 2000b). Utterances will be defined by independent clauses (statements with subject and predicate) with modifiers. Intonation and pauses can also define breaks (e.g., "You like that . Right?" is two utterances). Mothers' language-like sounds are typed out phonetically. Infant babbles are marked with "b" and non-linguistic vocalizations (cry, laugh, grunt) with "c." Unintelligible utterances are marked "xxx." Utterances will be time-locked to video, revealing overlaps in infant-mother speech, and co-occurrence and sequencing of speech with object interactions, emotions, and locomotion. Spanish transcriptions will follow the same rules.

Based on transcripts, mothers' utterances will be coded as declaratives (labels and descriptions of objects and events "Red"; "Puppy"), attention-imperatives that solicit infant attention ("Look at that"), action-imperatives that solicit infant action ("Put it there"), or prohibition-imperatives ("Stop it!"); interrogatives (open- and close-ended questions, "Is it hot?" with the exception of "tag" questions, which will be coded as declaratives, "That's a ball, right?"); affirmation/conversational fillers ("Yes!"; "What's next?"), and unintelligible. Infants' vocalizations will be categorized as language (sentences or words), prelinguistic vocalizations (babbling or vowels), non-linguistic vocalizations (e.g., cry, laugh, scream, grunt), or unintelligible.

Gestures will be categorized as points, show/hold up (deictic), conventional (wave bye-bye, thumbs-up), and representational (flapping arms to represent a bird).

Object interactions will be coded for onset and offset of manual engagement (touching, manipulating, carrying) with any manipulable, moveable object or part of an object that moves through space. Locomotion will be coded for onset and offset of self-generated locomotion of any form (e.g., crawling, walking, climbing, stepping in place). Coders also

score falls, and periods when the infant is held or constrained by furniture (e.g., highchair). Emotion will be coded for onset and offset of positive (smiling, laughing) and negative (crying, frowning, fussing) facial expressions. Inter-observer reliability:

To verify inter-observer reliability, PLAY staff will rescore 25% of each infant's natural play video (5 minutes randomly drawn from each 20-minute segment), blind to the original coders' output (categorical measures: kappas $> .85$; duration measures: % exact frame agreement $> 90\%$). If codes are not reliable, PLAY staff will reassign the videos to a new lab.

Data analysis. The launch group will jointly establish best practices for PLAY analyses. Our guidelines will include: recommendations to pre-register predictions and analyses; the use of procedures that use one portion of the data set to explore correlations among variables and a separate subsample to confirm it; the use of reproducible and transparent workflows for data processing and analyses (e.g., Ruby scripts in Datavyu; syntax instead of menu-driven commands for SPSS users; scripts and functions for R users); a commitment to openly sharing supplementary video codes and operational definitions to avoid unnecessary duplication of coding efforts; and open sharing of null results as well as positive findings. We will create means for communication (e.g., a Google group) among launch group members who wish to discuss, propose, and organize team efforts focused on answering specific video-based research questions. In addition, in keeping with emerging standards in research transparency, we will report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

General Discussion

The PLAY corpus will be a treasure trove of data, and it will all be made available openly to the research community at the end of the study. Our hope is that it will seed substantial new scholarship.

Researchers can examine real-time behavioral cascades among infant behaviors, among mother behaviors, and between infants and mothers. They can test whether particular infant

behaviors are temporally connected (e.g., vocalizations and gestures) or independent (vocalizations and locomotion). They can test infant-to-mother cascades and vice versa, such as whether infant emotional expressions affect real-time language input from mother. Prior correlational work, for example, shows that infants who express higher quantities of negative emotions display lower levels of language development on the MCDI and later language milestones (Bloom & Capatides, 1987; Salley & Dixon, 2007). But the evidence for these findings offers limited insight into the real-time behaviors that underlie the correlations (Bloom & Beckwith, 1989). With PLAY, researchers can examine real-time behavioral cascades by testing whether infants' negative emotions (Table 1 VE1) hinder interactions with objects (Table 1 VO1) and/or vocal and gestural communications (Table 1 VL2-3), and consequently, lead to low quantity and diversity of mother speech (Table 1 VL1-2). Infant emotions could also facilitate language learning: Emotional expressions might elicit mental state terms and emotion words from mothers (e.g., "You think mommy's leaving?", "Why are you sad?"). Regardless, whether and how infant emotions affect their language development requires data on the words mothers use preceding, during, and following infant emotional behaviors in real time.

PLAY's three age groups and measures of skill and experience (e.g., MCDI, walking experience) allow researchers to investigate developmental cascades in new ways. We can examine age-related changes in temporal coordination among infant, mother, and infant-mother behaviors—such as whether infants of different ages with different skills elicit different behaviors in mothers. For example, object interactions in 12-month-olds, who are typically at the cusp of conventional word use, might elicit declaratives from mothers ("That's a truck!"), whereas object interactions in 24-month-olds, who typically have substantial expressive vocabularies, might elicit interrogatives ("What's that?"). Alternatively, researchers might compare language cascades in infants of different ages but with similar skills—whether the vocalizations of 18- versus 24-month-olds matched on MCDI vocabulary size elicit similar or different language input from mothers. Finally, we might

compare real-time cascades in infants of the same age but with different skills—such as whether 18-month-olds who use isolated words versus those who combine words into simple sentences, elicit different language input from mothers. Comparisons of real-time contingencies by infant age and skill level provide a unique window into understanding developmental mechanisms that underpin behavioral change.

Environmental cascades. PLAY’s rich array of environmental measures, ranging from distal macro environmental characteristics (e.g., SES, geographic region) to proximal environmental features (e.g., clutter and chaos), will advance understanding of how environmental risks affect everyday opportunities for learning. Researchers might test proximal environmental cascades on the quantity and quality of infants’ object interactions and locomotion, for example by coding video home tours (Table 1 VH1) for object availability and using laser measurements of room dimensions (Table 2 NH2). We can relate environmental features of clutter, ambient noise, and so on, to mothers’ speech and infants’ language development. Researchers can expand the lens of environmental influences to consider how distal macro factors, such as family SES and geo-coded data on neighborhood poverty (Table 2 NH5) relate to proximal environmental measures—objects and space in the home—and in turn infant behaviors, mother behaviors, and infant language and skill.

Databrary (“Databrary,” n.d.), funded by NICHD/NSF, is a digital web-based library for sharing and reusing research videos, clips, and displays. Researchers can reuse shared videos to ask questions beyond the scope of the original study (K. E. Adolph, n.d.). They can use shared video clips to learn about procedures and to illustrate findings and displays for teaching (Rick O Gilmore & Adolph, 2017). Sharing is easy. To mitigate the onerous task of curating a dataset after the study is completed, Databrary developed an active curation system. Researchers can use Databrary as a file manager, lab server, and secure backup prior to sharing (“Databrary,” n.d.; A. S. Gordon, Millman, Steiger, Adolph, & Gilmore, 2015). When they are ready to share, they need only click a button.

Video contains personally identifiable information, so sharing poses special ethical

issues. Advised by ethics experts, IRB and grants/contracts administrators, and legal counsel, Databrary developed a policy framework^{53,54} for sharing identifiable data based on obtaining participants' permission to share and restricting access to authorized researchers under the oversight of their institutions. Since the Databrary website went live in 2014, 580+ researchers (including the launch group) and 245+ affiliates from 330+ institutions around the world are authorized. The repository contains 7820+ hours of video from 7830+ participants. Datavyu (K. E. Adolph, 2015; "Welcome! || datavyu," n.d.) is a powerful, flexible, coding tool that allows researchers to manipulate the temporal-spatial properties of behavior and to tag portions of the video for events and behaviors of interest. With fingertip control over video playback, they can run the video forward and backward at varying speeds ($\pm 1/32$ - 32x normal speed) or jog frame by frame to determine when behaviors began and ended, freeze frames to dissect behavior into its component parts, zoom in/out to focus on details or the larger context, and label behavioral events with categorical and qualitative codes. Each code is time-locked to the video to facilitate tests of behavioral cascades and real-time contingencies based on sequential order, duration, and begin/end times of events. A full scripting language allows researchers to manipulate the spreadsheet, error-check entries, import other data streams, and export data to their specifications for analyses. The latest Datavyu release has new features to reduce the notoriously high cost of transcribing infant and mother speech in noisy contexts, time locked to video, at the utterance level (from the typical 10-12 hours per hour of video to 7-9 hours).

Naturally, The creation of a large dataset with many variables raises the possibility that a particular statistically significant finding may be spurious. In particular, correlational analyses among non-video questionnaire data require special protection against spurious findings because of the large number of easily available measures (Table 2). The corpus includes a summary score for each infant on each instrument, subscores for standard scales, and raw data for each item. For example, researchers will have access to infants' total productive vocabulary on the MCDI, the number of words produced within specific

categories (e.g., animal words; action words), and production of each word. Similarly, researchers will have access to fully processed, ready-to-analyze data on infant temperament, locomotor experience, infant health, environmental chaos, media use, family demographics, and so on. Spurious results and duplication of analyses are especially likely from these “low-hanging fruit.”

In contrast to the ready-to-use questionnaire data, analyses of time-locked video codes raise other analytic issues. Data from the foundational coding passes will not be “ready to go.” Researchers will need to make decisions about how to process the data—whether to turn categorical codes into frequencies or rates; whether to convert onset/offset times into average durations, latencies from one behavior to another, sequences of behavior, or other analytic constructs. We will encourage individual launch group members to use their expertise and Datavyu training to mine the video corpus (by further coding of natural play and coding of structured play sessions and the home tour). Additional coding passes will be labor intensive, and duplication of coding effort would waste researchers’ time.

Of course, PLAY’s homogenous sampling strategy and cross-sectional design have limitations. Although the sample will not be nationally representative, it will capture important demographic variations and can easily grow. With only one session per dyad, we cannot test stability or predictive validity of behaviors. However, the protocol and codes can be easily extended to other populations and to longitudinal designs (several launch group members plan to do this). If labs assigned to data collection or coding cannot fulfill their tasks, we will replace them.

We will monitor ongoing data collections. If the data are too homogeneous, we will ask some sites to recruit more than their allotment. If a lab’s codes are not reliable, we will reassign the videos and retrain the coder.

In conclusion, the PLAY project represents an innovative, synergistic, cross-domain approach to developmental science that will facilitate scientific discovery, transparency, and reproducibility we hope for years to come. In creating the first, large-scale, sharable,

510 reusable, fully transcribed, coded, and curated video corpus of human behavior, we hope to
511 establish video sharing of procedures, codes, and findings as a new standard in
512 developmental and behavioral science. In so doing, we hope to help answer fundamental
513 cross-domain questions about behavioral, environmental, and developmental cascades as they
514 shape the playful work of infants.

References

- Adolph, K. E. (2015). Best practices for coding behavioral data from video.
<http://datavyu.org/user-guide/best-practices.html>. Retrieved from
<http://datavyu.org/user-guide/best-practices.html>
- Adolph, K. E. (n.d.). Video as data: from transient behavior to tangible recording.
APS Observer, in press.
- Adolph, K. E., Gilmore, R. O., & Kennedy, J. L. (2017, October). Video data and
 documentation will improve psychological science.
<http://www.apa.org/science/about/psa/2017/10/video-data.aspx>. Retrieved from
<http://www.apa.org/science/about/psa/2017/10/video-data.aspx>
- Adolph, K. E., Gilmore, R. O., & Tamis-LeMonda, C. (2016). Play & learning across a year
 (PLAY) workshop at NICHD. <https://nyu.databrary.org/volume/254>.
 doi:10.17910/B7.254
- Adolph, K. E., Tamis-LeMonda, C. L., & Gilmore, R. O. (n.d.). PLAY project: Webinar
 discussions on protocol and coding. <https://nyu.databrary.org/volume/232>.
 Retrieved from <https://nyu.databrary.org/volume/232>
- Adolph, K. E., Tamis-LeMonda, C., & Gilmore, R. O. (2018, March). PLAY pilot data
 collections. <https://nyu.databrary.org/volume/444>. Retrieved from
<https://nyu.databrary.org/volume/444>
- Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*.
 Retrieved from <https://github.com/crsh/papaja>
- Barbaro, K. de, Johnson, C. M., Forster, D., & Deák, G. O. (2015). Sensorimotor decoupling
 contributes to triadic attention: A longitudinal investigation of Mother-Infant-Object
 interactions. *Child Development*, 87(2), 494–512. doi:10.1111/cdev.12464
- Bloom, L. (1995). *The transition from infancy to language: acquiring the power of expression*.
 Cambridge University Press. Retrieved from

<https://market.android.com/details?id=book-ntvdZv1wttQC>

Bloom, L., & Beckwith, R. (1989). Talking with feeling: Integrating affective and linguistic expression in early language development. *Cognition and Emotion*, 3(4), 313–342.

doi:[10.1080/02699938908412711](https://doi.org/10.1080/02699938908412711)

Bloom, L., & Capatides, J. B. (1987). Expression of affect and the emergence of language. *Child Development*, 58(6), 1513–1522. doi:[10.2307/1130691](https://doi.org/10.2307/1130691)

Bornstein, J. and P., Marc H and Jager. (2013).

Sampling in developmental science: situations, shortcomings, solutions, and standards. *Developmental Review: DR*, 33(4), 357–370.

doi:[10.1016/j.dr.2013.08.003](https://doi.org/10.1016/j.dr.2013.08.003)

Bruner, J. S. (1975). Play is serious business. *Psychology Today*, 8(8), 80–83.

Bruner, J. S. (1976). *Play, its role in development and evolution*. Basic Books. Retrieved from <https://market.android.com/details?id=book-Zq8oAAAAYAAJ>

Cole, W. G., Lingeman, J. M., & Adolph, K. E. (2012). Go naked: Diapers affect infant walking. *Developmental Science*, 15(6), 783–790.

doi:[10.1111/j.1467-7687.2012.01169.x](https://doi.org/10.1111/j.1467-7687.2012.01169.x)

Databrary. (n.d.). <https://nyu.databrary.org/>. Retrieved from <https://nyu.databrary.org/>

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.

doi:[10.1016/j.cognition.2016.03.005](https://doi.org/10.1016/j.cognition.2016.03.005)

Gesell, A. (1946). Cinematography and the study of child development.

The American Naturalist, 80(793), 470–475. Retrieved from

<http://www.jstor.org/stable/2458189>

Gesell, A. (1991). Cinemanalysis: a method of behavior study.

The Journal of Genetic Psychology, 152(4), 549–562.

doi:[10.1080/00221325.1991.9914712](https://doi.org/10.1080/00221325.1991.9914712)

Gibson, E. J. (1994). Has psychology a future? *Psychological Science*, 5(2), 69–76.

doi:[10.1111/j.1467-9280.1994.tb00633.x](https://doi.org/10.1111/j.1467-9280.1994.tb00633.x)

Gilmore, R. O., & Adolph, K. E. (2017). Video can make behavioural science more reproducible. *Nature Human Behavior*, 1. doi:[10.1038/s41562-017-0128](https://doi.org/10.1038/s41562-017-0128)

Gilmore, R. O., & Adolph, K. E. (n.d.). *Open sharing of research video: Breaking the boundaries of the research team, in advancing social and behavioral health research through cross-disciplinary team science: Principles for success*. (Hall, Kara, Croyle, R., & Vogel, A., Ed.). Springer.

Gilmore, R. O., Adolph, K. E., & Millman, D. S. (2016). Curating identifiable data for sharing: The databrary project. In *2016 new york scientific data summit (NYSDS)* (pp. 1–6). ieeexplore.ieee.org. doi:[10.1109/NYSDS.2016.7747817](https://doi.org/10.1109/NYSDS.2016.7747817)

Gilmore, R. O., Adolph, K. E., Millman, D. S., & Gordon, A. (2016). Transforming education research through open video data sharing. *Advances in Engineering Education*, 5(2). Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/28042361>

Glenn, E. H. (2018). *Acs: Download, manipulate, and present american community survey and decennial data from the us census*. Retrieved from <https://CRAN.R-project.org/package=acs>

Gordon, A. S., Millman, D. S., Steiger, L., Adolph, K. E., & Gilmore, R. O. (2015). Researcher-library collaborations: Data repositories as a service for researchers. *Journal of Librarianship and Scholarly Communication*, 3(2). doi:[10.7710/2162-3309.1238](https://doi.org/10.7710/2162-3309.1238)

Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>

Hoff, E. (2013). *Language development*. Cengage Learning. Retrieved from <https://market.android.com/details?id=book-7MxTSQlfUVUC>

Iverson, J. M., & Wozniak, R. H. (2007). Variation in vocal-motor development in infant siblings of children with autism. *Journal of Autism and Developmental Disorders*,

37(1), 158–170. doi:[10.1007/s10803-006-0339-z](https://doi.org/10.1007/s10803-006-0339-z)

Karasik, L. B., Adolph, K. E., Tamis-LeMonda, C. S., & Zuckerman, A. L. (2012). Carry on: Spontaneous object carrying in 13-month-old crawling and walking infants.

Developmental Psychology, 48(2), 389–397. doi:[10.1037/a0026040](https://doi.org/10.1037/a0026040)

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(4),

1199–1209. doi:[10.1111/j.1467-8624.2011.01595.x](https://doi.org/10.1111/j.1467-8624.2011.01595.x)

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science*, 17(3),

388–395. doi:[10.1111/desc.12129](https://doi.org/10.1111/desc.12129)

Karasik, L. B., Tamis-LeMonda, C. S., Adolph, K. E., & Bornstein, M. H. (2015). Places and postures: A cross-cultural comparison of sitting in 5-month-olds. *Journal of*

Cross-Cultural Psychology, 46(8), 1023–1038. doi:[10.1177/0022022115593803](https://doi.org/10.1177/0022022115593803)

Lamstein, A. (2017). *ChoroplethrMaps: Contains maps used by the 'choroplethr' package*.

Retrieved from <https://CRAN.R-project.org/package=choroplethrMaps>

Lamstein, A., & Johnson, B. P. (2017). *Choroplethr: Simplify the creation of choropleth maps in r*. Retrieved from <https://CRAN.R-project.org/package=choroplethr>

Lang, D. T., & CRAN Team. (2018). *XML: Tools for parsing and generating xml within r and s-plus*. Retrieved from <https://CRAN.R-project.org/package=XML>

MacWhinney, B. (2000a). The CHILDES project: Tools for analyzing talk (third edition): Volume i: Transcription format and programs, volume II: The database.

Computational Linguistics, 26(4), 657–657. doi:[10.1162/coli.2000.26.4.657](https://doi.org/10.1162/coli.2000.26.4.657)

MacWhinney, B. (2000b). *The CHILDES project: Tools for analyzing talk*. Lawrence Erlbaum Associates. doi:[10.1162/coli.2000.26.4.657](https://doi.org/10.1162/coli.2000.26.4.657)

McCune-Nicolich, L., Fenson, L., & Others. (1984). Methodological issues in studying early pretend play. *Child's Play: Developmental and Applied*, 81–104.

Montessori, M. (1984). *The absorbent mind*. New York: Dell Publishing Co., Inc. Retrieved

- from <https://openlibrary.org/books/OL22287831M.opds>
- Müller, K. (2016). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from <https://CRAN.R-project.org/package=bindrcpp>
- Onis, M. (2006). WHO motor development study: windows of achievement for six gross motor development milestones. *Acta Paediatrica*, 95(S450), 86–95. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1651-2227.2006.tb02379.x/full>
- Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [Stat.CO]*. Retrieved from <https://arxiv.org/abs/1403.2805>
- Piaget, J. (1967). *Play, dreams and imitation in childhood*. London: Routledge & K. Paul. Retrieved from <https://openlibrary.org/books/OL22344377M.opds>
- PLAY project protocol manuscript. (n.d.). Retrieved from <https://github.com/PLAY-behaviorome/ibad-lets-PLAY>
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Robinson, A. K. &, SR. (2015). Motor development. In *Handbook of child psychology and developmental science* (Vol. 2, pp. 114–157). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781118963418.childpsy204
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, 83(5), 1762–1774. doi:10.1111/j.1467-8624.2012.01805.x
- Salley, B. J., & Dixon, W. E., Jr. (2007). Temperamental and joint attentional predictors of language development. *Merrill-Palmer Quarterly*, 53(1), 131–154. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/18080005>
- Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare

environments. *PloS One*, 8(11), e80646. doi:[10.1371/journal.pone.0080646](https://doi.org/10.1371/journal.pone.0080646)

Soska, K., Adolph, K. E., Tamis-LeMonda, C. T., & Gilmore, R. O. (2016, December). The PLAY project wiki. <https://dev1.ed-projects.nyu.edu/wikis/docuwiki/doku.php>.

Retrieved from <https://dev1.ed-projects.nyu.edu/wikis/docuwiki/doku.php>

Stevenson, M. B., Leavitt, L. A., Roach, M. A., Chapman, R. S., & Miller, J. F. (1986).

Mothers' speech to their 1-year-old infants in home and laboratory settings. *Journal of Psycholinguistic Research*, 15(5), 451–461. doi:[10.1007/BF01067725](https://doi.org/10.1007/BF01067725)

Welcome! || datavyu: video coding and data visualization tool. (n.d.). <http://datavyu.org>.

Retrieved from <http://datavyu.org>

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Retrieved from <http://ggplot2.org>

Wickham, H. (2017a). *Httr: Tools for working with urls and http*. Retrieved from

<https://CRAN.R-project.org/package=httr>

Wickham, H. (2017b). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*.

Retrieved from <https://CRAN.R-project.org/package=tidyr>

Wickham, H. (2017c). *Tidyverse: Easily install and load 'tidyverse' packages*. Retrieved

from <https://CRAN.R-project.org/package=tidyverse>

Wickham, H. (2018a). *Forcats: Tools for working with categorical variables (factors)*.

Retrieved from <https://CRAN.R-project.org/package=forcats>

Wickham, H. (2018b). *Stringr: Simple, consistent wrappers for common string operations*.

Retrieved from <https://CRAN.R-project.org/package=stringr>

Wickham, H., & Francois, R. (2016). *Dplyr: A grammar of data manipulation*. Retrieved

from <https://CRAN.R-project.org/package=dplyr>

Wickham, H., Francois, R., & Müller, K. (2017). *Tibble: Simple data frames*. Retrieved from

<https://CRAN.R-project.org/package=tibble>

Wickham, H., Hester, J., & Francois, R. (2017). *Readr: Read rectangular text data*.

Retrieved from <https://CRAN.R-project.org/package=readr>

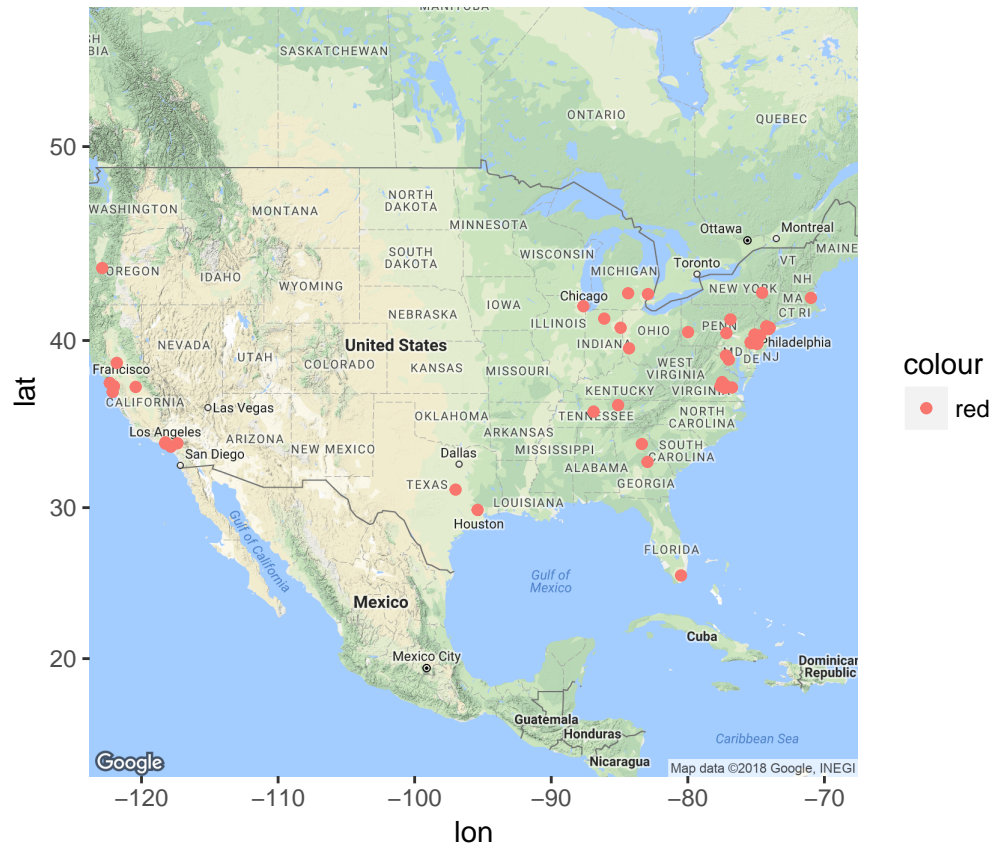


Figure 1. A map of the PLAY project data collection sites.

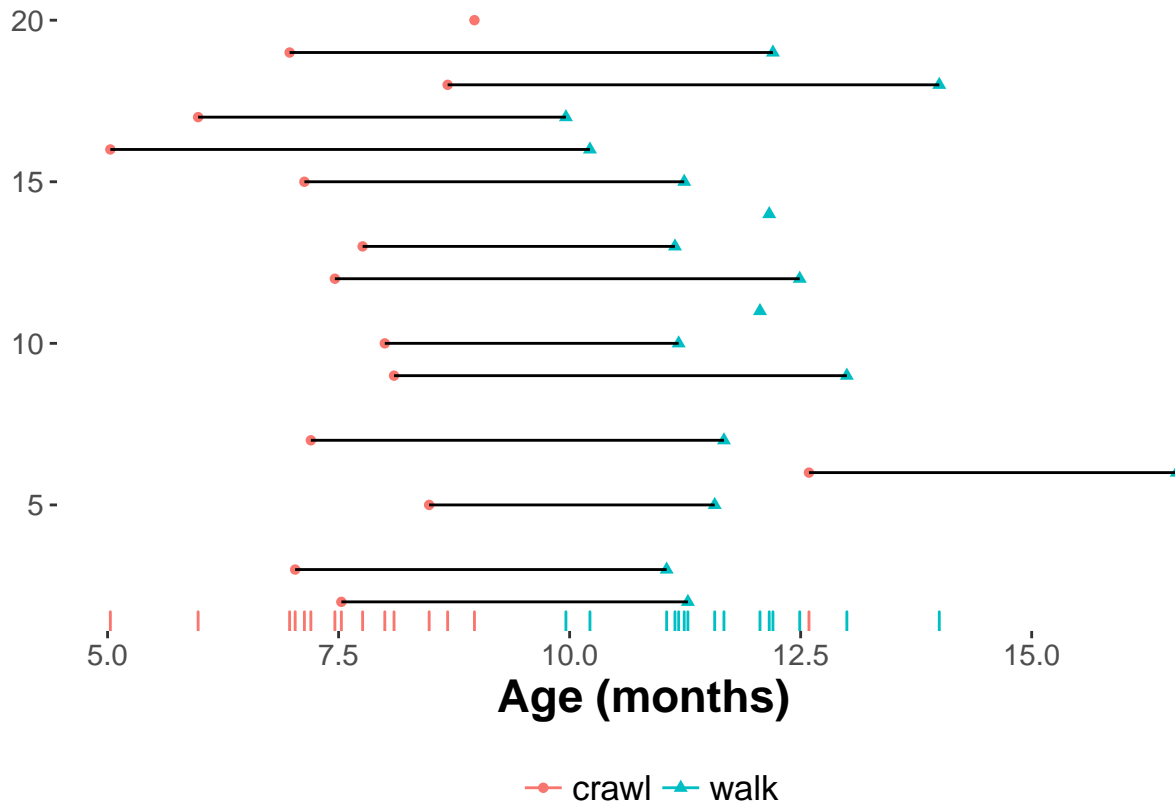


Figure 2. Parent-reported onset ages for hands and knees crawling and walking