

# Action prediction during real-time social interactions in infancy

**Claire Monroy (claire.monroy@osumc.edu)**

**Chi-hsin Chen (chi-hsin.chen@osumc.edu)**

**Derek Houston (derek.houston@osumc.edu)**

Department of Otolaryngology—Head and Neck Surgery

The Ohio State University Wexner Medical Center, Columbus, OH 43210 USA

**Chen Yu (chenyu@indiana.edu)**

Department of Psychological and Brain Sciences

Indiana University, Bloomington, IN 47405 USA

## Abstract

Developmental theory considers action prediction as one of several processes involved in determining how infants come to perceive and understand social events (Gredebäck & Daum, 2015). Action prediction is observed from early in life and is considered an important social-cognitive skill. However, knowledge about infant action prediction is limited to evidence from screen-based eye-tracking tasks. Little is known about action prediction in real-life action contexts. Our aim in the current study was to provide new evidence on whether and how infants anticipate actions in free-flowing parent-child interaction. Using dual head-mounted eye-tracking, we analyzed infants' visual anticipations of their parents' reaching actions while they played with objects together. Findings reveal that infants anticipate their parents' actions at a rate higher than would be expected by chance.

**Keywords:** dual head-mounted eye-tracking; action prediction; parent-child interaction; social-cognitive development

## Introduction

Action prediction refers to the ability to anticipate the outcome or endpoint of another person's goal-directed action (Flanagan & Johansson, 2003). This ability serves several important perceptual and cognitive functions: in a noisy and dynamic environment, anticipation allows the observer to direct visual attention to where important events will occur next (Gredebäck, Johnson, & von Hofsten, 2010). Action prediction also facilitates smooth, coordinated interactions. For instance, a simple interaction such as passing an object to another person requires planning a motor response at a precise moment in time and space to grasp the object successfully. Anticipating the other person's action and gazing to the location their hand will go next allows this kind of joint coordination to take place (Knoblich & Sebanz, 2012). For infants, whose developing system is solving the challenge of integrating their motor and visual systems, action prediction is an emerging skill (Falck-Ytter, Gredebäck, & von Hofsten, 2006; Kanakogi & Itakura, 2011; Monroy, Gerson, & Hunnius, 2017). In the current study, we investigated action prediction in 9-month-old infants, who are at the cusp of acquiring new fine motor skills and demonstrating rapid growth in their social-cognitive skills.

Prior research has demonstrated that infants exploit multiple cues to anticipate observed actions. For instance, infants can use kinematic cues from movement trajectories (Rosander & von Hofsten, 2011; Stapel, Hunnius, & Bekkering, 2012), the statistical regularities in familiar action sequences (Monroy, Gerson, & Hunnius, 2017), and knowledge about an actor's goal (Woodward, 1998). This ability develops within the first year of life: at 12 months, but not at 6 months, infants can anticipate unambiguous reaching actions (Falck-Ytter et al., 2006). By 9 months of age, infants can predict the endpoints of simple reaching actions based on motor cues from pincer and palmar grasps (Monroy et al., 2017; Senna et al., 2016).

The research described above is exclusively based on evidence from tightly controlled, yet artificial reaching paradigms. These paradigms have been useful in refining current theories about infants' action perception (Gredebäck & Daum, 2015). However, little is currently known about action prediction abilities 'in the wild', as infants interact with others while freely moving about in the environment. It is unknown whether infants' anticipatory behavior in laboratory contexts would generalize to the messier, more complex action contexts of real life. Here, we aimed to provide new evidence for whether and how frequently infants predict their parents' actions during free-flowing parent-child play.

In real-life contexts such as toy play, infants spend a great deal of time engaged with objects (e.g., almost 90% of the time; Yuan, Xu, Yu, & Smith, 2019) and their visual attention is characterized by long fixations to objects they are holding themselves (Yu & Smith, 2013). Based on these findings from recent head-mounted eye-tracking studies, our first question was whether infants do anticipate others' actions in real life, as they do in controlled laboratory contexts. If so, our second aim was to identify the frequency with which they do so and whether this frequency occurs at a rate higher than would be expected by chance. In the current study, we quantified the proportion of anticipated reaching actions during parent-child play and compared these to chance proportions.

To examine further the contexts in which action prediction can occur during parent-child play, we also analyzed infants' visual attention and manual activity during parents' reaching actions. For instance, to make a successful anticipation, do infants need to be disengaged from other

non-target objects? Do they exploit ostensive cues by attending to their parent’s face (Senju & Csibra, 2008)? Given the limitations of infants’ visual attention and their tendency to focus on their own manual actions at this age (Yu & Smith, 2017), one possibility is that infants demonstrate anticipations when they are less active themselves (i.e., better opportunity to anticipate) or when they are more socially engaged with their parent (e.g., more face looking).

## Method

### Participants

The sample consisted of 32 parent-infant dyads (mean age = 9.3 months, range = 9-9.7; 18 females). All children were born full-term and had no developmental diagnoses.

### Procedure

Infants and parents were seated at a child-sized table across from one another. Both dyad members were fitted with head-mounted eye-trackers from Positive Science, LLC (Figure 1). Each eye-tracker has an infrared camera that records the right eye and a head camera that records the field of view. Two additional cameras recorded a third-person view of each dyad member. All six cameras recorded at 30Hz and were synchronized offline using custom-written Python scripts.



Figure 1: Experimental setup. A parent and her infant are seated across from one another playing with familiar objects. The crosshair indicates the estimated gaze direction.

To calibrate the eye-trackers, an engaging toy was placed in 15 unique locations on the tabletop to capture the infant’s attention. Parents were instructed to attend as well. This phase was used for offline calibration using Yarbus software by marking the locations on the corresponding video frames when the eye was directed at the target.

Following calibration, participants were presented with six familiar, engaging toys (a car, cup, a train, a duck, a plane, and a boat). Toys were grouped into two sets of three, with each set containing one red, one green and one blue toy. Parents were instructed to play with their infants “as they normally would at home”. Dyads played with each toy set twice for 90 seconds, yielding six possible minutes of interaction. The order of toy sets was counterbalanced across dyads.

### Data processing

After offline calibration, gaze direction was superimposed onto the head camera recording with a

crosshair, yielding an additional recording of the calibrated gaze. All camera recordings were then exported into a series of single frames. Each camera contributed a maximum of 10,800 frames per dyad (six minutes of recording at 30 frames per second).

Infants’ gaze direction and parents’ reaching actions were then manually coded frame-by-frame. For gaze, two independent coders used frames from the calibrated recording to determine whether the crosshair fell within one of four regions of interest (ROIs): the three novel objects and the parent’s face. Frames were excluded whenever the eye-tracker failed to capture the eye (e.g., the child knocked the camera out of place), in between trials, or whenever the child was off-task. The second coder annotated a random 10% of the frames. Reliability ranged from 82-95% (Cohen’s kappa = .81).

Additional coders annotated parent reaching actions: for each frame, the coder determined whether the parent was reaching for an object and, if so, which one. Reaching was defined as any movement towards an object that ended when contact was made. Right and left hands were coded separately and then merged to yield one data stream.

To identify infants’ action prediction—anticipatory looks to the targets of their parents’ reaching actions—the two data streams from the infant gaze and the parent reaching were aligned. Action prediction was defined as a gaze to an object that occurred after the onset of a parent reach to that same target, but before the reach was completed (Figure 2). This represents the time window in which the infant had enough information to predict the observed action, but before the hand reached the target. The number of anticipations per interaction was then summed and divided by the total number of valid parent actions to yield the proportion of anticipated actions.

### Example data streams

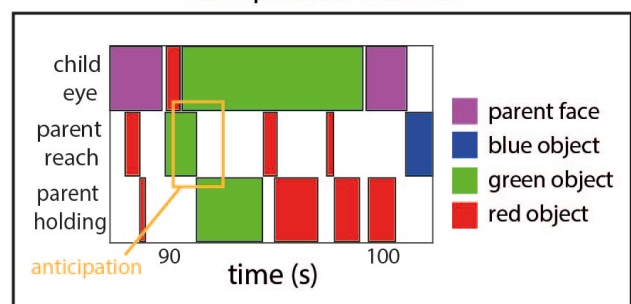


Figure 2: A sample of the aligned gaze and reaching data streams from a representative dyad. The yellow box highlights an example of an anticipation: the infant looks to the green object after the reach onset and prior to the end of the reach. Parent holding is included here for visualization purposes.

Not all parent reaches represented fair opportunities for anticipation. To estimate rates of anticipation out of the child’s actual opportunities to anticipate—rather than total

number of reaches—we categorized all parent reaching actions as *valid* or *invalid* opportunities (Table 1).

Table 1: Criteria for categorizing parent reaches as *invalid* opportunities to anticipate.

Criterion
1. <200ms (to account for the time needed to program an eye movement)
2. Subsequent contacts in cases of multiple object contacts (e.g., tapping or switching object from one hand to another)
3. Infant reaching for the object at the same time
4. Experimenter was reaching for or touching the object at the same time
5. Both parent and object were entirely out of the child's view for the entire duration of the reach (e.g., child's eyes were closed, or object was underneath the table)
6. Child threw or rolled the object to the parent and the parent simply received it <sup>1</sup>

## Results

### Action prediction

As a group, infants made 78 total anticipations out of 3640 gaze fixations. Per infant, they made an average of 2.44 anticipations throughout the interaction (range = 0-7,  $SD = 1.97$ ). Parents made 1176 total reaching events, an average of 36.75 reaches per parent (range = 17-67,  $SD = 12.14$ ). Of these, 563 represented valid opportunities to anticipate (average = 17.59 per parent, range = 7-33,  $SD = 5.08$ ). The 78 total infant anticipations corresponded to valid reaches; there were no anticipations that corresponded to an invalid reach. Therefore, the mean proportion of anticipated reaches out of all valid reaches across infants was .13 ( $SD = .11$ ).

These results indicate that infants do demonstrate action prediction at 9 months of age during free-flowing interaction, though infrequently (Figure 3). There was a substantial amount of variability among infants: while some infants never anticipated ( $n = 8$ , or 25% of the sample), others anticipated more than 40% of their parents' actions. After excluding infants who never anticipated, the mean proportion of anticipated reaches was .18 ( $SD = .09$ ), which is consistent with the findings reported above from all infants.

Out of 563 total reaching events across all parents, in 94 of these events the infant was already looking at the target object when the parent initiated their reach. In these cases, the parent was most likely responding to the child's visual attention by reaching for what the infant is looking at. When these reaching events are removed from the total count—they can also be considered invalid opportunities to anticipate, since the child cannot anticipate a target they are

already looking at—the average proportion of anticipated reaches increases to 0.16 ( $SD = 0.13$ ).

Given the low frequency of this behavior, we tested whether infants' anticipations could have been due to chance overlaps between infant gaze and parent reaching behavior. For each infant, we created 1000 randomized time-series by shuffling the sequence of gaze fixations while preserving their overall duration. We then aligned each randomized gaze sequence with the sequence of parents' reaching actions, calculated the number of anticipations that could occur by chance, and averaged over these 1000 values to yield a baseline anticipation rate for each infant. This resulted in a mean of 1.31 baseline anticipations across infants (range = 0.31-3.33,  $SD = 0.65$ ). A paired-samples  $t$ -test revealed that the average number of anticipations was significantly *higher* than baseline (mean difference = 1.13,  $t(31) = 3.84$ ,  $p = .001$ ). This result was the same when comparing the proportion of anticipated reaches with the chance proportion of .07, (mean difference = .09,  $t(31) = 4.34$ ,  $p < .001$ ). This finding reveals that infants' action prediction did not simply occur from chance overlaps between looking and parent reaching to the same object.

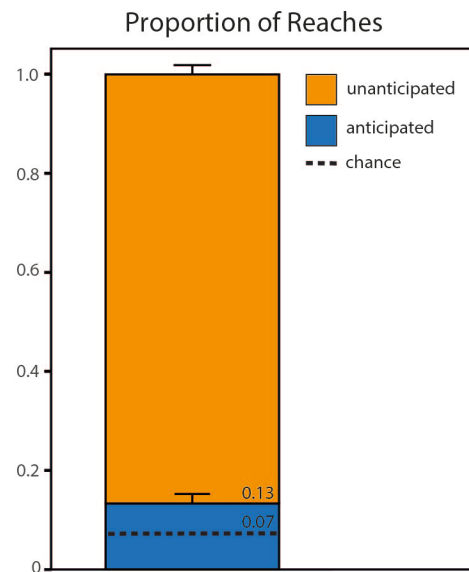


Figure 3: The proportions of reaching actions that were anticipated vs. unanticipated, with the dotted line representing chance. Error bars represent the s.e.m.

### Infant visual attention and manual activity during parent reaching

To explore the characteristics of the parent-child interaction during reaching events, we examined the infant and parent behaviors that were occurring during each reaching event. Figure 4 illustrates the proportion of all valid parent reaches in which the infant was attending to or manipulating a different object from the target of the reach. Parents were also holding another object in their other hand during 37.5% of their reaches. In fact, there was not *one single* reaching event across all dyads with no concurrent

<sup>1</sup>Here, the child may be anticipating the causal outcome of their own action or the movement trajectory of the ball rather than their parents' action goal.

visual and/or manual activity to a non-target object from either the child or the parent.

Given this finding, we conducted an additional analysis on the rate of anticipation with the aim of only considering opportunities that did not overlap with infants' own actions. We therefore excluded reaches during which the infant was holding or reaching for a non-target object. In these instances, infants were occupied with planning their own manual actions, which requires vision. After excluding infants' manual activity, this resulted in 251 valid reaches actions across parents ( $mean = 7.84$  per parent,  $SD = 3.62$ ). The mean proportion of anticipated actions was 0.36 ( $SD = .31$ ), which remained significantly higher than the chance level of 0.20 ( $SD = .12$ ),  $t(31) = 3.61$ ,  $p = .001$ . In other words, in a less demanding observation context (i.e., without competing goals from their own manual actions) infants will anticipate close to 40% of their parents' actions.

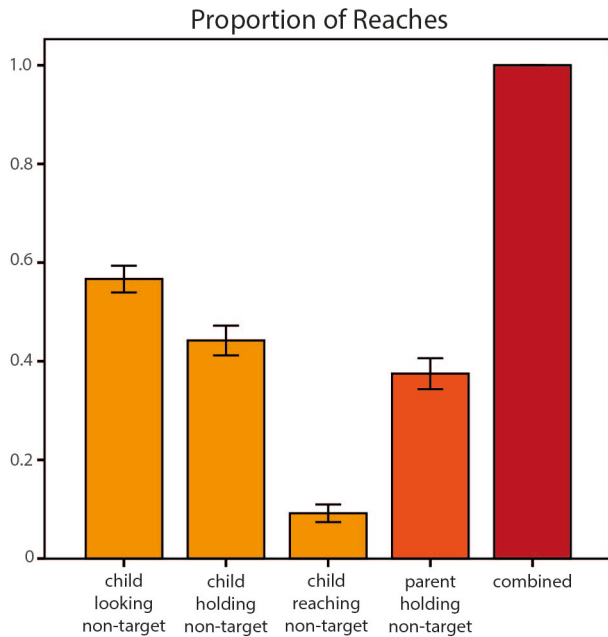


Figure 4: The proportion of reaching events in which infants or parents were concurrently attending to or manipulating a non-target object. From left to right, reaches during which 1) infants were looking at non-target object; 2) infants were holding a non-target object; 3) infants were reaching for a non-target object; 4), parents were holding a non-target object in their other hand, and 5) at least one of the above was occurring. (Note: 1-4 are not mutually exclusive and therefore add up to more than 1.)

### Anticipation latency

The mean duration of parents' anticipated reaches were 611.46ms ( $SD = 314.90$ ,  $range = 200-1770$ ). Figure 5 displays a bar chart of the time-course of infants' anticipations: the latencies between the onset of the reaching actions, infants' looks to the target object, and the contact with the goal (i.e., the end of the reach). The mean latency from the start of the reach to the moment the child looked to the target was 328.88ms ( $SD = 234.67$ ). The mean latency

from the gaze onset to the moment the hand reached the target was 282.58ms ( $SD = 282.31$ ). In other words, on average infants required just over 300ms to detect and process their parents' movements and then anticipate. This includes the time required for infants to program an eye movement in response to a visual stimulus (Gredebäck et al., 2010). This finding suggests that movement cues were a strong cue for anticipation.

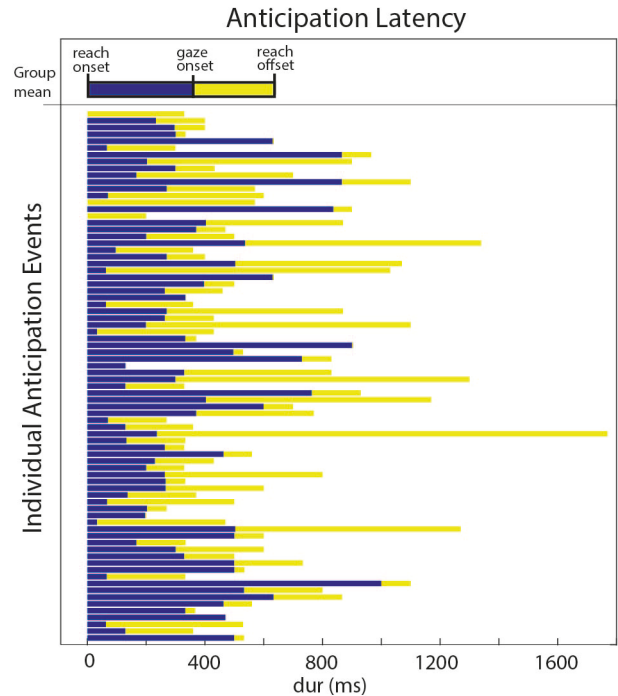


Figure 5: A histogram of the time course of infants' anticipations: blue bars indicate the latency between the start of the reach and the moment the child looked to the target; yellow bars indicate the latency between the child's look and the moment the parent's hand reached the target.

### Looks to Parent Faces

If an infant is looking at their parent's face, they could be more likely to perceive the onset of their parent's reach and its trajectory. To determine whether attending to their parents facilitated anticipations, we calculated the proportion of anticipations that were immediately preceded by a fixation to the parent's face. Out of the 78 anticipations performed across all subjects, 18 were immediately preceded by a face look. For the remaining 60 anticipations, 42 of them did not have a face look within a 3-second window before or after the anticipation. This finding suggests that looking to parents' faces was not a strong cue for anticipation.

### Discussion

The world of the developing infant is dynamic and constantly changing in both time and space. In the first year of life, infants become increasingly proficient actors and make rapid gains in their abilities to perceive and understand social events. Action prediction reflects an

important part of this social-cognitive process (Gredebäck & Daum, 2015; Hunnius & Bekkering, 2014) and has been widely studied using screen-based eye-tracking methods (Robson & Kuhlmeier, 2016). In the current study, we used dual head-mounted eye-tracking to investigate whether infants' predict their parents' object-directed reaching actions during free-flowing parent-child interaction.

Our primary finding is that, as a group, infants do anticipate their parents' actions at a rate that was significantly higher than what would be expected by chance. This finding demonstrates that infants' action prediction skills are not limited to the unambiguous, controlled action contexts that are typical in laboratory paradigms—they also demonstrate this ability during free-flowing parent-child play while they are also acting themselves.

On the other hand, the low rate of anticipation is also consistent with recent work investigating the real-time dynamics of parent-child interactions. New evidence from head-mounted eye-tracking studies have revealed, for instance, that infants actually rarely look to their parents' faces (Franchak, Kretch, Soska, & Adolph, 2010) and achieve joint attention through their own manual actions rather than through gaze following (Yu & Smith, 2016). The current study adds to this growing literature by revealing that infants also attend less to the goals of their parent's reaching actions. However, it is difficult to evaluate whether proportions of .13-.16 should actually be considered low, as there is no existing data to compare to these values. Future work could, for instance, quantify the rate of parents' anticipations of their infants' actions to provide a reference point.

A second finding was that infants anticipated their parents' actions on a rapid timescale—the mean latency after the reach onset was 328ms. In one recent study that also reported the “disengagement time”—i.e., the  $\text{reach}_{\text{onset}} - \text{gaze}_{\text{onset}}$  latency—on average infants required 344ms (SD = 209ms) to look to the target (Rosander & von Hofsten, 2011). These authors interpreted this finding as evidence that infants were able to use movement trajectories on a rapid timescale to accurately predict their parents' targets. In that study, infants were seated in a high chair and observed an experimenter move a small ball into a cylinder. Interestingly, infants demonstrated a similar timescale despite the increase in complexity of the toy play context.

A third finding that emerged is that there was not *one single* object-directed reaching action, in the entire sample, in which infants or parents were not simultaneously looking at or holding a different object than the target of the reach. Nevertheless, infants were still able to generate successful anticipations. One recent study may shed some insight into this finding. De Barbaro et al., (2016) investigated the qualitative shift from infant-guided object play to the triadic joint object play that emerges around 9-12 months of age. Their research highlights a “decoupling” between infant gaze and manual activity: in their studies, infants frequently directed their visual and sensorimotor attention to different objects (i.e., they do not look at what they are holding).

Likewise, they frequently shifted their gaze from the objects in their own hands and those in their parents' hands. These authors propose that this “sensorimotor decoupling” in hand-eye coordination contributes to the emergence of triadic interactions, by enabling infants to manipulate objects while still attending to the objects in their parents' hands.

This finding also highlights the dissociation between screen-based action contexts and real life: infants rarely experience isolated, unambiguous moments without competing cues or distractors, unlike discrete trials in experimental studies. Although examining what infants can or cannot do in a controlled laboratory setup is an effective approach to understanding infant cognition, it is also critical to examine how behaviors emerge from complex contexts with competing goals. For example, in the toy play context examined here, infants need to efficiently control their visual attention on a rapid timescale to serve multiple tasks—guiding their own manual actions, predicting their partner's actions, and sometimes using gaze to send social signals to their partner. In such real-life contexts, the key question is how the infant cognitive system operates with multiple ongoing tasks and how they distribute cognitive resources (e.g., attention and memory) to coordinate and manage these tasks.

What are the functional consequences of action prediction? Anticipating the actions of their social partners may help infants form associations between other peoples' actions and their goals or intentions. This pathway has been proposed to provide a potential explanation for how infants transition from forming associations between the behaviors they observe and more complex social understanding skills within the first years of life (Hunnius & Bekkering, 2014; Ruffman, Taumoepeau, & Perkins, 2012). In fact, a related study in our lab found correlations between the frequency of infants' action prediction and their vocabulary size, both at the same age and up to 6 months later (Monroy et al., under review). This finding provides preliminary evidence that action prediction may not only reflect infants' current social information-processing skills (Gredebäck & Daum, 2015) but also provide learning opportunities that support their developing social-cognitive system.

Our study represents a first attempt to investigate action prediction in naturalistic parent-child play. In our experimental set-up, parents and infants engaged in unstructured object play limited to three objects. However, this paradigm is also limited in the kinds of information available to infants. For instance, parents' actions did not lead to any meaningful action goal, as our everyday actions do (e.g., making a sandwich or building a Lego tower). In addition, there were no regularities in parents' reaching actions that they could use to anticipate their next target, which is one cue that infants use to generate predictions (Monroy et al., 2017). In future work, we plan to investigate action prediction in a broader range of action contexts—for instance, when infants and parents are engaged in joint activities that feature structure and shared goals.

## Acknowledgments

Thank you to Alexis Allard for data coding and to Steven Elmlinger, Melissa Hall, Charlene Tay, Charlotte Wozniak, Melissa Elston, and Seth Foster for data collection. We would also like to thank three anonymous reviewers for helpful comments on an earlier draft of this manuscript. Research reported in this publication was supported by the National Institute on Deafness and Other Communication Disorders under award number F32DC017076 to CM, and National Institutes of Health Grant R01HD074601 and R01HD093792 to CY.

## References

- de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother-infant-object interactions. *Child Development, 87*(2), 494–512. <https://doi.org/10.1111/cdev.12464>
- Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience, 9*(7), 878–879. <https://doi.org/10.1038/nn1729>
- Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature, 424*(6950), 769–771. <https://doi.org/10.1038/nature01861>
- Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2010). Head-mounted eye-tracking: A new method to describe infant looking. *Learning, 82*(6), 1–9. <https://doi.org/10.1111/j.1467-8624.2011.01670.x>. Head-mounted
- Gredebäck, G., & Daum, M. M. (2015). The Microstructure of Action Perception in Infancy: Decomposing the Temporal Structure of Social Information Processing. *Child Development Perspectives, 9*(2), 79–83. <https://doi.org/10.1111/cdep.12109>
- Gredebäck, G., Johnson, S., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology, 35*(1), 1–19. <https://doi.org/10.1080/87565640903325758>
- Hunnius, S., & Bekkering, H. (2014). What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1644), 20130490. <https://doi.org/10.1098/rstb.2013.0490>
- Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications, 2*, 341. <https://doi.org/10.1038/ncomms1342>
- Knoblich, G., & Sebanz, N. (2012). The Social Nature of Perception and Action. *Current Directions in Psychological Science, 15*(3), 99–104. Retrieved from <http://cdp.sagepub.com/content/15/3/99.short>
- Monroy, C., Gerson, S., & Hunnius, S. (2017). Infants' motor proficiency and statistical learning for actions. *Frontiers in Psychology, 8*(DEC). <https://doi.org/10.3389/fpsyg.2017.02174>
- Monroy, C., Gerson, S., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology, 157*, 14–28. <https://doi.org/10.1016/j.jecp.2016.12.004>
- Monroy, C., Chen, C., Houston, D., Yu, C. (under review). Action prediction during real-time parent-infant play predicts language development.
- Robson, S. J., & Kuhlmeier, V. A. (2016). Infants' Understanding of Object-Directed Action: An Interdisciplinary Synthesis. *Frontiers in Psychology, 7*(February), 111. <https://doi.org/10.3389/fpsyg.2016.00111>
- Rosander, K., & von Hofsten, C. (2011). Predictive gaze shifts elicited during observed and performed actions in 10-month-old infants and adults. *Neuropsychologia, 49*(10), 2911–2917. <https://doi.org/10.1016/j.neuropsychologia.2011.06.018>
- Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology, 30*(1), 87–104. <https://doi.org/10.1111/j.2044-835X.2011.02045.x>
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology, 18*, 668–671.
- Senna, I., Addabbo, M., Bolognini, N., Longhi, E., Macchi Cassia, V., & Turati, C. (2016). Infants' visual recognition of pincer grip emerges between 9 and 12 months of age. *Infancy, 22*(3), 389–402. <https://doi.org/10.1111/inf.12163>
- Stapel, J. C., Hunnius, S., & Bekkering, H. (2012). Online prediction of others' actions: The contribution of the target object, action context and movement kinematics. *Psychological Research, 76*(4), 434–445. <https://doi.org/10.1007/s00426-012-0423-2>
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1–34. [https://doi.org/10.1016/S0010-0277\(98\)00058-4](https://doi.org/10.1016/S0010-0277(98)00058-4)
- Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE, 8*(11), e79659. <https://doi.org/10.1371/journal.pone.0079659>
- Yu, C., & Smith, L. B. (2016). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science, 41*, 5–31. <https://doi.org/10.1111/cogs.12366>
- Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development, 88*(6), 2060–2078. <https://doi.org/10.1111/cdev.12730>
- Yu, C., Suanda, S. H., & Smith, L. B. (2018). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science, e12735*. <https://doi.org/10.1111/desc.12735>

Yuan, L., Xu, T. L., Yu, C., & Smith, L. B. (2019). Sustained visual attention is more than seeing. *Journal of Experimental Child Psychology*, 179, 324–336. <https://doi.org/10.1016/j.jecp.2018.11.020>