Using automated detections to quantify how postural developments shape the social information in view for young infants

Bria L. Long[1], Alessandro Sanchez[1], Allison M. Kraus[1], Ketan Agrawal[1], & Michael C. Frank[1]

[1] Department of Psychology, Stanford University

Author Note

Correspondence concerning this article should be addressed to Bria L. Long, 450 Serra Mall, Stanford CA 94305. E-mail: bria@stanford.edu

Abstract

Social information plays a key role in infants' early linguistic and cognitive development. Yet infants locomotor abilities also develop continuously as they learn to sit, crawl, stand, and walk – influencing the way that infants see the world around them. How do these postural developments affect infants access to the social information relevant for early learning? To address this question, we first created a rich annotated dataset of naturalistic play sessions between infants and their caregivers with sets of novel and familiar toys (N=36, 8-16 months of age), recording both egocentric and third-person video. We then developed an automated analysis method using a pose detection model (Zhang et al., 2017) to detect the faces and hands in the infant view, allowing us to analyze the entirety of this dataset. We also applied this automated method to second egocentric video dataset where one-year-olds explored a large play area with their caregivers (Franchak et al., 2017). We found that infants' posture and orientation to their caregiver changed dramatically across this age range and modulated infants access to social information; infants who were sitting or standing with their caregiver at a close distance tended to have the most faces and hands in their visual field. We also found convergence across both datasets, confirming the finding that motoric developments play a significant role in shaping the social information that infants have access to over development, and suggesting that they may play a role in emergence of infants' cognitive and linguistic abiltiies. We suggest that the combined use of head-mounted cameras and the application of new computer vision techniques is a promising avenue for understanding the statistics of infants' visual and linguistic experience as they change over development and for re-utilizing rich video datasets collected across different labs and diverse contexts.

Keywords: Postural developments, headcam, social information, face detection, pose detection

Word count: XXX

Using automated detections to quantify how postural developments shape the social information in view for young infants

From their earliest months, infants are deeply engaged in learning from others. Even newborns tend to prefer to look at faces with direct vs. averted gaze (Farroni, Csibra, Simion, & Johnson, 2002) and young infants follow overt gaze shifts (Bruner, 1975; Gredeback, Fikke, & Melinder, 2010). And as infants reach their first birthday, they also tend to follow (Yu & Smith, 2013, 2017) and imitate the gestures of their caregivers (e.g., pointing).

Infants ability to process these social cues may provide strong scaffolding for early word learning. Longitudinal studies provide some evidence for this link: children's level of joint engagement with their mother at 9-12 months predicts both their receptive and productive vocabularies (Carpenter, Nagell, & Tomasello, 1998) and 10 month-olds who follow an adult's gaze (in an experimental context) have larger vocabularies at 18 months and throughout the second year of life (Brooks & Meltzoff, 2005, 2008). While the relationship between hand-following and language development has been less well characterized, infants who follow their caregivers' hands tend to be those who spend more time jointly attending to events with their caregivers (Yu & Smith, 2017).

Relatively little work, however, has quantified how often infants see and use these kinds of social cues in naturalistic learing environments. By using head-mounted cameras to record what infants see, researchers have begun to document the infant egocentric perspective (Yoshida & Smith, 2008) and to quantify the information – social and otherwise – available to infants as they learn. While head-mounted camera data do not provide explicit information on where infants are attending, some work suggests that infants orient their head towards what they are focusing on – putting those people or objects in view (Yoshida & Smith, 2008). Initial recordings from in-lab play sessions revealed a different view than many imagined: instead of being dominated by their faces, the infant perspective contained close up views of primarily toys and hands (Franchak, Kretch, Soska, & Adolph, 2011; Yoshida & Smith, 2008; Yu & Smith, 2017). Subsequent research has revealed that the infant view is far from a unitary construct, undergoing dramatic changes as infants grow. Recordings from home environments suggest that the viewpoints of very young infants – less than 4 months of age – do indeed contain persistent and frequent views of faces (Fausey, Jayaraman, & Smith, 2016; Jayaraman, Fausey, & Smith, 2017) but that the infant view tends to contain more and more hands as infants grow older.

These changes in perspective are likely the downstream consequence of a myriad factors, chief of which may be infants' evolving locomotive abilities: an infants ability to sit, crawl, stand, or walk structures the way they interact with the things and people in their world. These changes can be rapid: the physical body of an infant can lengthen up to 2cm in 24 hours (Lampl, 1993), leading to changes in their motoric abilities from one day to the next. These motoric developments can be thought of as gateways that open up entirely new phases of development (Iverson, 2010), and cause a cascade of changes in infants ability to interact with their world and the people in it (Karasik, Tamis-LeMonda, & Adolph, 2014).

Thus, infants' changing locomotor abilities may shape the social cues that infants' see and seek out, in turn impacting their cognitive and linguistic abilities. Supporting this idea, infants experience with sitting predicts their success at 3D object completion tasks (Soska, Adolph, & Johnson, 2010) as well as their receptive vocabulary (Libertus & Violi, 2016). And as children begin crawling (Adolph, Vereijken, & Denny, 1998) – or scooting or cruising (Patrick, Noah, & Yang, 2012) – they are no longer constrained to the same spot that their caregivers last placed them in. Yet while crawlers can choose where to go and what they see to a much greater degree, they also appear to spend much of their time in a world populated by floors and knees; during spontaneous play, toddlers are more likely to look at the floor while crawling than while walking (Franchak et al., 2011), when they have full visual access to their environment and the people in it (Kretch, Franchak, & Adolph, 2014).

On one view, it is children's ability to stand and walk that fundamentally changes their ability to access social information (e.g., facial expressions, gaze cues, pointing) relative to children who are still crawling and sitting (Walle, 2016). In turn, this ability to access more detailed social information could allow infants to learn words quicker and more efficiently, facilitating language growth (Walle, 2016). Supporting this idea, walking infants make different kinds of object-related bids for attention from their caregivers than crawling infants, and tend to hear more action directed statements (e.g., "open it") (Karasik et al., 2014); further, in an observational study, Walle and Campos (2014) also found that children who were able to walk had both higher receptive and productive vocabularies. However, not all evidence supports this view: some parental report data suggest a weak relationship between walking and the onset of language (Moore, Dailey, Garrison, Amatuni, & Bergelson, 2019), and head-mounted eye-tracking data have highlighted the role of infants' in-the-moment posture – and that of their caregivers' – in what infants' see, finding that these factors are better predictors of the social information in view than their locomotor status per se (i.e., "walker" vs "crawler") (Franchak, Kretch, & Adolph, 2017).

Understanding this nuanced relationship between the social information in the infant view and their motoric and linguistic development has remained challenging. Indeed, different kinds of play sessions (e.g., exploring novel environments vs. playing with novel objects) might modulate the degree to which infants' posture changes the social information in view. Furthermore, no work to date has directly examined how the availability of social information changes in naturalistic language learning contexts. For example, it is unclear if and how the social information in view changes during the naming events for different objects (e.g, "Look, a [ball]").

More broadly, the field is in need of computational tools to reutilize these rich head-mounted camera datasets and to understand the generalizability of findings across different datasets, populations, tasks, and age-ranges. While the field has assembled many head-mounted camera datasets, conducting new analyses on these videos has remained prohibitively time-consuming due to a lack of computational tools for annotations. Instead, hundreds of hours of manual annotations have been required to analyze a fraction of the available frames for a given analysis. Thus, despite containing a wealth of information about the structure of parent-child interactions, these datasets have thus gone dramatically underused.

Recent innovations in computer vision hold promise for automated annotations of the infant view. Over the past decade, deep neural networks have become dramatically better at a wide range of visual tasks, including object classification, Simonyan and Zisserman (2014)], scene categorization (Zhou, Lapedriza, Khosla, Oliva, & Torralba, 2017), and pose detection (K. Zhang, Zhang, Li, & Qiao, 2016), arguably facilitating our understanding of visual perception (Peterson, Abbott, & Griffiths, 2018, VanRullen (2017)) and improving computational neuroscience (Kietzmann, McClure, & Kriegeskorte, 2018). However, as most models have been trained on photographs or videos taken from the adult perspective, it is still unclear how easily these models can be applied to videos taken from the infant perspective.

In this paper, we thus first validate a model for the automated detection of faces and hands in the infant view, using a open-source model of pose detection – OpenPose(Cao, Simon, Wei, & Sheikh, 2017; K. Zhang et al., 2016) – that provides pose, face, and hand keypoints for every person in an image. We compare the detection accuracy of OpenPose with that of both older and more specialized models of face detection and demonstrate the usability of this off-the-shelf model for quantifying the faces and hands in the infant view.

We then apply this automated method to a head-mounted camera dataset collected in our lab to directly examine the role of postural developments on the social information in the infant view during language learning. Specifically, we constructed a dataset where infants and caregivers played freely using a set of novel and familiar objects; 8, 12, and 16 months-olds (N=26; N=12 in each age group) participated to ensure a wide range of locomotor abilities. Infants' in-the-moment posture and orientation relative to their caregiver were hand annotated and all play sessions were transcribed. This cross-sectional design thus allowed us to directly examine the relative contributions of age vs. postural developments on children's visual access to social information, and how the avaliability of social cues changes relative to naming events (e.g., "Yes, you like the [ball!]"). Broadly, we predicted that there would be differential access to social information based on children's postural developments: crawling infants would see fewer faces/hands because they would primarily be looking at the ground, while walking toddlers would have access to a richer visual landscape with greater access to the social information in their environment.

To examine the robustness of our automated method and the generalizability of our findings, we applied OpenPose (Cao et al., 2017) to the dataset collected in Franchak et al. (2017), where one-year-olds wore head-mounted eye-tracking cameras during a play session and their in-the-moment posture was hand-annotated. Unlike the present dataset, infants and caregivers roamed a large, open playroom and explored different toys placed throughout. We thus analyzed this dataset with the goal of both (1) validating our automated method on a very different kind of dataset, attempting to replicate their primary findings originally obtained with head-mounted eye-tracking (2) extending their findings to ask whether whether the proportion of hands in view also varies with infants' posture in an exploratory play session.

Methods

Free-play video dataset

A primary goal of this project was to provide a rich, open dataset for other researchers to examine the effects of postural developments and labeling behavior during naturalistic parent-child interactions. We thus invited caregivers of 8, 12, and 16-month-olds to participate in play sessions where they were provided with pairs of novel and familiar objects (e.g., a ball and a microfiber duster, called a "zem") in a small playroom in a lab (approximately 10' x 10'). Infants wore head-mounted cameras equipped with a fish-eye lens (see Head-mounted camera), a third-person camera captured a birds-eye view of the play session. Using these video data, infants' posture and orientation to their caregiver were hand-coded and annotated for the entirety of the play session. All videos were transcribed, and we collected collected MacArthur CDIs for all children who particpated. All materials have been made publicly avaliable on Databrary for whom the parents provided sharing consent (29/36 dyads) via https://nyu.databrary.org/volume/101

Participants. Our final sample consisted of 36 infants and children, with 12 participants in three age groups: 8 months (6 F), 12 months (7 F), and 16 months (6 F). Participants were recruited from the surrounding community via state birth records, had no documented disabilities, and were reported to hear at least 80 percent English at home. Demographics and exclusion rates are given in the table below (see Table 1)/

To obtain this final sample, we tested 95 children, excluding 59 children for the following reasons: 20 for technical issues related to the headcam (e.g., failure to record, ran out of battery), 15 for failing to wear the headcam, 10 for fewer than 4 minutes of headcam footage, 5 for having multiple adults present, 5 for missing Communicative Development Inventory (CDI) data, 2 for missing scene camera footage, 1 for fussiness, and one for sample symmetry. 16-month-olds tolerated the head-mounted camera less well than younger infants, leading to a higher exclusion rate. All inclusion decisions were made independent of the results of subsequent analyses.

Head-mounted camera. We used a small, head-mounted camera ("headcam") that was constructed from a MD80 model camera attached to a soft elastic headband. Videos captured by the headcam were 720x480 pixels with 25 frames per second.[1] A fisheye lens was attached to the camera to increase the view angle from 32° horizontal by 24° vertical to 64° horizontal by 46° vertical (see Figure 1, above).

Even with the fish-eye lens, the vertical field of view of the camera is still considerably reduced compared to the child's field of view, which spans around 100–120° in the vertical dimension by 6-7 months of age (Cummings, Van Hof-Van Duin, Mayer, Hansen, & Fulton, 1988; Mayer, Fulton, & Cummings, 1988). As we were originally primarily interested in the presence of faces in the child's field of view, we chose to orient the camera upwards to capture the entirety of the child's upper visual field where the child is likely to see adult faces, understanding that this decision limited our ability to detect hands (especially those of the child, which are typically found at the bottom of the visual field).

Procedure. All parents signed consent documents while children were fitted with the headcam. If the child was uninterested in wearing the headcam or tried to take it off, the experimenter presented engaging toys to try to draw the child's focus away from the headcam. When the child was comfortable wearing the headcam, the child and caregiver were shown to a playroom for the free-play session. Parents were shown a box containing three pairs of familiar and novel objects. These pairs consisted of a ball paired with a microfiber duster (a "zem"), a toy car paired with a cheese grater (a "manu"), and a brush paired with a back massager (a "tima"). Parents were instructed to play with the object pairs with their child one at a time, "as they typically would." [ˆ3]:The first few participants played with a different pair of objects (a toy cat and pedicure foam piece) that was replaced as some infants chewed persistently on the foam.

All parents confirmed that their child had not previously seen the novel toys and were instructed to use the novel labels to refer to the toys. The experimenter then left the

[1] Detailed instructions for creating this headcam can be found at http://babieslearninglanguage.blogspot.com/2013/10/how-to-make-babycam.html.

playroom for approximately 15-20 minutes, during which a tripod-mounted camera in the corner of the room recorded the session and the headcam captured video from the child's perspective.

Data processing and annotations. Headcam videos were trimmed such that they excluded the instruction phase when the experimenter was in the room and were automatically synchronized with the tripod-mounted videos using FinalCut Pro Software. These sessions yielded 507 minutes (almost a million frames) of video, with an average video length of 14.07 minutes (min = 4.53, max = 19.35).

Posture and caregiver orientation annotations. We created custom annotations to describe the child's physical posture (i.e. standing) and the orientation of the caregiver relative to the child (e.g. far away). The child's posture was categorized as being carried, prone (crawling or lying), sitting, or standing. The caregiver's orientation was characterized as being close, far, or behind the child (independent of distance). For the first two annotations (close/far from the child), the caregiver could either be to the front or side of the child. All annotations were made by a trained coder using the OpenSHAPA/Datavyu software (Adolph, Gilmore, Freeman, Sanderson, & Millman, 2012). Times when the child was out of view of the tripod camera were marked as uncodable and were excluded from these annotations; similarily, times when the child was being carried or the caregivers were out of the frame were marked as uncodable for caregiver orientation. On average, posture or orientation was uncodable from 1-2 minutes of data in each child (seconds excluded from analysis for posture;. M =94.01, SD =225.18; orientation; M =94.01, SD =225.18), and these rates did not vary substantially with the age of the child. To assess the reliability of these annotations, a second coder coded videos from three different children to calculate Cohen's kappa (posture, irr = XX, orientation, irr = XX).

Naming event annotations. One coder listened to all of the audio from the play sessions and marked the exact timestamps whenever one of the novel or familiar objects was named in any instance (e.g., “Look at the [ball]”, “Can you say [zem]?”); a second coder listened to the majority of the play sessions (N=23 sessions) and also annotated all naming events. To asesess reliability, we calculated the proportion of naming events detected by the first coder that were also annotated by a second coder within a sliding window. We found that 82.1% of naming events were detected within a 4 second window (+/- 2s), and 70.9% of namings events were detected within a 2 second window. We also obtained full text transcriptions of the entire play sessions (with time stamps marking 10s intervals). While these full transcriptions are not used in the present analyses, they have been made publically avaliable for future research.

Face and hand detection

We evaluated three detection systems for the ability to measure infants’ access to faces. The first of these is the most commonly-used and widely available face detection algorithm: Viola-Jones (Viola & Jones, 2004). We used this algorithm as a benchmark for performance, as while it can achieve impressive accuracy in some situations, it is notoriously bad at dealing with occluded faces (Scheirer, Anthony, Nakayama, & Cox, 2014). We next tested the performance of two face detectors that both made use of relatively recently developed Convolutional Neural Networks (CNNs) to extract face information. The first algorithm was specifically optimized for face detection, and the second algorithm was optimized to extract pose information of all the individuals in an image, operationalized as information about the position of 18 different body parts. For this second algorithm (OpenPose; Cao et al., 2017), we used the agent’s nose (one of the body keypoints detected) to operationalize the presence of faces, as any half of a face necessarily contains a nose.

The OpenPose detector also provided us with the location of an agent’s wrists, which we used as a proxy for hands for two reasons. First, as we did not capture children’s entire visual field, the presence of a wrist is likely often indicative of the presence of a hand within the field of view. Second, hands are often occluded by objects when caregivers are interacting with children, yet still visually accessible by the child and part of their joint interaction.

Algorithms. Viola Jones, the first face detection system, made use of a series of Haar feature-based cascade classifiers (Viola & Jones, 2004) applied to each individual frame. The second algorithm (based on work by K. Zhang et al. (2016)) uses multi-task cascaded convolutional neural networks (MTCNNs) for joint face detection and alignment, built to perform well in real-world environments where varying illuminations and occlusions are present. We used a Tensorflow implementation of this algorithm available at https://github.com/davidsandberg/facenet.

The CNN-based pose detector (OpenPose; Cao et al., 2017; Simon, Joo, Matthews, & Sheikh, 2017; Wei, Ramakrishna, Kanade, & Sheikh, 2016) provided the locations of 18 body parts (ears, nose, wrists, etc.) and is available at https://github.com/CMU-Perceptual-Computing-Lab/openpose. The system uses a convolutional neural network for initial anatomical detection and subsequently applies part affinity fields for part association, producing a series of body part candidates. The candidates are then matched to a single individual and finally assembled into a pose; here, we only made use of the body parts relevant to the face and hands (nose and wrists), though the entire set of keypoints is publicly avaliable. Each keypoint was accompanied by a confidence score made by the detector.

Detector evaluation. To evaluate face detector performance, we hand-labeled a "gold set" of frames extracted from the video dataset. To account for the relatively rare appearance of faces in the dataset, we hand-labeled two types of samples: a sample containing a high density of faces (half reported by MTCNN, half by OpenPose) and a random sample from the remaining frames. Each sample was comprised of an equal number of frames taken from each child's video. For wrist detections, the "gold set" was constructed in the same manner, except frames with a high density of wrists came only from detections made by OpenPose. Faces were classified as present if at least half of the face was showing; wrists were classified as present if any part of the wrist was showing. Two authors labelled the frames independently and resolved disagreements on a case-by-case basis. Precision (hits / hits + false alarms), recall (hits / hits + misses), and F-score (harmonic mean of precision and recall) were calculated for all detectors.

Results

First, we report the accuracy of the automated detectors, as assessed by comparison to hand-labelled frames from the free-play video dataset described above. We then apply one of these automated detectors (OpenPose) to the entirety of this video dataset, and use these outputs to examine how postural developments influence children's visual access to faces and hands from 8-16 months of age, as well as how access to these social cues changes during labeling events (e.g., do you see the [zem]?). Finally, we apply this same automated detector to another video dataset (Franchak et al., 2017), replicating and extending their findings on the effects of posture on visual access to social information.

Accuracy of automated detections

For face detection, we found that both OpenPose and MTCNN dramatically outperformed ViolaJones, our baseline model, especially with respect to the random sample, where ViolaJones generated a relatively high proportion of both false negatives and false positives. When considering only the composite F-score across all frames, MTCNN slightly outperformed OpenPose (0.89 MTCNN vs. 0.83 OpenPose), and MTCNN and OpenPose performed comparably with the random sample. Generally, MTCNN exhibited higher precision, whereas OpenPose exhibited higher recall, and these differences were most pronounced on the randomly sample frames. In other words, while OpenPose generated slightly more false positives than MTCNN, MTCNN missed several faces that were accurately detected by OpenPose. When we restricted our analysis to high-confidence detections from OpenPose (>.5 confidence; default threshold for visualization), we found very high precision (P = 0.97), but much lower recall (R = 0.64) and thus overall performance (F = 0.77), indicating that these low-confidence detections often indexed actual faces that were in the infant view. Figure 2 shows example successful detections from OpenPose in each age group, and Figure 3 shows examples of missed faces as well as false positives for context.

We next analyzed the viability of OpenPose as a hand detector. Despite the fact that hand detection is a more computationally challenging problem (Bambach, Crandall, Smith, & Yu, 2017), and the fact that we used wrist keypoints as a proxy for hands, OpenPose performed moderately well as a hand detector (F = 0.74). OpenPose achieved relatively high precision – generating relatively few false positives – but showedlow recall on the randomly sampled frames (see Table 1). As with face detections, when we restricted our analysis to high-confidence detections, we found much higher precision (P = 0.95), but much lower recall (R = 0.37) and thus overall performance (F = 0.53).

Thus, one major advantage of OpenPose relative to specialized face detectors, such as MTCNN, is that it allows the analysis of both the faces and hands in the infant view with the outputs of only one algorithm, and analyzing the results of all detections (regardless of confidence) yielded reasonably accurate results. Going forward, we analyze face and wrist detections using all detections from OpenPose, with the caveat that we are likely underestimating the proportion of hands in the dataset given the lower recall for hand detections.

Access to social information during postural developments

Developmental changes in infant posture and caregiver orientation. We report developmental shifts in infants' posture and their orientation relative to their caregiver, consistent with previous literature (Adolph & Berger, 2006; Franchak et al., 2017). The proportion of time infants spent sitting decreased with age, and the proportion of time infants spent standing increased with infants' age. As children got older, their locomotor abilities allowed them to become more independent. Both 8-month-olds and 12-month-olds spent relatively equivalent amounts of time lying/crawling (i.e., "prone") which was markedly decreased in the 16-month-olds, who spent most of their time sitting or standing (see Figure 4). We also observed changes in infants' orientation relative to their caregivers: the 8-month-olds spent more time with their caregiver behind them supporting their sitting positions than did children at other ages (see Figure 4). However, we also saw considerable variability across children: some infants spent almost their entire time sitting at a close distance from their caregiver, whereas others showed more considerable variability (see Figure 5).

Access to faces and hands. First, we examined the proportion of face and hand detections as a function of infants' age without considering their posture (see Figure 6). While faces tended to be in the field-of-view overall more often than hands, infants' head-mounted cameras were angled slightly upward to capture the presence of faces, and hand detections suffered from somewhat lower recall than face detections. We thus only considered differences in the relative proportion of faces or hands in view as a function of age, posture, and orientation, rather than comparing them directly. Overall, we observed that 12-month-olds appeared to have visual access to slightly fewer faces than 8 or 16-month-olds, creating a slight U-shaped function in face detections; conversely, hand detections were showed a slight increase across this age range, as reported in prior work (Fausey et al., 2016).

However, these age-related trends were much smaller than the effect of infant's postural developments on infants' visual access to faces and hands. Infants' in-the-moment posture was a major factor both in how many faces and hands were in view during the play session, as was their orientation relative to their caregiver. This was quantified using two generalized linear mixed-effect models estimating the proportion of faces and hands that were in view, with orientation, posture, their interaction, and scaled participant's age as fixed effects, and with random slopes for infants' orientation and posture. A summary of the coefficients of the models can be found in Tables 2 and 3.

The interaction between infants' posture and their caregiver's orientation had the most dramatic effect on the social information in view. When caregivers were behind their infants, supporting their infants' sitting or standing positions, infants saw fewer faces. When caregivers were relatively close to their infants, infants who were sitting or standing had more faces in view (Face detections; infant sitting and CG close, b=0.99, SE = 0.07, Z = 14.54, P = 0; infant standing and CG close, b=1.36, SE = 0.08, Z = 16.95, P = 0 than infants who were lying down/crawling (i.e. prone). When caregivers were far away from their infants, face detections were similarity higher (Face detections; infant sitting and CG far, b = 0.75, SE = 0.07, Z = 10.48, P = 0; infant standing and CG far, b = 1.40, SE = 0.09, Z = 16.16, P = 0. Overall, we found that age did not remain a significant predictor in accounting for the faces in view when modeling infants posture, their orientation to their caregiver, and the interaction between these two factors (Face detections; Age (scaled), b = 0.05, SE = 0.12, Z = 0.39, P = 0.70).

We found a similar pattern of results for wrist detections, even though there were fewer wrist detections overall in the dataset. Infants saw fewer wrists when caregivers were behind their infants, supporting their infants' sitting or standing positions vs when caregivers were relatively closer to their infants (Hand detections; infant sitting and CG close, b=0.31, SE = 0.08, Z = 3.72, P = 0.00; infant standing and CG close, b=0.76, SE = 0.09, Z = 8.46, P = 0 than infants who were lying down/crawling (i.e. prone). Wrist detections were clearly highest when caregivers were far away from their infants (Wrist detections; infant sitting and CG far, b=0.13, SE = 0.10, Z = 1.27, P = 0.21; infant standing and CG far, b=1.76, SE = 0.11, Z = 15.87, P = 0. As with faces, age did not remain a significant predictor in these models (Wrist detections; Age (scaled), b = 0.16, SE = 0.11, Z = 1.48, P = 0.14).

We directly examined the contributions of posture and orientation by fitting a reduced version of the full model (Nakagawa & Schielzeth, 2013) without their fixed effects (both models were run with the maximal random effects structure). The fixed effects in a model with only the age of the participants accounted for relatively little variance in the proportion of faces (marginal $R^2 < 0.01$) or hands in view (marginal $R^2 = 0.02$). However, when adding infants' posture and orientation to their caregiver to the model (and their interaction) the marginal R^2 were higher for both faces (marginal $R^2 = 0.28$) and wrists (marginal $R^2 = 0.31$).

Overall, these results suggest that infants' visual access to social information is largely modulated by their posture and orientation to their caregiver, which is in turn a function of their general locomotor development.

Social information during naming events. Our play session was designed to provide parents with opportunities to label objects – both familiar and novel – such that we could examine whether children saw different kinds of social information around naming events. In a set of exploratory analyses, we thus analyzed how face and hand detections changed during object labeling events relative to baseline, analyzing a four-second window around each labeling event (e.g., "Look at the [zem]!"). Every utterance of one of the labelled objects (e.g., "ball") was counted as a "labeling event"; timestamps of the beginning of each word were hand-annotated and synchronized with the frame-by-frame detections.

To assess whether there were differences in the social information in view during naming events, we first calculated the proportion of detections that were in view during this four second window, and averaged across naming events for each subject as function of whether the named object as a novel or a familiar objet; this was then then compared to the baseline proportion of faces in view for each subject in linear mixed-effect models, with random effects of subjects and fixed effects of (scaled) age.

Face detections were numerically slightly higher around these novel naming events relative to baseline, with similar effects across age groups (8 m.o., $M_{fam-baseline} = $ =0.02, 12 m.o., $M_{nov-baseline}$ =0.05, 16 m.o. $M_{nov-baseline}$ =0.02) however, they were not higher during familiar naming events vs. baseline (8 m.o., $M_{fam-baseline}$ =0.00, 12 m.o., $M_{fam-baseline}$ =0.00, 16 m.o. $M_{fam-baseline}$ =0.02, Conversely, wrist detections were higher during both familiar (8 m.o., $M_{fam-baseline}$ =0.07, 12 m.o., $M_{fam-baseline}$ =0.04, 16 m.o. $M_{fam-baseline}$ =0.08) and novel naming events relative to baseline across all age groups (Novel; 8 m.o., (8 m.o., $M_{fam-baseline}$ =0.09, 12 m.o., $M_{fam-baseline}$ =0.08, 16 m.o. $M_{fam-baseline}$ =0.08). These results were confirmed by a linear mixed-effect model with scaled aged as a fixed effect and random intercepts for each subject (Wrist detections; familiar objects vs. baseline, b=0.06, SE = 0.01, t = 4.33, P = 0.00'; Novel objects vs. baseline, b=0.08, SE = 0.02, t = 5.21, P = 0.

Overall, these exploratory results suggest that children may tend to see more hands around naming events. This finding is consistent with the possibility that caregivers may change how they interact with their infant when presenting them with objects (Gogate, Bahrick, & Watson, 2000, Gogate, Bolzani, and Betancourt (2006)). For example, caregivers may tend to simultaneously name objects when demonstrating their affordances or simply when pointing to them. In turn, infants may be sensitive to these naming events and orient their attention towards their caregiver, consistent with other accounts positing infants' sensitivity to social cues in early word learning environments (Yurovsky, 2018, Yurovsky and Frank (2017), Frank, Simmons, Yurovsky, and Pusiol (2013)).

Extension to Franchak et a., 2017

In the present work, we found that infant's in-the-moment posture changed with their age, as did infants' orientation relative their caregiver. In a related study with 12-month-olds, Franchak et al. (2017) also found that infants' in-the-moment posture changed the proportion of time infants spent looking at faces. Here, we sought to replicate these findings using our automated methodology (OpenPose detections), using the footage from their head-mounted cameras (D. A. Simon, Gordon, Steiger, & Gilmore, 2015) for two reasons. First, we sought to validate our novel method, which could fail to generalize to scenes from these more complex environments, where detecting faces and hands could arguably be a much harder task. Second, we sought to replicate the effects of infants' in-the-moment posture on differences in visual access to hands in an independent dataset.

Comparison between video datasets. First, despite having different head-mounted camera setups, the view angle of the two head-mounted cameras were only slightly different (52.2° horizontal by 42.2° in Franchak et al. (2017), here 64° horizontal by 46° vertical). However, the environment that infants were immersed in with their caregivers (and experimenters) was much larger and more varied than the play room used in the present dataset, containing multiple structures and toys in different parts of the room for infants to climb, explore, and interact with, and infants were unconstrained and allowed to freely wander the room. In contrast, the play room used in Study 1 was relatively small (approximately 10' x 10') and was set-up with focused play with the pairs of novel and familiar objects. Further, multiple people were present during the play session in Franchak et al. (2017)– including their caregiver and two experimenters – whereas in Study 1 the experimenters left the room during the play session.

Replication using automated detections. Overall, we found convergence between our two methodologies, replicating the main results from Franchak et al. (2017), and finding that the proportion of faces detected was greater when infants were sitting or standing vs. prone. We found this result regardless of whether we used all detections (see Figure 9, panel A; (proportion of frames with face detections; Prone: M =0.13%, Sitting: M =0.20%, Upright: M =0.22%) or restricted our analyses to high-confidence detections (percentage of frames with face detections; Prone: M =0.03, Sitting: M =0.06, Upright: M =0.05). These results were confirmed in generalized linear-mixed models, as in the first study (Sitting vs. Prone, b=0.34, SE = 0.02, Z = 23.37, P = 0. Upright vs. Prone, b=0.38, SE = 0.02, Z = 24.01, P = 24.01).

The most noticeable difference between our analysis methods is the overall proportion of frames with faces detected by OpenPose vs. the proportion of frames in which infants were judged to be looking at faces in Franchak et al. (2017). Across the entire session in Franchak et al. (2017), infants looked at faces on 4.7% of frames (Franchak et al., 2017). When we used all detections from OpenPose, we found a higher proportion of faces: M = 0.22, or 21.80%. When we restricted our results to only high-confidence detections, we found M = 0.06, closer to the original values reported by Franchak et al. (2017). However, the above analyses on the accuracy of these two methods suggest that high-confidence detections dramatically underestimate the number of faces in view. Thus, we suspect that these differences are due to both (1) the fact head-mounted eye-trackers may sometimes underestimate the proportion of faces attended due to calibration issues, and that (2) OpenPose may sometimes detect faces that are in view for infants but that infants may not be foveating.

We also found that infants' in-the-moment posture also modulated the proportion of hands that were in view (i.e., wrist detections), though these were not originally annotated in Franchak et al. (2017) (see Figure 9, (proportion of frames with wrist detections; Prone: M =0.15; Sitting: M =0.21; Upright: M =0.22). These results were confirmed in generalized linear-mixed models, as in the first study (Sitting vs. Prone, b=0.27, SE = 0.01, Z = 19.09, P = 0. Upright vs. prone, b=0.30, SE = 0.02, Z = 19.73, P = 19.73).

Overall, these analyzes extend and validate previous work, replicating the results in Study 1 that infants' in-the-moment posture modulated the proportion of hands in view, and suggesting that posture is a major factor that structures infants' access to visual information broadly construed.

General Discussion

What social cues do infants see as they learn language, and how does this change as they grow and are start to locomote themselves in their environments? We explored this question using video data from head-mounted cameras from two naturalistic datasets of parent-child interactions: a cross-sectional database of naturalistic play sessions from 8-16 month-olds with sets of novel and familiar toys, and database of head-mounted camera videos from one-year-olds who explored a large play area from Franchak et al. (2017). To analyze these datasets, we developed a novel method using a pose detection model to automate the annotation of the social information in the infant view, here operationalized as the presence of the faces and hands of their caregiver; these annotations were then synced with manual annotations of infants in-the-moment posture from third-person videos.

Despite being trained on the adult perspective, the pose detector we used (OpenPose, Cao et al. (2017)) was able to generalize relatively well to the infant viewpoint, achieving comparable precision and accuracy as a face detector relative to a state-of-the-art model optimized specifically for detecting faces in natural scenes (MTCNN, K. Zhang et al. (2016)]. While OpenPose had relatively low recall as a hand detector – missing some hands that were in the infant view – it made comparable rates of false alarms. In both cases, we found that overall performance was maximized when all detections were included, regardless of their confidence, suggesting that some low-confidence face and hand detection still index actual faces and hands that were seen by infants. Thus, while imperfect, we suggest that OpenPose can be applied to infant egocentric videos for the extraction of the social information in the infant viewpoint, reducing the burden of manual annotations and promoting the reusability of rich video datasets for further analyses. The use of this automated methodology allowed us to easily annotate the entirety of our dataset – additionally analyzing the social information around naming events – and to re-analyze the data from Franchak et al. (2017) to replicate our findings in a very different kind of play session.

Broadly, our results replicate and extend previous work, first by showing systematic changes in infants' in-the-moment posture and their orientation relative to their caregivers (Adolph & Franchak, 2017); older children spent more time standing and less time sitting, and older infants' caregivers spent less time supporting their standing or sitting postures. Motor development changes dramatically at the same time that children are breaking into language learning. Using these automated detections, we found that infants' changing posture and orientation to their caregiver jointly shaped the amount of social information that was in their view during one-on-one play sessions with their caregivers: Children saw the most faces/hands when they were sitting or standing and close to their caregiver vs. crawling or prone. These same findings were recapitulated in a second dataset collected by Franchak et al. (2017) with one-year-olds: sitting and upright infants saw more faces–and hands–than infants who were prone. Motor development appears to modulate how infants experience their visual world and the social information in it.

While exploratory, our results also suggest that infants saw a greater proportion of hands during naming events, hinting that children may have been orienting towards their caregiver when they heard labels for objects. While this effect was not present for faces, other work (Yoshida & Smith, 2008; Yu & Smith, 2013, 2017), including Franchak et al. (2017), has found that infants spend much more time looking at the toys vs. their caregiver's faces during these play sessions, and highlighted the importance of hand-following as a component of joint attention (Yu & Smith, 2017). Further, given that there were only two possible referents in the room at a time—-and one of them was always a familiar category (i.e., car, kitty) – this particular play session did not present many opportunities where children would need to use gaze cues to disambiguate referents.

Importantly, however, all of these findings come from observational, in-lab datasets, posing limits on their generalizability. Future work is needed to relate these the slices of experience captured within in-lab play sessions and infants' everyday experiences (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Fausey et al., 2016). More broadly, though observational findings allow us to document developmental change and identify potential causal pathways, they cannot confirm them. Locomotive abilities are of course only part of a cascade of changes in infants' abilities and experiences, and these analyzes document only a fraction of this broader, multifaceted trajectory. As children grow and change, the activities in which they engage with their caregivers are likely to also vary, leading to differences in the distribution of social cues that they experience that may not be captured in these in-lab play sessions.

Nonetheless, we think that this approach hold promise for documenting these developmental trajectories and for generating hypotheses about potential causal pathways. Indeed, understanding the relationship between different domains of developmental changes in naturalistic contexts has been a persistent challenge for developmental psychology. However, the field of computer vision has advanced dramatically in recent years, creating a new generation of algorithmic tools that deal better with noisier, more complicated datasets and extract richer information. We hope that these new tools can now be leveraged to examine the consequences of the changing infant perspective for linguistic, cognitive, and social development.

Acknowledgements

Thanks to Kaia Simmons, Kathy Woo, and Aditi Maliwal for help in recruitment, data collection, and annotation. This work was funded by a Jacobs Foundation Fellowship to MCF, a John Merck Scholars award to MCF, and NSF #1714726 to BLL. An earlier version of this work was presented to the Cognitive Science Society in Frank et al. (2013) and Sanchez, Long, Kraus, and Frank (2018).

References

Adolph, K. E., & Berger, S. E. (2006). Motor development. *Handbook of Child Psychology*.

Adolph, K. E., & Franchak, J. M. (2017). The development of motor behavior. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1-2), e1430.

Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry*, *23*(3), 244–247.

Adolph, K. E., Vereijken, B., & Denny, M. (1998). Roles of variability and experience in development of crawling. *Child Development*, *69*(1299), 312.

Bambach, S., Crandall, D. J., Smith, L. B., & Yu, C. (2017). An egocentric perspective on active vision and visual object learning in toddlers. In *Proceedings of the seventh joint ieee conference on development and learning and on epigenetic robotics*.

Brooks, R., & Meltzoff, A. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220.

Bruner, J. (1975). From communication to language: A psychological perspective. *Cognition*, *3*(3), 255–287.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. In *CVPR*.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society* *for Research in Child Development*, *63*(4).

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Phil. Trans. R. Soc. B*, *372*(1711), 20160055.

Cummings, M., Van Hof-Van Duin, J., Mayer, D., Hansen, R., & Fulton, A. (1988). Visual fields of young children. *Behavioural and Brain Research*, *29*(1), 7–16.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605.

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.

Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2017). See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental Science*.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*(6), 1738–1750.

Frank, M. C., Simmons, K., Yurovsky, D., & Pusiol, G. (2013). Developmental and postural changes in children's visual access to faces. In *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 454–459).

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894.

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word–object relations. *Infancy*, *9*(3), 259–288.

Gredeback, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science*, *13*(6), 839–848.

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.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2017). Why are faces denser in the visual experiences of younger than older infants? *Developmental Psychology*, *53*(1), 38.

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.

Kietzmann, T. C., McClure, P., & Kriegeskorte, N. (2018). Deep neural networks in computational neuroscience. *BioRxiv*, 133504.

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*(4), 1503–1518.

Lampl, M. (1993). Evidence of saltatory growth in infancy. *American Journal of Human Biology*, *5*(6), 641–652.

Libertus, K., & Violi, D. A. (2016). Sit to talk: Relation between motor skills and language development in infancy. *Frontiers in Psychology*, *7*, 475.

Mayer, D., Fulton, A., & Cummings, M. (1988). Visual fields of infants assessed with a new perimetric technique. *Investigative Ophthalmology & Visual Science*, *29*(3), 452–459.

Moore, C., Dailey, S., Garrison, H., Amatuni, A., & Bergelson, E. (2019). Point, walk, talk: Links between three early milestones, from observation and parental report. *Developmental Psychology*.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Patrick, S. K., Noah, J. A., & Yang, J. F. (2012). Developmental constraints of quadrupedal coordination across crawling styles in human infants. *Journal of Neurophysiology*, *107*(11), 3050–3061.

Peterson, J. C., Abbott, J. T., & Griffiths, T. L. (2018). Evaluating (and improving) the correspondence between deep neural networks and human representations. *Cognitive Science*, *42*(8), 2648–2669.

Sanchez, A., Long, B., Kraus, A. M., & Frank, M. C. (2018). Postural developments modulate children's visual access to social information.

Scheirer, W. J., Anthony, S. E., Nakayama, K., & Cox, D. D. (2014). Perceptual annotation: Measuring human vision to improve computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(8), 1679–1686.

Simon, D. A., Gordon, A. S., Steiger, L., & Gilmore, R. O. (2015). Databrary: Enabling sharing and reuse of research video. In *Proceedings of the 15th acm/ieee-cs joint conference on digital libraries* (pp. 279–280).

Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

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.

VanRullen, R. (2017). Perception science in the age of deep neural networks. *Frontiers in Psychology*, *8*, 142.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154.

Walle, E. A. (2016). Infant social development across the transition from crawling to walking. *Frontiers in Psychology*, *7*, 960.

Walle, E. A., & Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Developmental Psychology*, *50*(2), 336.

Wei, S.-E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. In *CVPR*.

Yoshida, H., & Smith, L. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*, 229–248.

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).

Yu, C., & Smith, L. B. (2017). Hand–eye coordination predicts joint attention. *Child*

Development, *88*(6), 2060–2078.

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, *50*, 73–79.

Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*, *20*(2), e12349.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, *23*(10), 1499–1503.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1452–1464.

Group	N	% incl.	Avg age	Avg video length (min)
8 mo.	12	0.46	8.71	14.41
12 mo.	12	0.40	12.62	12.71
16 mo.	12	0.31	16.29	15.10

Table 1

Exclusion rates and summary demographics for the infants included in the study.

Algorithm	Sample Type	P	R	F
MTCNN-Faces	High density	0.89	0.92	0.90
MTCNN-Faces	Random	0.94	0.62	0.75
OpenPose-Faces	High density	0.78	0.93	0.84
OpenPose-Faces	Random	0.72	0.80	0.76
ViolaJones-Faces	High density	0.96	0.44	0.60
ViolaJones-Faces	Random	0.44	0.38	0.41
OpenPose-Wrists	High density	0.66	1.00	0.79
OpenPose-Wrists	Random	0.88	0.29	0.44

Table 2

Detector performance on both high density samples (where proportion of targets detected was high) and random samples (where frames were randomly selected). P, R, and F denote precision, recall, and F-score, respectively.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-4.29	0.18	-23.87	0.000
Sit	0.59	0.18	3.20	0.001
Stand	0.53	0.30	1.76	0.079
Close	0.53	0.19	2.75	0.006
Far	0.15	0.26	0.56	0.574
Age (Scaled)	0.16	0.11	1.48	0.140
Sit*Close	0.31	0.08	3.72	0.000
Stand*Close	0.76	0.09	8.46	0.000
Sit*Far	0.13	0.10	1.27	0.205
Stand*Far	1.76	0.11	15.87	0.000

Table 3

Model coefficients from a generalized linear mixed models predicting the proportion of wrists seen by infants.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-3.25	0.18	-18.04	0.000
Sit	0.05	0.17	0.33	0.744
Stand	-0.24	0.18	-1.34	0.180
Close	0.01	0.19	0.07	0.942
Far	0.45	0.27	1.63	0.102
Age (Scaled)	0.05	0.12	0.39	0.697
Sit*Close	0.99	0.07	14.54	0.000
Stand*Close	1.36	0.08	16.95	0.000
Sit*Far	0.75	0.07	10.48	0.000
Stand*Far	1.40	0.09	16.16	0.000

Table 4

Model coefficients from a generalized linear mixed models predicting the proportion of faces seen by infants.

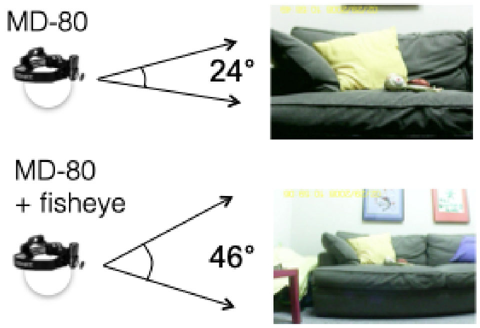

Figure 1. Vertical field of view for two different headcam configurations (we used the lower in our current study).

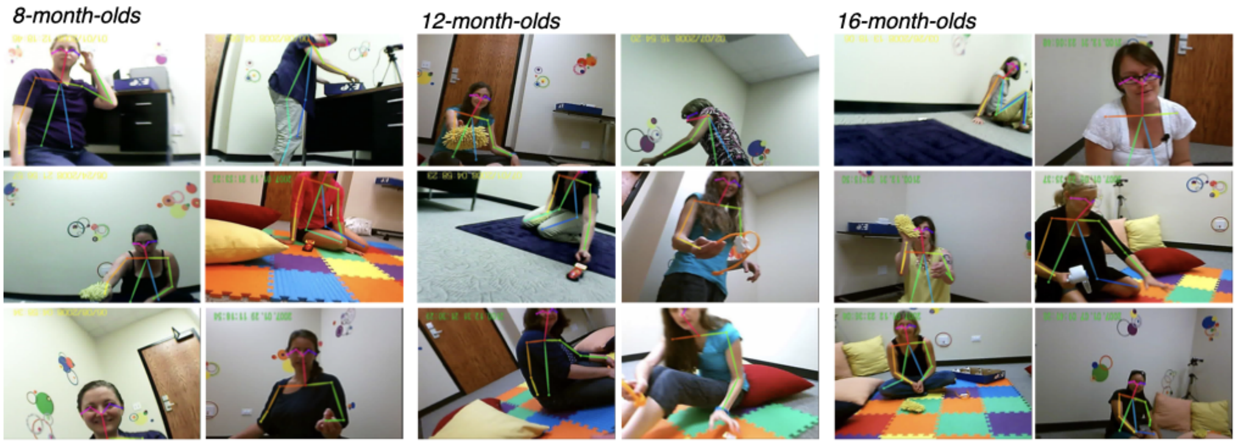

Figure 2. Example detections made by OpenPose from children in each age group.

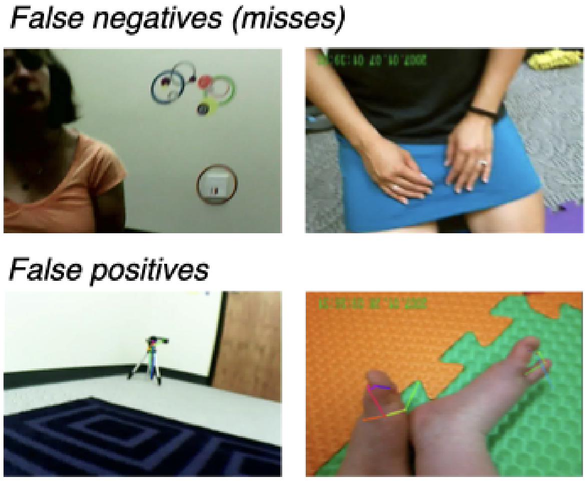

Figure 3. Example failed detections from OpenPose, showing both false positives and false negatives (i.e. missed detections).

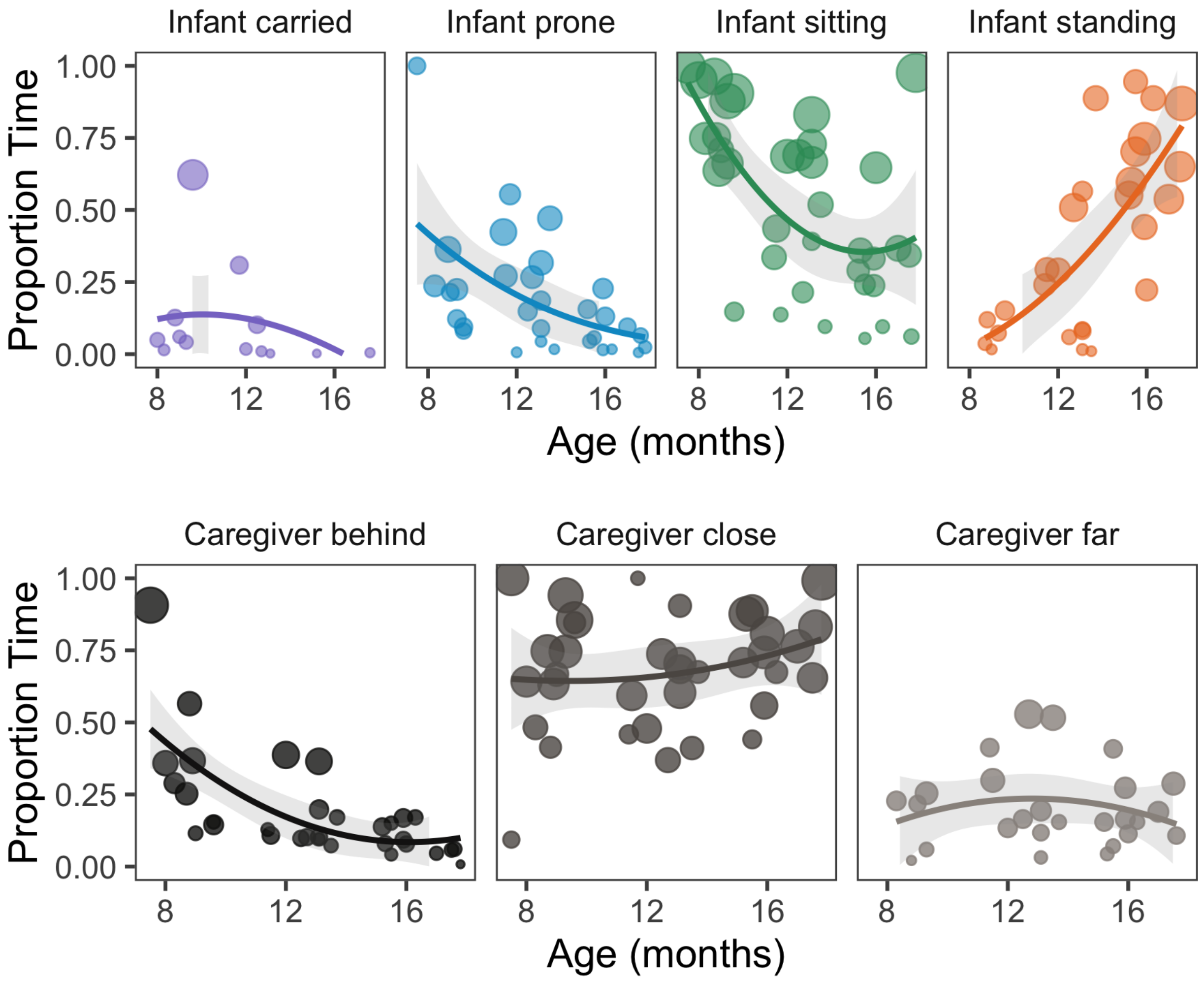

Figure 4. Proportion of time spent by each infant in different postures and orientations relative to their caregivers (CG); times where posture was not codable are ommitted for visualization purposes

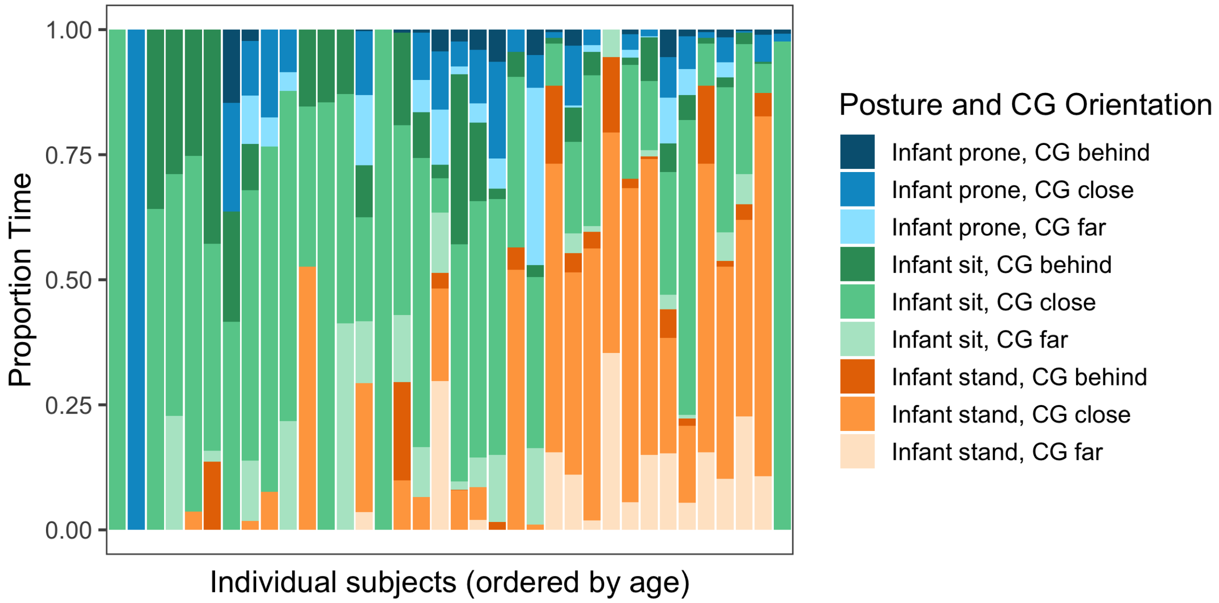

Figure 5. Proportion of time spent by each infant in different postures and orientations relative to their caregivers (CG); times when infant was carried or when posture/orientation were not codable are ommitted for visualization purposes

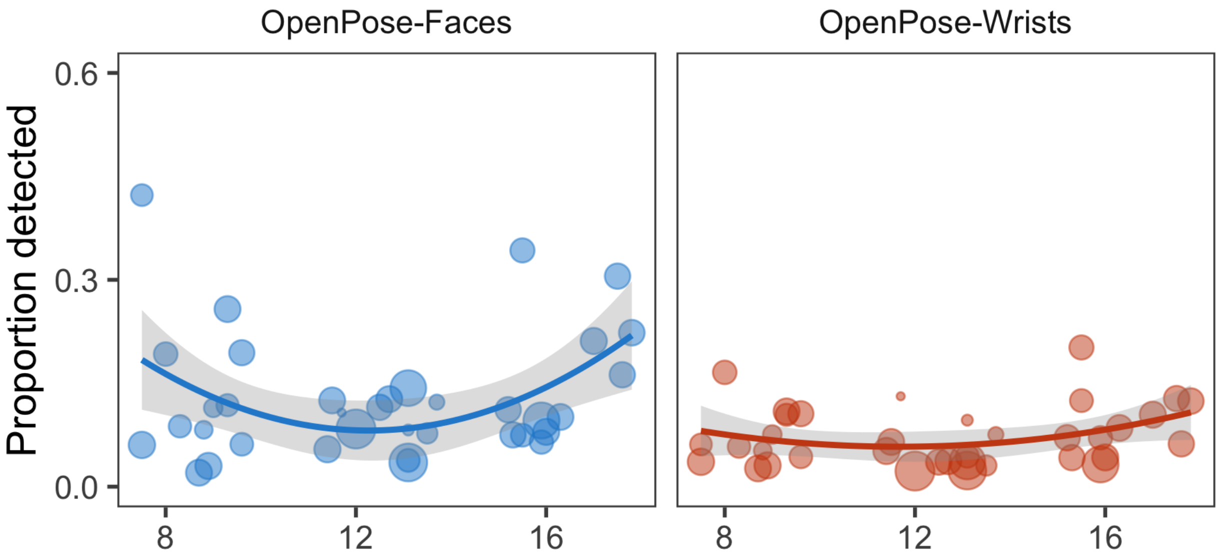

Figure 6. Proportion of faces (left) and wrists (right) detected by the OpenPose model as a function of child's age. Larger dots indicate children who had longer play sessions and thus for whom there was more data.

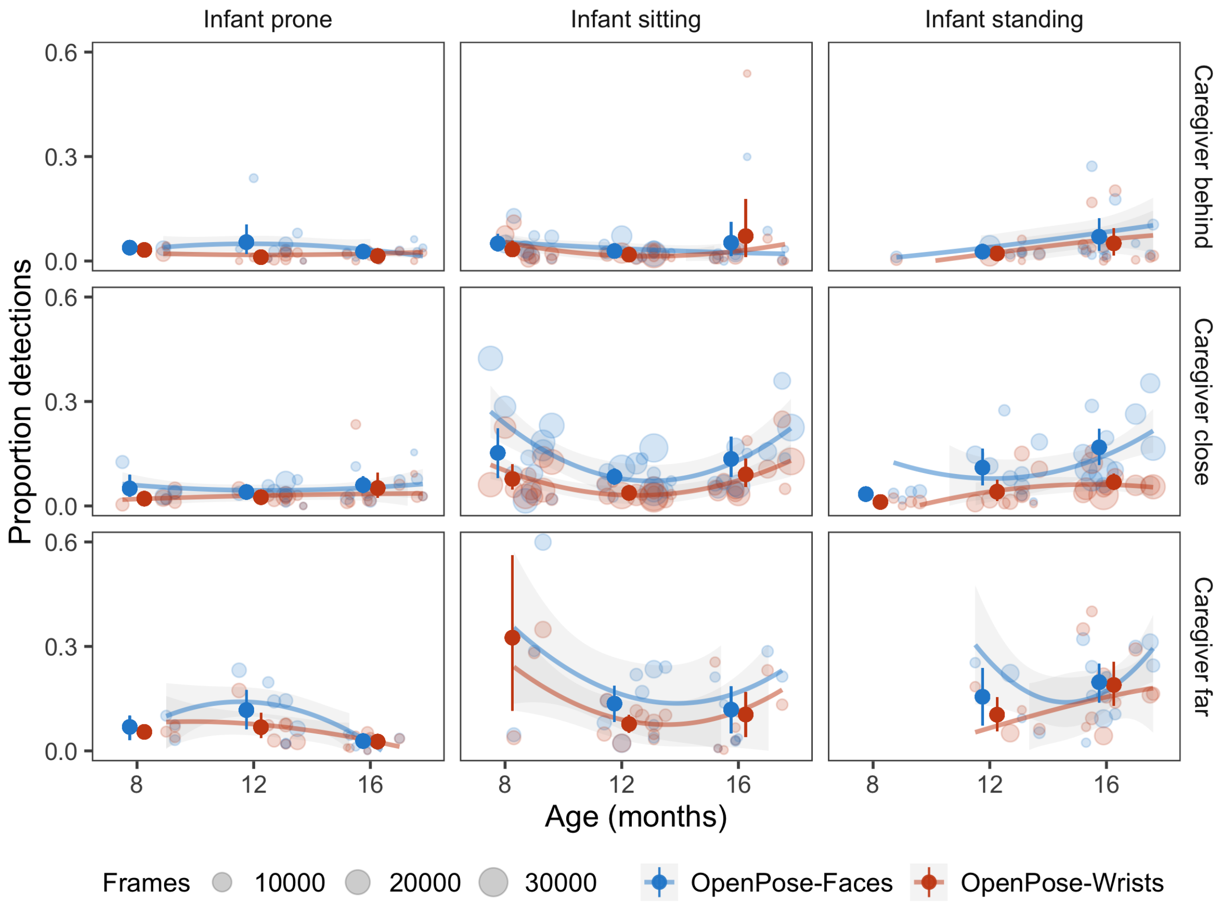

Figure 7. Proportion of face / wrist detections by children's age, their posture, and their caregivers orientation. Data points are scaled by the amount of time spent in each orientation/posture combination; times when posture/orientation annotatinos were unavaliable or the infant was carried are not plotted. Error bars represent 95% bootstrapped confidence intervals

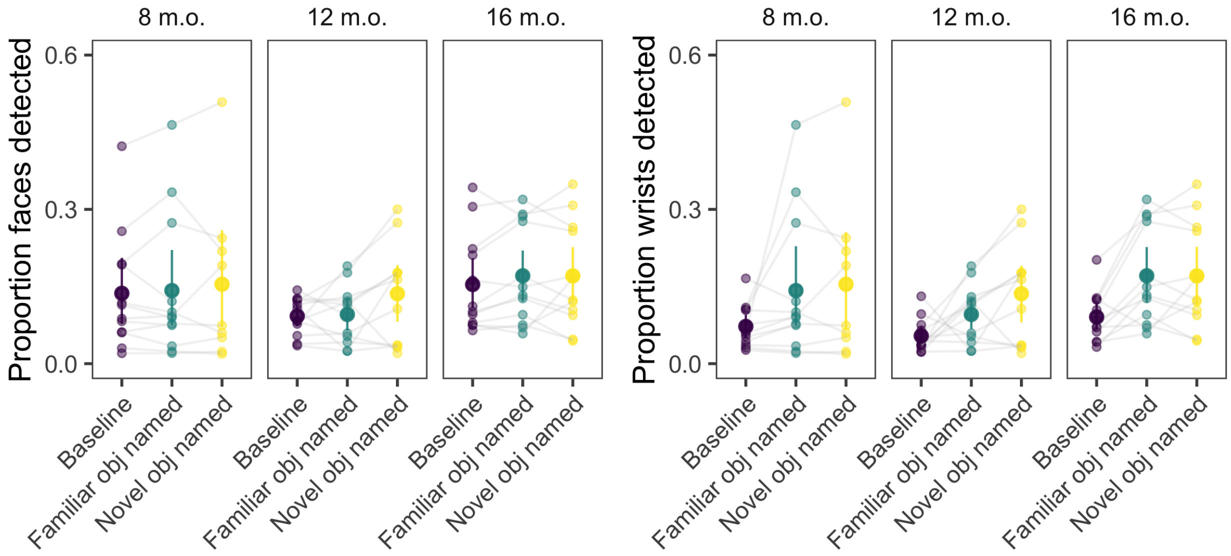

Figure 8. Proportion of face / wrist detections during naming events (+/- 2 seconds around label) for familiar and novel objects; these rates are put into context relative to baseline. Error bars represent 95% bootstrapped confidence intervals. Grey lines connect points from individual subjects.

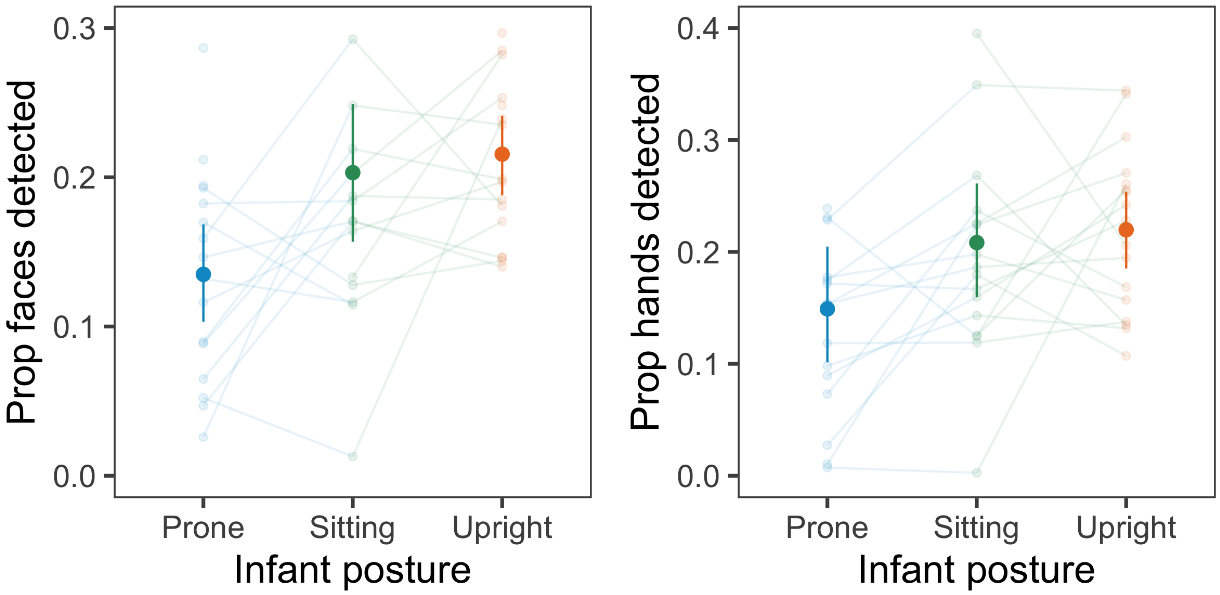

Figure 9. Proportion of face / wrist detections for 12-month-olds in Franchak et al., 2017 as a function of children's in-the-moment posture. Error bars represent 95 percent bootstrapped confidence intervals.