PAPER

Developmental Science WILEY

Action prediction during real-time parent-infant interactions

Claire Monroy[1] | Chi-Hsin Chen[1] | Derek Houston[1,2] | Chen Yu[3]

[1]Department of Otolaryngology, The Ohio State University Wexner Medical Center, Columbus, Ohio, USA

[2]Nationwide Children's Hospital, Columbus, Ohio, USA

[3]Department of Psychological and Brain Sciences, University of Texas at Austin, Austin, Texas, USA

Correspondence
Claire Monroy, Department of Otolaryngology, The Ohio State University Wexner Medical Center, Columbus, OH 43210, USA.
Email: claire.monroy@osumc.edu

Abstract

Social interactions provide a crucial context for early learning and cognitive development during infancy. *Action prediction*—the ability to anticipate an observed action—facilitates successful, coordinated interaction and is an important social-cognitive skill in early development. However, current knowledge about infant action prediction comes largely from screen-based laboratory tasks. We know little about what infants' action prediction skills look like during real-time, free-flowing interactions with a social partner. In the current study, we used head-mounted eyetracking to quantify 9-month-old infants' visual anticipations of their parents' actions during free-flowing parent–child play. Our findings reveal that infants do anticipate their parents' actions during dynamic interactions at rates significantly higher than would be expected by chance. In addition, the frequency with which they do so is associated with child-led joint attention and hand-eye coordination. These findings are the first to reveal infants' action prediction behaviors in a more naturalistic context than prior screen-based studies, and they support the idea that action prediction is inherently linked to motor development and plays an important role in infants' social-cognitive development. A video abstract of this article can be viewed at https://www.youtube.com/watch?v=9HrmcicfiqE

KEYWORDS
action prediction, head-mounted eyetracking, parent–child interaction, sensorimotor coordination, social-cognitive development

1 | INTRODUCTION

From playing with blocks to mealtime exchanges, infants' social experiences largely consist of interactions with their parents. These experiences provide critical opportunities for the early communicative exchanges that give rise to the rapid social-cognitive and language development that happens in infancy. Typical joint activities—for instance, passing a ball back and forth on the playground or handing over a spoon at the dining table—require action coordination between two social partners. One process that is especially important for fostering coordinated behaviors between two social partners is action prediction: anticipating a social partner's action target or goal is critical for planning an appropriate behavioral response (Sebanz & Knoblich, 2009).

Infants demonstrate action prediction starting from early in life. This ability is commonly measured in infants by analyzing anticipatory gaze during passive observation of discretized action events shown on a computer screen (Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Hunnius & Bekkering, 2010; Monroy et al., 2017a). These studies have yielded important insights into infants' early action understanding and their object knowledge. For instance, Hunnius and Bekkering (2010) showed 6-month-old infants videos of an actor performing familiar (e.g., bringing a phone to the ear or a cup to the mouth) or unfamiliar actions (e.g., bringing a phone to the mouth or a cup to the ear). Using screen-based eyetracking, they were able to show that infants at 6 months of age already expect that a person will bring a cup to their mouth and a phone to their ear, and that they reveal these expectations by looking to where the ongoing action will unfold. Infants can also learn the statistical regularities within action sequences and accurately predict a future action before it begins (Monroy et al., 2017a). From as early as 2 months of

age, infants will make postural adjustments prior to being picked up, demonstrating that they anticipate actions that involve them in addition to those they simply observe (Reddy et al., 2013). These studies, among many others, highlight the many ways in which infants reveal their expectations about action events and their growing knowledge about the behavior and goals of other people (Hunnius & Bekkering, 2014).

1.1 | Action prediction 'in the wild'

Prior research on infant action prediction primarily consists of findings from controlled laboratory paradigms. These paradigms have the advantage of allowing the experimenter to highlight or manipulate specific action cues. For instance, a common technique is to reveal only the actor's hand against a plain background to eliminate distraction from the actor's face and eyes, drawing infants' attention to the action itself (e.g., Falck-Ytter et al., 2006). Some paradigms introduce 'occluders' into the display to elicit anticipations from infants and discourage them from simply tracking the moving agent (e.g., Paulus et al., 2017). These experimental designs serve to maximize the opportunities for infants to anticipate. At the same time, these paradigms display action contexts that considerably simplify the everyday action contexts that infants observe in daily life, such as their parent preparing a meal in a cluttered kitchen. Furthermore, in these 'clean' paradigms, the infant is typically a passive observer of action stimuli that are pre-segmented for them into discrete trials. However, in real-life interactions, infants are also actors themselves. They need to plan and execute their own goal-directed actions in real time, while also observing their social partner's movements and responding appropriately. To achieve this, infants must dynamically distribute their visual attention to serve two concurrent tasks: guiding their own actions and attending to their social partner's actions. Little is known about whether infants anticipate others' actions 'in the wild', while interacting naturally with their parent in a social context. In the current study, we apply head-mounted eyetracking to the study of infant action prediction to determine whether and when infants predict their parents' reaching actions during parent–child play.

Head-mounted eyetracking provides a view of the world from the child's perspective (Slone et al., 2018). Recent studies using head-mounted eye-trackers to investigate infant visual attention during social interactions have shed new insights into their early social and language environments. For instance, recent studies on joint attention revealed that infants rarely look at their parents' faces during free-flowing object play (Deák et al., 2018; Franchak et al., 2010; Yu & Smith, 2013, 2017). These findings challenge the established theory that joint attention is achieved primarily through gaze following (Brooks & Meltzoff, 2005; Carpenter et al., 1998; Meltzoff & Brooks, 2007). This work also illustrates that observations of infant behavior which are reliably demonstrated in the lab may not always generalize to the more complex contexts of daily life. In the current study, we used head-mounted eyetracking to record the

Research Highlights

- We used head-mounted eye-tracking to measure action prediction during parent–child play.
- Nine-month-old infants predict their parents' actions during free-flowing social interactions.
- Action prediction occurred at rates higher than what would be expected by chance.
- Action prediction was correlated with infants' hand-eye coordination and joint attention.

visual and motor behaviors of both infants and their parents, making it possible to measure sensorimotor coordination across modalities and across social partners.

1.2 | Linking action prediction 'in the wild' to other developmental processes

Given that action prediction 'in the wild' happens while infants are themselves acting upon the world, a complementary question to *if* action prediction happens during more naturalistic contexts is *how* action prediction in those contexts relates to other developmental processes. A substantial body of research reveals that infant action prediction skills are coupled with their own motor capabilities. Infants become more precise at anticipating an observed action once they have acquired the requisite motor skill (Monroy et al., 2017b; Senna et al., 2016; Stapel et al., 2016), a finding that is not explained by general development. For instance, Monroy et al. (2017b) took advantage of the natural variation in the emerging motor skills of young infants to show that those infants who had acquired a specific motor skill (e.g., a pincer grasp) were more precise at anticipating that action than infants of the same age who had not yet mastered that motor skill. Infants also demonstrate activation over their motor cortex when anticipating an action outcome (Monroy et al., 2019; Southgate et al., 2010). Together, these findings provide strong evidence that infants' motor capabilities facilitate their ability to process and predict the actions they observe. Here, we aimed to extend these findings by examining whether action prediction during parent–child play is correlated with hand-eye coordination (looking at and holding the same object). This hypothesis is supported by prior research demonstrating robust coupling between gaze and hand movements for actions that we execute as well as actions that we observe (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003).

A second developmental process of relevance to the current study is joint attention.[1] Social interaction requires joint attention—attending to objects and events that one's partner is attending to. Joint attention is generally agreed to be a critical social-cognitive milestone (Tomasello, 1995). Sebanz et al. (2006) provides a theoretical explanation for potential links between action prediction and joint attention, based on empirical findings from adult research. First,
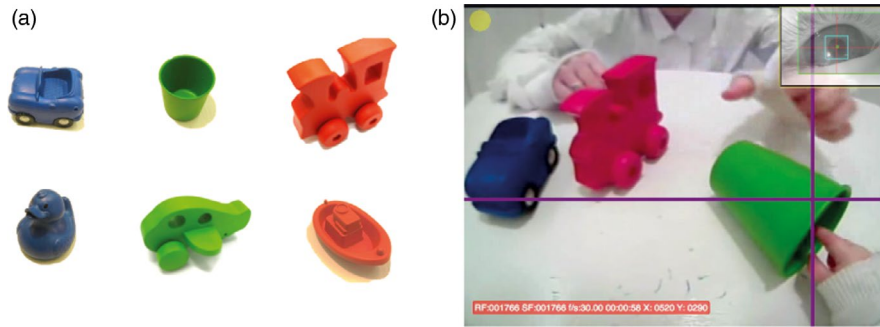
FIGURE 1 Stimuli and experimental set-up. (a) Familiar object sets. (b) Example frame from an infant's head camera while the parent is reaching for an object, with the crosshair indicating estimated direction of gaze

knowledge about what the social partner is attending to provides important cues about their action targets and can facilitate predictive eye movements (Sebanz et al., 2006). Second, in both adults and infants, eye movements are similar both when we execute actions and when we observe others' actions (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003), in that our gaze shifts to the target of a reach just before the actor's hand makes contact. This similarity increases the likelihood that we would align our visual attention with a social partner. Currently, there is some evidence for links between action prediction and social-cognitive development in toddlers. For example, Krogh-Jespersen et al. (2015) showed that faster action prediction (i.e., shorter gaze latencies to the goal of an actor's reach) is associated with better social competence (i.e., perspective-taking) in 20-month-old toddlers, independ of toddlers' general social engagement. Another study by Meyer et al. (2015) showed that action prediction was correlated with more successful turn-taking skills in 2.5-year-olds. These studies suggest that action prediction is associated with other measures of social cognition. However, to our knowledge, no empirical study has examined links between action prediction and joint attention.

1.3 | The current study

We analyzed data from parent–child play sessions when infants were 9 months of age. We focused on this young age because screen-based eyetracking studies have shown that infants can anticipate reaching actions from kinematic cues by 9 months (Ambrosini et al., 2013). Parent–infant dyads played freely with three familiar objects while their eye movements and actions were recorded using head-mounted eye-trackers. Our primary aim was to identify whether infants demonstrate anticipatory looking to parents' actions during free-flowing play, and, if so, to examine the contexts in which they do so. Our secondary aim was to examine whether action prediction correlates with infants' hand-eye coordination and joint attention at 9 months and at 15 months of age. While we expect that action prediction in general may be related to joint attention and hand-eye coordination, in the current study we focus exclusively on reaching actions because they provide a clean opportunity to measure anticipatory looking in a free-flowing parent–infant play context.

2 | METHOD

2.1 | Participants

Data represent a corpus from a longitudinal dataset, collected between 2014 and 2016. Thirty-two parent–child dyads contributed eyetracking data when infants were 9 and 15 months old (mean infant age = 9.3 months at first visit; range = 9–9.7; 18 females). The sample size of 32 is consistent with prior studies using similar high-density eyetracking measures (Franchak et al., 2010; Kretch & Adolph, 2017; Ossmy et al., 2020; Yu et al., 2019). All infants were healthy and born full-term.

2.2 | Stimuli

Parents and infants were presented with six familiar, engaging toys—a car, a 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 object (Figure 1).

2.3 | Experimental set-up and procedure

Parents and infants were seated opposite one another at a small table (61 cm × 91 cm × 64 cm). Parents sat on the floor approximately eye-level to their child. Both dyad members were fitted with head-mounted eye-trackers (Positive Science, Inc). Each eye-tracker featured an infrared camera directed towards the right eye and a head camera that recorded 90° of the visual field. Two additional cameras recorded third-person views of each dyad member. All six cameras recorded at 30 Hz and were synchronized offline using ffmpeg (https://ffmpeg.org).

To calibrate the eye-trackers, an experimenter drew infants' attention by placing an engaging toy in 15 unique locations on the tabletop. Parents were asked to attend to the toy as well. These moments were used to calibrate eye gaze relative to the head camera recording offline using Yarbus software (Positive Science, Inc). Yarbus uses a specialized algorithm to map each position of the pupil and corneal reflection from the eye-tracker recording to corresponding

locations in the head camera recording. This yields a calibrated video with the estimated direction of gaze indicated by a crosshair and superimposed on the head camera recording. This experimental procedure of head-mounted eye tracking has been successfully used in many studies on adults (Hayhoe et al., 2003; Jovancevic-Misic & Hayhoe, 2009; Land & Hayhoe, 2001) and infants (Bambach et al., 2013; Chen et al., 2020; Franchak et al., 2010; Pérez-Edgar et al., 2020; Suarez-Rivera et al., 2019). Additional details of the calibration procedure and useful practices to improve eye-tracking quality can be found in Yu and Smith (2017) Supplemental Information.

Following calibration, we presented participants with each set of familiar toys in alternating order. Parents were instructed to play with their infant "as they normally would at home". Dyads played with each toy set twice for 90 s, resulting in four 'trials' and six minutes of total interaction. The order of toy sets was counterbalanced across dyads.

2.4 | Data processing

After offline calibration, a crosshair indicating the estimated focus of infant gaze was superimposed onto the head camera recording, creating an additional recording of the calibrated gaze. All seven recordings (the six camera recordings plus the calibrated gaze recording) were then exported into a series of single frames. Each camera contributed a maximum of 10,800 frames per dyad (6 min of recording at 30 frames per second). Subsequent data processing and statistical analyses were performed using custom-written code in Matlab (see https://github.com/lingerxu/timevp).

2.4.1 | Manual activity: reaching and holding

Parent reaching was defined as any hand movement towards an object that ended with an object contact. A trained coder used frames from the head camera images and the two third-person view cameras to determine, on every frame, whether a hand was reaching towards a toy and, if so, which one. Right and left hands were coded separately and then merged to yield one data stream.

Two additional trained coders annotated infant and parent holding behaviors with the toys, defined as any contact between the hand and an object (Figure 2). As before, right and left hands were coded separately and then merged to yield one data stream for parent actions and one for infant actions. The second coder also annotated a random 10% of the frames, with reliability ranging from 91% to 100% (Cohen's kappa = 0.94, indicating almost perfect agreement).

2.4.2 | Opportunities to anticipate

Not all parent reaches provided fair opportunities for the infant to anticipate. To estimate rates of anticipation out of all actual
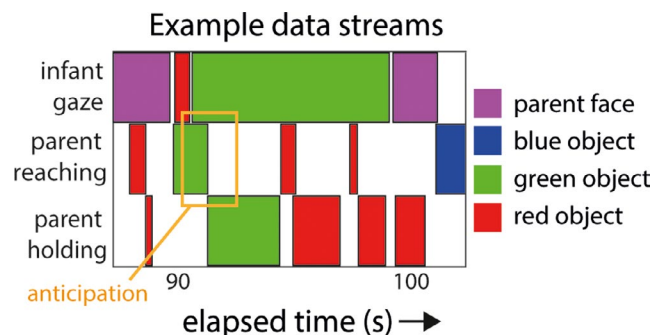

FIGURE 2 A sample of the aligned gaze and reaching data streams from a representative dyad. The yellow box highlights an example of an anticipation: the infant looks to the green object after the reach onset and prior to the end of the reach. Parent holding is included here for visualization purposes

opportunities, rather than simply the total number of coded reaches, we determined whether reaching events provided a valid or an invalid opportunity to anticipate. This determination was performed in two steps. First, we automatically rejected reaching events based on temporal properties (see Table 1 for detailed explanations of each criterion). These included reaches during which infant was already holding/touching the object, reaches that lasted less than 200 ms (the time needed to program an eye movement), or reaches that were actually subsequent contacts in cases of multiple object contacts (e.g., tapping the object). Second, we manually rejected reaching events that could not be identified automatically. These included reaches with experimenter interference (i.e., a toy fell onto the floor and the experimenter replaced it), reaches where the infant was reaching for the object simultaneously, reaches where the parent and object were entirely of the infant view, or reaches where the infant threw/rolled the object to the parent. Of 1164 reaching events across all parents, 509 were categorized as valid reaching events. The remaining 655 were considered invalid and excluded from further analyses. A second experimenter coded reaching events for validity for 20% of all participants, with an interrater reliability of 90% (Cohen's kappa = 0.71, indicating substantial agreement).

2.4.3 | Infant gaze

Gaze was coded frame-by-frame by two trained, independent coders. Four regions-of-interest (ROIs) were defined from the calibrated head-camera videos: the three objects and the parent's face.[2] To determine whether gaze fell within these ROIs, coders watched the calibrated video with a cross-hair indicating gaze direction and annotated for every frame whether the cross-hair fell on a pixel identified as any part of the four ROIs. 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. A second coder annotated a random 10% of the frames. Reliability ranged from 82% to 95% (Cohen's kappa = 0.81).

TABLE 1 Criteria for categorizing a parent reach as an *invalid* opportunity to anticipate

	Criterion	Method	N
1.	The infant was already looking at the target object (in this case, the infant cannot anticipate something they are already looking at). These events were identified based on whether the infant's gaze fixation began before the reach onset and ended after reach onset	A	60
2.	The reach lasted less than 200 ms (the time needed to program an eye movement)	A	97
3.	Subsequent contacts in cases of multiple object contacts (e.g., tapping the object or moving the object from one hand to the other), based on if the reach onset was less than 3 s after the offset of the previous holding event	A	123
4.	The infant was already holding the object that the parent was reaching for. These events were identified based on temporal overlap between parent reaching events and infant holding events for the same objects	A	215
5.	The infant was reaching for the object at the same time (in this case, it impossible to determine whether the infant is anticipating their own action or their parent's)	M	20
6.	Experimenter interference	M	20
7.	Both parent and object were entirely out of the infant's view for the entire duration of the reach (e.g., parent was retrieving the object from underneath the table)	M	19
8.	The infant threw or rolled the object to the parent and the parent received it (in this case, it is impossible to determine whether the infant is tracking the ball's trajectory rather than anticipating the reaching event)	M	87

An "A" in the column *Method* indicates that these criteria were implemented automatically using Matlab functions. An "M" indicates that these criteria were manually coded by two independent coders.

Additional details regarding the coding procedure are reported in Yu and Smith (2017) Supplemental Information.

2.4.4 | Predictive gaze

To identify infants' anticipatory looks to their parents' actions, the two data streams from infant gaze and parent reaching were aligned (Figure 2). We operationally defined action prediction as a gaze shift to an object that occurred after the onset of a parent reach to that same object, but before the reach was completed. This represents the time window in which infants potentially have enough information to predict the goal of their parents' reach and make an anticipatory eye movement to the target object. The total number of gaze anticipations were summed and divided by the total number of valid parent actions to yield the proportion of predicted actions. Proportions of gaze anticipations were entered as the dependent variable in subsequent analyses (Monroy et al., 2017a; Stapel et al., 2015). The numbers of anticipations were also used to perform correlation analyses with the numbers of joint attention and hand-eye coordination bouts (see next section).

2.4.5 | Joint attention and hand-eye coordination

Joint attention and hand-eye coordination were derived from the aligned data streams of gaze and manual activity. Overall joint attention was defined as any time period in which the parent and child were looking at the same object. Joint attention events were further divided into whether they were child-led or parent-led, based on which dyad member first looked to the jointly attended object (Chen et al., 2020; Yu & Smith, 2016). The total number of each type

of event (child-led or parent-led) per infant served as our measure of joint attention. Hand-eye coordination was defined as any period of time in which the infant was looking at and holding the same object (Abney et al., 2018; Yu & Smith, 2017). As before, the total number of hand-eye coordination events per infant served as our measure of hand-eye coordination. We conducted correlation analyses to test for associations between the number of anticipatory looks, the number of joint attention bouts, and the number of hand-eye coordination events. To control for differences across infants in the duration of hand-eye coordination events, mean duration was included as a covariate.

3 | RESULTS

3.1 | Action prediction

The primary aim of this study was to determine whether infants predict their parents' actions during parent–child play. To do so, we quantified the frequency of infants' gaze anticipations to their parents' actions. The total number of gaze anticipations to parent actions was 78 across all infants, and the total number of valid parent reaches was 509. Per dyad, the mean number of anticipations was 2.44 (SD = 1.97) and the mean number of parent reaches was 15.91 (SD = 4.60). The mean proportion of anticipatory gaze, of total reaches, was 0.153 (range = 0–0.38; SD = 0.11; Figure 3). On average, infants therefore experienced 3.95 opportunities to anticipate per minute, and performed 0.58 anticipations per minute.

These results suggest that infants do anticipate their parents' actions at 9 months of age during free-flowing parent–child play. However, given the low frequency of this behavior, one possibility is that infants' object looks coincided with parent reaches to those
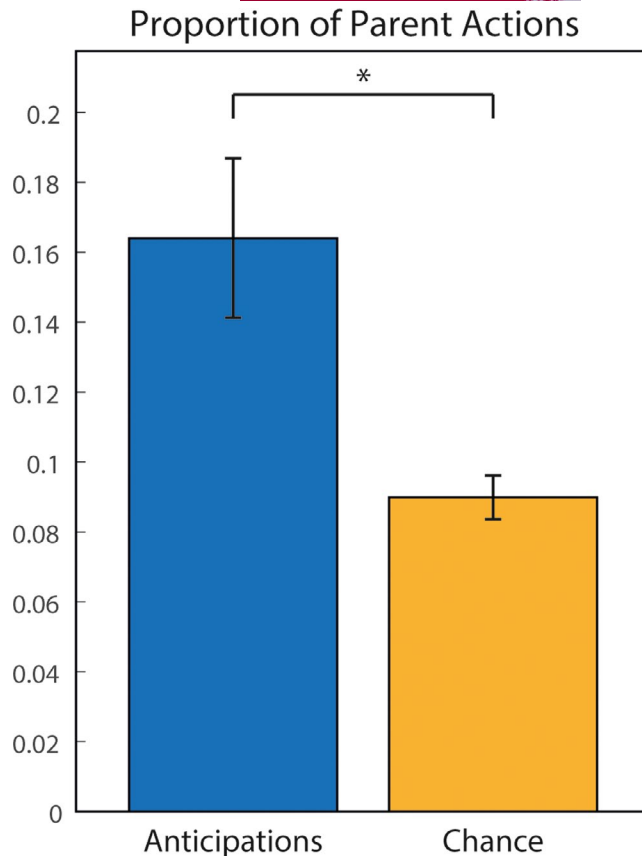
Proportion of Parent Actions

FIGURE 3 The proportion of reaches that were anticipated compared with the chance proportion. Error bars represent the standard errors of the means

objects simply by random chance. To evaluate this possibility, we calculated individual 'chance' levels of action prediction. For each infant, we created 1,000 randomized time-series by shuffling the sequence of infants' gaze fixations while preserving their overall duration and ROI category (Dale et al., 2011). Each randomized gaze sequence was aligned with the sequence of the parent's reaching actions. We then calculated the number of random overlaps between gaze and reaches that occurred in every shuffled data stream, and averaged across these 1,000 values to yield a chance anticipation rate for each infant. In other words, we performed 1,000 simulations to calculate how often randomly shuffled gaze fixations would align with parent reaches and averaged over these 1,000 simulations. This resulted in a mean of 1.31 baseline anticipations across infants (range = 0.32–3.33, SD = 0.65) and a mean chance anticipation proportion of 0.082 (SD = 0.03). A paired-samples t test revealed that the average number of observed anticipations (2.44) was significantly higher than chance number of 1.31 (mean difference = 1.13, $t(31) = 3.84$, $p = .001$; Figure 3). The same was true when comparing the chance proportion of 0.082 with the observed proportion of anticipated reaches (mean difference = 0.07, $t(31) = 3.48$, $p = .002$). These results indicate that action prediction was not simply due to random temporal overlap between infant looking and parent reaching to the same object.

To investigate the possibility that the duration of parent reaches exerted significant influence over infants' abilities to anticipate, we conducted an ANOVA on the duration of parent reaches that were anticipated (M = .61 s, SD = 0.31 s), unanticipated (M = 0.57 s, SD = 0.28 s), or reacted to (i.e., infant gaze arrived after the reach ended; M = 0.57 s, SD = 0.27 s). This yielded no significant main effect of reach type, indicating that the duration of reaches did not differ based on whether infants anticipated them, $F(2, 717) = 0.76$, $p = .47$. There was also no significant correlation between the number of infant anticipations and the mean duration of anticipated reaches, $r(24) = .23$, $p = .29$, revealing that infants who made more anticipations did not simply have parents who made slower reaches.

Finally, we examined whether the target of each anticipated reach was ambiguous at reach onset, to determine whether infants had clear information about their parent's goal from early on. We operationally defined ambiguity as whether the target object (i.e., the to-be-grasped object) was touching another object (ambiguous) or was not touching another object (unambiguous; see Figure S1). We used the frame from the child head camera recording that corresponded to reach onset to make this determination. This analysis revealed that 18 of the 78 anticipated reached events (23%) were ambiguous, in that the target object was touching another object at the onset of the reach. The remaining 60 reaches (77%) were unambiguous, revealing that most of the time infants had clear information about the target of their parents reach from early on.

3.2 | Infant visual attention and manual activity during parent reaching

To better understand what may have facilitated or prevented action prediction events, we examined the behaviors surrounding reaching events (i.e., the windows of opportunity to anticipate). For instance, if infants anticipated just 15.3% of their parents' reaching actions, what were infants looking at during the *unanticipated* reaches? Were infants more generally reactive, or were they simply inattentive to their parents' actions altogether?

First, infants could detect the goal of their parents' reaches through gaze following, if parents look to the target of their own action prior to initiating their reach. To test this, we repeated our primary analysis to determine whether parent gaze shifts (instead of reaches) to target objects aligned with infant gaze shifts at levels above chance. We first extracted parent gaze fixations to target objects that coincided with (valid) reaches to those objects, and then calculated the proportion of parent gaze shifts that preceded an infant gaze shift to that object. Next, we calculated a chance proportion (as described above) of how frequently parent gaze shifts would randomly align with infant gaze. This analysis revealed that the proportion of parent gaze shifts followed by an infant gaze shift was 0.076 (SD = 0.11) which was not significantly different from the chance proportion of 0.070 (SD = 0.052). These findings suggest that gaze following was not a primary cue for action anticipation, although it did occur in some reaching events.

Instead, during *anticipated* reaching events, infants were most likely to be looking at a non-target object (i.e., one of the two remaining objects the parent was not reaching for; 85.9% of reaching events) prior to shifting their gaze to the target object. Infants were manually actively engaged throughout the interactions: during 42 of the 78 anticipated reaches (53.85%), infants were manipulating a different object in their own hands. There was no reaching event during which the parent or the child was not looking at or holding a different object from the parent's target.

During *unanticipated* reaching events, infants were most likely to be looking at a different object from their parent's target (59.16%), and then at their parents' face (22.74%). Infants were not attending to any target on 18.1% of unanticipated reaching events (e.g., looking elsewhere in the room). These findings suggest that it was not always simply lack of attention altogether that caused failure to anticipate. Infants *reacted* to their parents' reaches on 41.06% of all events—that is, they looked to the target object within two seconds after their parent touched it (Figure 4 right). About as frequently (43.81%), infants never looked to the target object during reaches nor within this two-second window. During 29.23% (126 of 431) of unanticipated reaches, infants were manipulating a different object in their own hands.

To further clarify whether rates of anticipation differed depending on infants' own manual activity, we separated reaching events depending on whether infants were holding another (i.e., nontarget) object. When infants were holding a different object, there were 234 opportunities to anticipate and infants generated 29 anticipatory looks (14.15%). When infants were not holding a different object, there were 275 opportunities and they generated 49 anticipatory

looks (17.82%). To determine whether infants anticipated more when they were not concurrently holding another object, we conducted a binary logistic model-based generalized estimating equation (GEE) with an unstructured working correlation matrix (Zeger et al., 1988). Our dependent variable was anticipation, with each reaching event assigned a 1 if it was anticipated and a 0 if it was not. "Holding" was entered as a predictor variable, with each reaching event assigned a 1 if the child was concurrently holding another object and a 0 if the child was not. This analysis revealed a significant main effect of holding, $\chi^2(1) = 25.41$, $p < .001$, indicating that infants were more likely to make an anticipation when they were not manually engaged with another object at the same time.

3.3 | Correlations with hand-eye coordination and joint attention

Figure 5 illustrates the correlations between measures of action prediction, joint attention, and hand-eye coordination. Action prediction was significantly, positively correlated with infants' hand-eye coordination for their own actions at 9 months ($r = .408$, $p = .023$), after controlling for the duration of hand-eye coordination events. Infants who demonstrated more frequent moments of coordinated gaze and manual activity—i.e., they were looking at and touching the same object—also predicted their parents' actions more frequently. The proportion of anticipated actions was not, however, correlated with the proportion of overall time spent in hand-eye coordination. Finally, there was also a trend toward significance for the correlation between action prediction at 9 months and hand-eye coordination
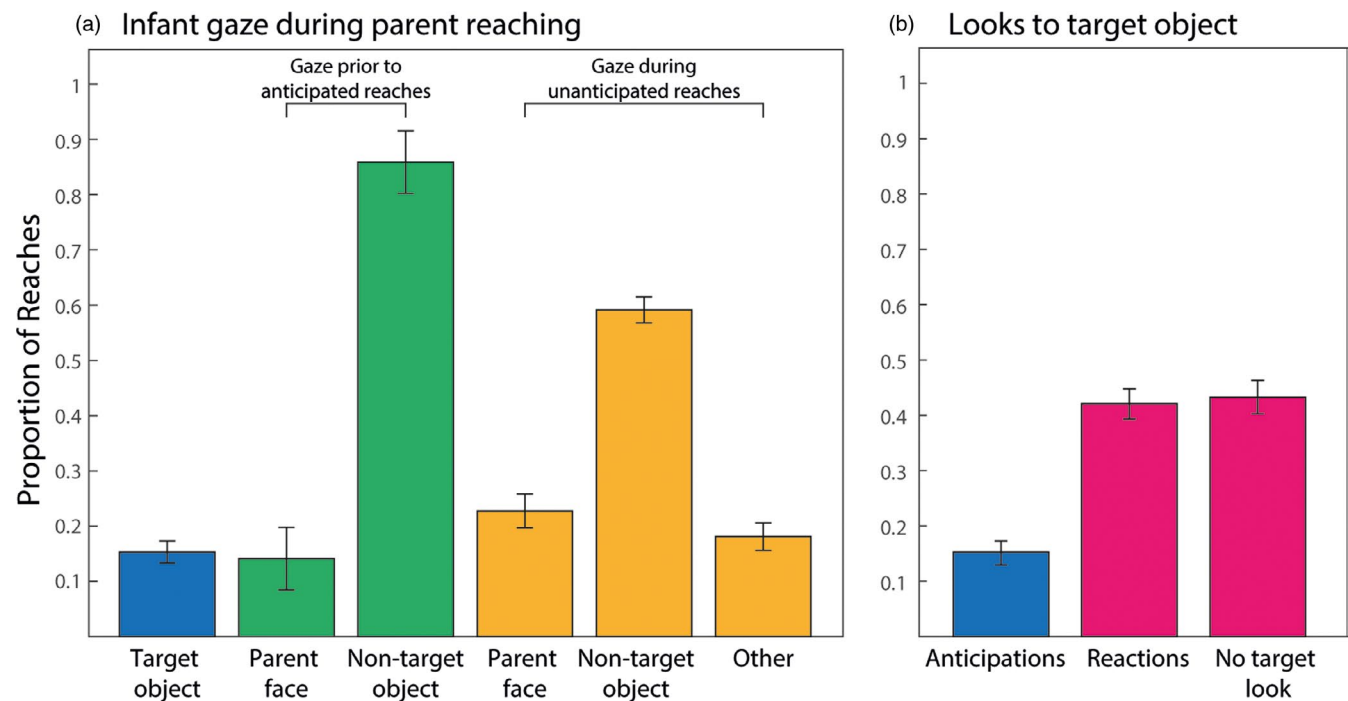

FIGURE 4 (a) Infant gaze behaviors prior to or during anticipated and unanticipated reaches. (b) Proportion of reaches that infants anticipated, reacted to, or never attended to the target of the parents' reach. Error bars represent the standard errors of the mean
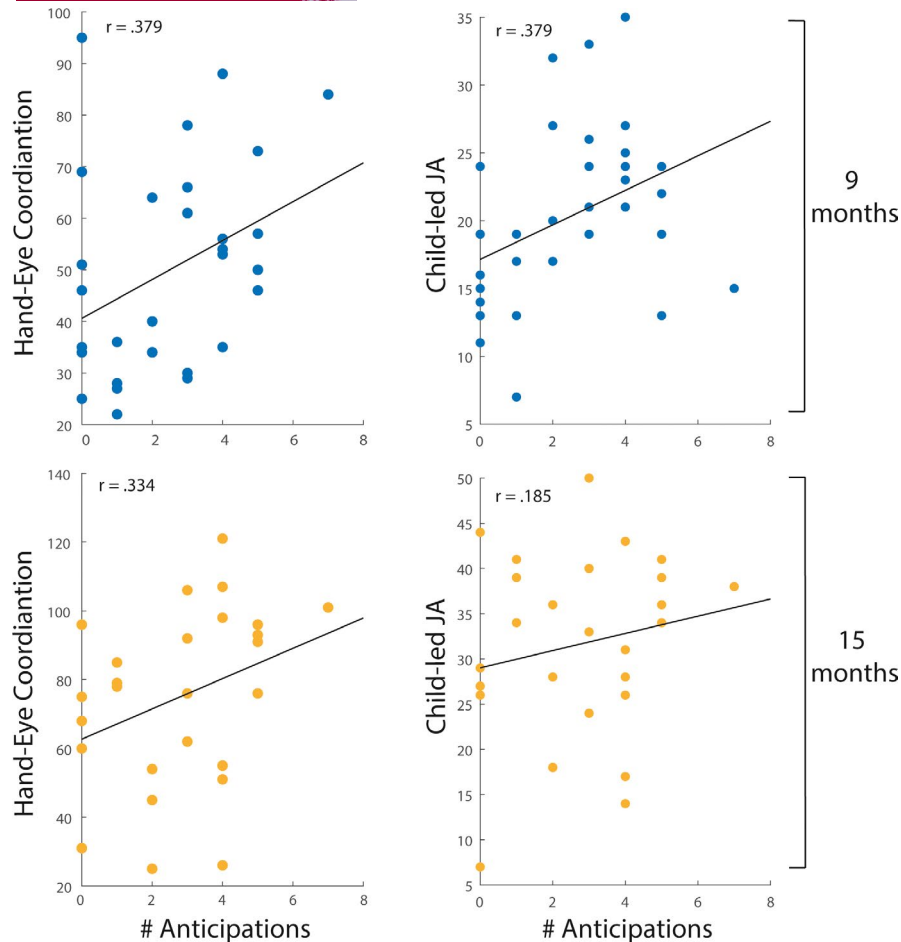
FIGURE 5 Scatterplots (with least-squares lines) depict correlations between action prediction at 9 months of age, and joint attention and hand-eye coordination at 9 months (upper) and at 15 months (lower)

at 15 months of age ($r = .372$, $p = .062$, $n = 32$). This finding suggests that action prediction is associated with the general coordination of visual attention with manual actions.

Action prediction was also significantly, positively correlated with the number of bouts of child-led joint attention (bouts of joint attention that were initiated by the child and 'joined into' by the parent), $r = .379$, $p = .032$, $n = 32$. Action prediction was not correlated with parent-led or overall joint attention ($ps > .17$). Infants who experienced more bouts of child-led joint attention also predicted their parent's actions more frequently at 9 months of age. There were no differences in the number of child-led versus parent-led joint attention bouts: infants experienced an average of 20.25 ($SD = 6.59$) child-led joint attention bouts and 18.50 ($SD = 6.29$) parent-led joint attention bouts ($t(31) = 1.22$, $p = .233$. Action prediction was not correlated with joint attention at 15 months of age ($p = .37$).

Finally, we examined the relationship between hand-eye coordination and joint attention (Figure 6). Hand-eye coordination was significantly, positively correlated with child-led joint attention at 9 months of age ($r = .380$, $p = .032$, $n = 32$) and was even more strongly correlated at 15 months ($r = .630$, $p < .001$, $n = 32$). Children with more coordinated visual and manual activity also experienced
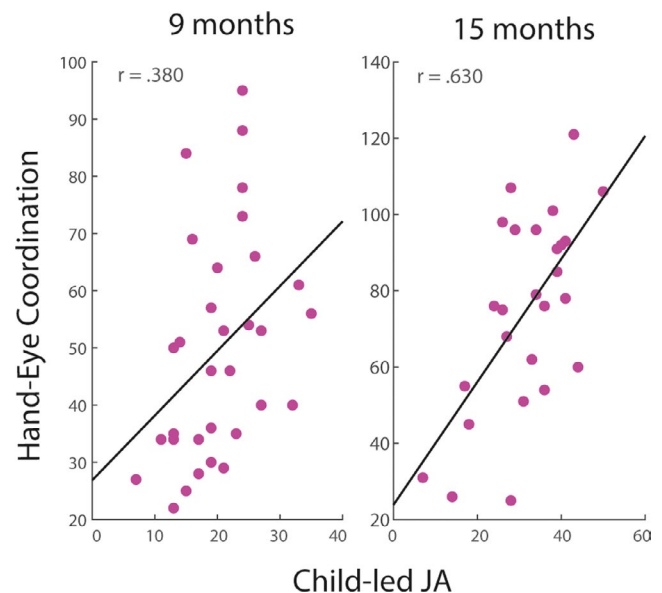

FIGURE 6 Scatterplots (with least-squares lines) depict correlations between hand-eye coordination and child-led joint attention at 9 and 15 months

higher numbers of child-led joint attention bouts, and these developmental processes remain coupled over time.

4 | DISCUSSION

This study is the first to investigate infant anticipatory looking during free-flowing parent–child play. Using head-mounted eyetracking, we characterized the rate of action prediction in 9-month-old infants. As predicted, at a group level 9-month-old infants do anticipate their parents' actions. The rate of action prediction was significantly above a chance level, which indicates that these findings reflect infants' genuine anticipation of their parents' goal-directed movements during real-time social interactions.

Nevertheless, anticipations were infrequent—less than one anticipation per minute of interaction on average. This low rate is consistent with some prior studies on infant action prediction: for instance, one screen-based experiment using action stimuli found anticipation rates of 20%–30% in toddlers (out of all gaze fixations), while a similar study using more complex visual stimuli with infants found anticipation rates of only 5% (Monroy et al., 2017b). In our study, infants were engaged in their own manual actions while monitoring those of their parents, and they were significantly less likely to anticipate when holding objects of their own. Given this added complexity, it is not surprising that infants anticipate relatively infrequently during unstructured play. We expect that action prediction would be higher for structured tasks (e.g., building a tower) or familiar daily activities (e.g., making peanut butter and jelly sandwiches), a question we are pursuing in ongoing work.

Our findings also highlight the dynamic nature of the world of the infant. As others before us have shown, we found that infants experience complex, multimodal inputs from their parent's visual and manual activity and their own sensorimotor behaviors (Chang et al., 2016; Franchak et al., 2010). Our findings reveal that infants never experience discrete, unambiguous action events like those typically presented in controlled laboratory paradigms. Instead, the events that infants observe in natural interactions are always coupled with overlapping activity: during every reaching event in our study, the parent or infant was looking at or holding a different object than the parent's target. It is therefore unsurprising that infants anticipate less frequently than in traditional laboratory experiments, in which action events are presented in clean, unambiguous contexts.

For infants who never anticipated, they may have been absorbed in their own object exploration rather than attending to their parent's actions. This is consistent with recent work investigating real-time dynamics of parent–child interaction: in the first year of life, infants demonstrate less face-looking and mutual gaze than previously thought from studies using traditional paradigms (Franchak et al., 2010; Yu & Smith, 2013, 2017). One contributing factor is that the visual fields of young infants are dominated by objects directly in front of them because of their physical characteristics (Yoshida & Smith, 2008). A target for future research is to examine the developmental trajectory of action prediction. Research has shown that the dynamic properties of parent–child interactions become more tightly coupled as infants grow older (Xu et al., 2018)—for instance, gaze, verbal cues, and actions become more coordinated and synchronized. With increased synchrony across modalities, older infants and toddlers may have access to additional cues that facilitate anticipation, including language (Gampe & Daum, 2014).

4.1 | Links between action prediction and sensorimotor development

Action prediction was associated with hand-eye coordination and joint attention. Specifically, 9-month-olds who made more anticipations also demonstrated stronger hand-eye coordination and experienced more bouts of child-led joint attention. Hand-eye coordination and joint attention were strongly correlated both at 9 and at 15 months of age. Interestingly, action prediction was correlated with the frequency but not the proportion of time spent in hand-eye coordination. This could indicate that action prediction skills are related to the ability to establish hand-eye coordination, but not to maintain it.

One explanation for this pattern of associations between action prediction, joint attention, and hand-eye coordination is that a general attention mechanism drives the development of these related cognitive skills. Decades of research have shown that attention plays an important role across multiple developmental domains, including learning (Markant & Amso, 2016), memory (Reynolds & Romano, 2016), and language (Kannass & Oakes, 2008). Markant and Amso (2016) point out that visual attention is the primary way in which young infants explore their environment and create opportunities for learning. As infants' visual attention skills develop, they likely become better able to coordinate their eye movements with their own actions (hand-eye coordination) and with the actions and eye movements of other people (action prediction and joint attention).

An alternative possibility is that action prediction relates to joint attention through hand-eye coordination. Our findings show that action prediction is correlated with hand-eye coordination, and that hand-eye coordination strongly correlates with joint attention both at 9 and at 15 months of age. Action prediction also correlates with joint attention, but the relationship is weaker. It could therefore be that emerging sensorimotor coordination skills facilitate infants' abilities to attend to and anticipate their parents' actions and action goals. This explanation is consistent with recent evidence from Yu and Smith (2017). These researchers showed that joint attention—also defined as synchronous looking to an object–emerges from the coordination between gaze and manual actions with objects rather than from gaze following, as traditionally assumed (Yu & Smith, 2017). Their central hypothesis is that actions, when coordinated with gaze, provide clear and redundant cues that help teach infants social and communicative behavior. This idea is also supported by the body of research showing that

sensorimotor experiences drive the broader development of infant social cognition and learning abilities (von Hofsten & Rosander, 2018). Although we cannot draw strong conclusions about causal mechanisms given that our data are correlational, our findings suggest that action prediction plays an important role in early social cognition and learning. Future work could extend these correlational findings with experimental studies to test the causality of the links between these developmental factors.

A noteworthy finding that emerged from our data is that action prediction was associated with child-led, but not parent-led, joint attention. Infants who anticipated their parents' actions more frequently also experienced more child-led joint attention. Child-led joint attention represents moments in which the parent followed the focus of their child's attention, suggesting that the association between action prediction and joint attention relates to some aspect of the _parent's_ behavior, rather than the child's behavior. For instance, parents who are responsive to their child's attentional shifts can provide more opportunities for action anticipation. Several recent studies have also reported similar patterns of parental responsivity that have been linked to word learning in young children (Bornstein et al., 2008; Smith & Yu, 2012; Wass et al., 2018). Taken together, this work highlights the importance of parents' contributions to their infant's early sensory environment by providing optimal learning moments in the action domain.

These findings imply that action prediction in the wild is jointly created by the infant observer and the parent actor, and depends upon the coupling between them. If parents are attuned to the readiness of their infants and generate an action at the right moment, it will increase the chance that their infant can make a successful action prediction. Therefore, we should consider action anticipation during social interactions in a broader context—as a joint action between developing infant observers and developed caregiver actors, rather than only as the infant's ability predict observed actions. This raises an interesting question for future research: whether there are behavioral patterns from infants and/or parents immediately preceding anticipation opportunities that could reliably separate anticipated versus unanticipated actions.

A limitation of the present study is that while our paradigm is intended to capture naturalistic parent–child interactions, it is nevertheless limited to tabletop object play with three toys. At home, infants likely experience many interactions in which prediction plays a critical role, such as everyday routines like feeding or getting dressed. In future work, we plan to investigate action prediction during more structured action contexts, which may elicit different patterns of anticipatory behaviors. A second limitation of our study is that we are not able to identify the exact cues that triggered each action anticipation. We found that gaze following occurred during some action events, though it is unlikely to be a primary cue that facilitated anticipations. Another potential source of information are kinematic cues from parent hands at the onset of reaching, which we were not able to analyze in this paradigm. Prior research has addressed this question in screen-based experiments by systematically controlling available action cues. These studies have demonstrated that

infants can exploit various information sources to predict actions, including statistical structure (Monroy et al., 2017a), kinematic cues (Stapel et al., 2012), goal information (Cannon & Woodward, 2012), and their own motor system (Southgate et al., 2009). Although the key aim and contribution of our study is to show that infants make action predictions during naturalistic parent–infant interactions, a future step is to identify the underlying cues infants rely on to make these predictions.

4.2 | Conclusion

This study is the first to show that infants visually anticipate their parents' actions during free-flowing parent–infant play interactions. Our findings emphasize the rich, multisensory nature of infants' early sensorimotor experiences and suggest that action prediction is associated with other important developmental processes. They also highlight the importance of using methodologies that allow researchers to capture infant behaviors during live parent–child interactions, as a complement to traditional, well-controlled laboratory experiments. In sum, this study contributes to the growing literature demonstrating that action prediction reflects an important component in infants' social and cognitive development.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Claire Monroy https://orcid.org/0000-0002-5044-5185

ENDNOTES

[1] Though operational definitions of joint attention differ across studies, in the current study we adopt the definition from Yu & Smith (2013) which defines joint attention as simultaneous gaze fixations to the same location. Joint attention is further divided into bouts, based on whether the child or parent initiated the gaze shift to the jointly attended object (child-led joint attention vs. parent-led joint attention).

[2] We did not include parent hands as an ROI, because head-mounted eyetracking data does not allow human coders to reliability determine the focus of gaze when (dynamically moving) ROIs are overlapping, as is often the case when parents are holding objects.

REFERENCES

Abney, D., Karmazyn, H., Smith, L., & Yu, C. (2018). Hand-eye coordination and visual attention in infancy. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, 1268-1273.

Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., de Looper, A., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PLoS One*, 8(7), e67916.

Bambach, S., Crandall, D., & Yu, C. (2013). Understanding embodied visual attention in child-parent interaction. In *The Third IEEE International Conference on Development and Learning and on Epigenetic Robotics*, 1-6.

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C.-S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology*, 44(3), 867-874.

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

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, 15(2), 292-298.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. In C. Moore & P. Dunham (Eds.), *Joint attention: its origins and role in development*, Vol. 63. Psychology Press.

Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology*, 41(5–8), 342-361.

Chen, C., Castellanos, I., Yu, C., & Houston, D. M. (2020). What leads to coordinated attention in parent–toddler interactions? Children's Hearing Status Matters. *Developmental Science*, 23(3). https://doi.org/10.1111/desc.12919

Dale, R., Kirkham, N. Z., & Richardson, D. C. (2011). The dynamics of reference and shared visual attention. *Frontiers in Psychology*, 2, 355.

Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy*, 23(1), 4-28.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, 9(7), 878-879.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, 424(6950), 769-771.

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.

Gampe, A., & Daum, M. M. (2014). Productive verbs facilitate action prediction in toddlers. *Infancy*, 19(3), 301-325.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1), 6.

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, 46(2), 446-454.

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.

Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *Journal of Neuroscience*, 29(19), 6234-6238.

Kannass, K. N., & Oakes, L. M. (2008). The development of attention and its relations to language in infancy and toddlerhood. *Journal of Cognition and Development*, 9(2), 222-246.

Kretch, K. S., & Adolph, K. E. (2017). The organization of exploratory behaviors in infant locomotor planning. *Developmental Science*, 20(4), e12421.

Krogh-Jespersen, S., Liberman, Z., & Woodward, A. L. (2015). Think fast! The relationship between goal prediction speed and social competence in infants. *Developmental Science*, 5, 815-823.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41, 3559-3565.

Markant, J., & Amso, D. (2016). The development of selective attention orienting is an agent of change in learning and memory efficacy. *Infancy*, 21(2), 154-176.

Meltzoff, A. N., & Brooks, R. (2007). Eyes wide shut: The importance of eyes in infant gaze following and understanding other minds. In R. Flom, K. Lee & D. Muir (Eds.), *Gaze following: Its development and significance* (pp. 217-241). Erlbaum.

Meyer, M., Bekkering, H., Haartsen, R., Stapel, J. C., & Hunnius, S. (2015). The role of action prediction and inhibitory control for joint action coordination in toddlers. *Journal of Experimental Child Psychology*, 139, 203-220.

Monroy, C., Gerson, S., & Hunnius, S. (2017a). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology*, 157, 14-28.

Monroy, C., Gerson, S., & Hunnius, S. (2017b). Infants' motor proficiency and statistical learning for actions. *Frontiers in Psychology*, 8, 2174.

Monroy, C., Meyer, M., Schröer, L., Gerson, S. A., & Hunnius, S. (2019). The infant motor system predicts actions based on visual statistical learning. *NeuroImage*, 185, 947-954.

Ossmy, O., Han, D., Cheng, M., Kaplan, B. E., & Adolph, K. E. (2020). Look before you fit: The real-time planning cascade in children and adults. *Journal of Experimental Child Psychology*, 189, 104696.

Paulus, M., Schuwerk, T., Sodian, B., & Ganglmayer, K. (2017). Children's and adults' use of verbal information to visually anticipate others' actions: A study on explicit and implicit social-cognitive processing. *Cognition*, 160, 145-152.

Pérez-Edgar, K., MacNeill, L. A., & Fu, X. (2020). Navigating through the experienced environment: Insights from mobile eye tracking. *Current Directions in Psychological Science*, 29(3), 286-292.

Reddy, V., Markova, G., & Wallot, S. (2013). Anticipatory adjustments to being picked up in infancy. *PLoS One*, 8(6), e65289.

Reynolds, G. D., & Romano, A. C. (2016). The development of attention systems and working memory in infancy. *Frontiers in Systems Neuroscience*, 10, 15.

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1(2), 353-367.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in cognitive sciences*, 10(2), 70–76.

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.

Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C.-H., Franchak, J. M., Pearcy, D., Suarez-Rivera, C., Xu, T. L., Zhang, Y., Smith, L. B., & Yu, C. (2018). Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior. *Journal of Visualized Experiments*, 141, e58496.

Smith, L. B., & Yu, C. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244-262.

Southgate, V., Johnson, M. H., Karoui, I. E., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science*, 21(3), 355-359.

Southgate, V., Johnson, M. H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology Letters*, 5(6), 769–772.

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 Psychologische Forschung*, 76(4), 434-445.

Stapel, J.C., Hunnius, S., & Bekkering, H. (2015). Fifteen-month-old infants use velocity information to predict others' action targets. *Frontiers in psychology*, 6, 1092.

Stapel, J. C., Hunnius, S., Meyer, M., & Bekkering, H. (2016). Motor system contribution to action prediction: Temporal accuracy depends on motor experience. *Cognition*, 148, 71-78.

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, 55(1), 96-109.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. J. Dunham (Eds.). *Joint attention: Its origins and role in development*. Lawrence Erlbaum Associates Inc.

von Hofsten, C., & Rosander, K. (2018). The development of sensorimotor intelligence in infants. *Advances in Child Development and Behavior*, 55, 73-106.

Wass, S. V., Clackson, K., Georgieva, S. D., Brightman, L., Nutbrown, R., & Leong, V. (2018). Infants' visual sustained attention is higher during joint play than solo play: Is this due to increased endogenous attention control or exogenous stimulus capture? *Developmental Science*, 21(6), e12667.

Xu, T., Abney, D., & Yu, C. (2018). Discovering multicausality in the development of coordinated behavior. In *The 39th Annual Meeting of the Cognitive Science Society* (pp. 1369-1374).

Yoshida, H., & Smith, L. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, 13(3), 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), e79659.

Yu, C., & Smith, L. B. (2016). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science*, 41, 5-31.

Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development*, 88(6), 2060-2078.

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, 18, e12735.

Zeger, S.L., Liang, K.Y., & Albert, P.S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 1049–1060.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Supplementary Material

How to cite this article: Monroy C, Chen CH, Houston D, Yu C. Action prediction during real-time parent-infant interactions. *Dev Sci*. 2021;24:e13042. https://doi.org/10.1111/desc.13042