The Sensorimotor Dynamics of Joint Attention

Sara E Schroer (saraschroer@utexas.edu)
Chen Yu (chen.yu@austin.utexas.edu)
Department of Psychology
University of Texas at Austin, Austin, TX

Abstract

Social interactions are composed of coordinated, multimodal behaviors with each individual taking turns and sharing attention. By the second year of life, infants are able to engage in coordinated interactions with their caregivers. Although research has focused on the social behaviors that enable parent-infant dyads to engage in joint attention, little work has been done to understand the sensorimotor mechanisms underlying coordination. Using wireless head-mounted eye trackers and motion sensing, we recorded 31 dyads as they played freely in a home-like laboratory. We identified moments of visual joint attention, when parent and infant were looking at the same object, and then measured the dyad's head and hand movements during and around joint attention. We found evidence that both parents and infants still their bodies during joint attention. We also compared instances of joint attention that were led by the parent or by the infant and identified different sensorimotor pathways that support the two types of joint attention. These results provide the foundation for continued exploration of the critical role of sensorimotor processes on coordinated social behavior and its development.

Keywords: action, attention, children, eye tracking, interactive behavior

Introduction

The "social dance" of coordinating multimodal behaviors and attention with social partners plays an important role in early development. Turn-taking, attending to one another's actions, and the bidirectional influences needed for successful social interactions are similar to the coordination of two individuals dancing. Of course, the social dance is not just a metaphor. Recent work has shown locomotor and sensorimotor coordination between infants and their parents. Synchrony, in the form of high similarity in spatiotemporal paths of movement, has been observed between mothers and infants in the second year of life as they explored a lab environment (Hoch, Ossmy, et al., 2021). By 3-months-old, infants can predict when their parent will pick them up and will adjust their posture in anticipation (Reddy, Markova, & Wallot, 2013). And 9-month-old infants can accurately predict their parents' actions during play, such as visual anticipation of parents' reaches (Monroy et al., 2020).

Research on adult-adult dyads has shown that coordinated behaviors are apparent across modalities, including speech, posture, and gaze (reviewed in Sebanz, Bekkering, & Knoblich, 2006; Shockley, Richardson, & Dale, 2009). Even while conversing with a stranger in a different room, adult social partners coupled their eye movements to matching screen-based stimuli (Richardson, Dale, & Kirkham, 2007). Coordination and mimicry of behaviors happen spontaneously in dyads and promote positive feelings about social partners (Chartrand & Bargh, 1999). Coordination requires that social partners share attention to and knowledge about the environment, be able to predict the other's actions, and adjust their actions based on the behavior of their partner (Sebanz et al., 2006). There is ample evidence that infants are capable of these "requirements" by the second year of life (e.g., Yu & Smith, 2013; Reddy et al., 2013; Monroy et al., 2020). In both developmental and adult research, it is apparent that social interactions are built from these coordinated sensorimotor behaviors.

One type of early coordination that is often studied is joint attention (JA), when an infant and parent share attention to the same object or task. As outlined in Siposova & Carpenter (2019) as well as Gabouer & Bortfeld (2021), the term JA has been used by researchers to describe a wide variety of behaviors. Recent work studying behavior at the level of milliseconds and seconds has examined how dyads engage in JA during play, defined as moments when parent and infant gaze at the same object (Yu & Smith, 2013; 2016; 2017). Head-mounted cameras and eye trackers measured visual attention at a rate of 30 frames/sec (the "micro level"), allowing researchers to study the attention of the "leader" and "follower" of JA in the moments before JA. Although decades of work have suggested gaze following is necessary to establish JA (e.g., Mundy & Newell, 2007), this hypothesis is predominantly supported by laboratory studies using videos of a social partner (e.g., Byers-Heinlein et al., 2020). One "surprising" finding from head-mounted eye tracking studies is that infants rarely look at their parent's face during naturalistic toy play, and infant gaze following only contributes to the establishment of 10% of JA instances (Yu & Smith, 2016; 2017; Deák et al., 2018). Other pathways must be used to establish JA in real-time naturalistic interactions – and hands are the key.

During the majority of JA instances, the attended object is held by the infant or parent. Crucially, not only do parents and infants attend to their own hands, but the hands of each other as well. This interpersonal hand-eye coordination and ability of a dyad to "hand follow" is predictive of how often dyads enter into JA (e.g., Yu & Smith, 2013, 2017). If parents

use multimodal behaviors (touching the object, talking), the dyad will enter JA more readily, stay in JA for longer, and the infant's visual attention, manual activities, and hand-eye coordination are all extended (a result seen both at home and in the lab; Deák et al., 2018; Suarez-Rivera, Smith, & Yu, 2019; Schroer, Smith, & Yu, 2019). Recent work also suggests that explicit cues of being in JA exist in infant's field of view (or what they can see, as captured by a head camera) *without* needing to look at a social partner's face, as a model was able to accurately classify frames into instances of JA or not JA solely based on the images (Peters et al., 2021). One pitfall of third-person coding to determine whether a dyad "knows" they are in JA (as specified in Siposova & Carpenter, 2019), is the potential for subjective, adult-centric judgements about whether a cue exists. To better understand the real-time sensorimotor dynamics underlying the coordination of attention, we will define JA at the micro-level as moments of shared visual attention to an object.

Despite the importance of hands and object holding for establishing JA, the role of infants' and parents' entire bodies and their patterns of movement has not been studied. Evidence from adult research suggests that there should be coordination beyond visual attention. To gain a richer understanding of how parent-infant dyads engage in JA, we studied sensorimotor processes at a timescale far more granular than most developmental research. In a home-like lab, dyads moved freely around while wearing wireless head-mounted eye trackers and motion sensors. Infants and their parents crawled, walked, and climbed throughout the experiment. At the level of milliseconds, we measured the movement of parents' and infants' heads and hands. In particular, we were interested in the distance between dyads and their speed in the moments leading up to and during JA.

We hypothesized that there would be temporally aligned changes in parent and infant body movement. Relationships between child attention and movement have been previously studied during solo-play activities. Long bouts of attention, or sustained attention, are accompanied by sensorimotor and physiological changes – including stilling the body, an intent expression, manipulation of the object being attended to (e.g., rotating the object), and a deceleration in heart rate (Ruff & Lawson, 1990; Lansink, Mintz, & Richards, 2000). Based on this work, we expected to see a decrease in infant's head movement during JA. Regarding hand movement, we proposed two alternative hypotheses: 1) hand speed would decrease, in line with the idea that the body stills while in sustained attention, or 2) hand speed would increase because hands and object manipulation play a major role in shaping JA. We also hypothesized that parents would show similar body movements as their infants, in line with work on dyadic spatiotemporal coordination. In addition, we chose to measure distance between parent and infant because the effects of interpersonal distance on social behaviors in parent-infant interactions have not been widely studied.

Methods

Participants and Data Collection

31 infants and a parent were recruited from a Midwest college town. Infants were all in the second year of life (mean: 16.6mo, range: 12.6-23.5mo). Based on previous research, we expected that most infants in this age range would be crawling and/or walking (Adolph & Berger, 2007) as well as readily engaging in JA (e.g., Yu & Smith, 2013). Dyads were given 10 objects to play with for 10min. We told parents that they were not required to play with the provided toys and that they and their child could move freely around the play area (approximately 3m x 3m, Figure 1a). Although not all participating infants could walk at the time of the experiment,

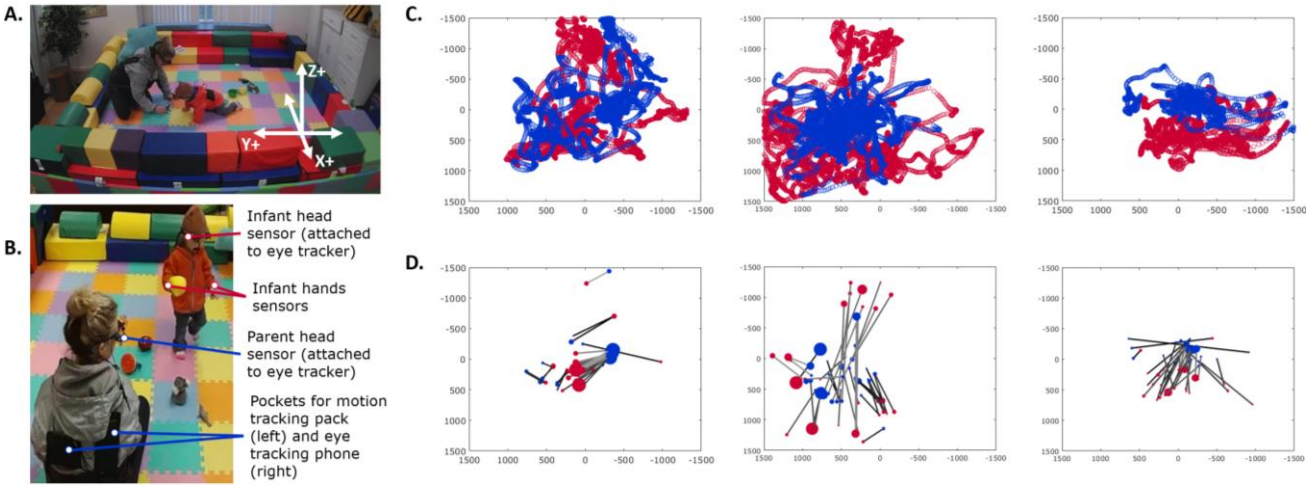

Figure 1: A) The play area with x-, y-, z-axes overlayed (origin is the center of the area). B) Parent and infant wore eye trackers and jackets equipped with 3 motion sensors. C) Spatiotemporal plots showing the movement of 3 different infant (red) and parent (blue) dyads. D) The locations of the same dyads when in JA. Location of infant (red) and parent (blue) during JA are represented with two dots connected by a line. The size of the dots represents how long the JA bout lasted. The line's color corresponds to when the bout occurred during the play session, with shade getting lighter as time passed.

all were capable of independent locomotion. Parent and infant wore head-mounted eye trackers, as well as three motion sensors, one on each wrist and head. To accommodate the wearable sensors, participants wore jackets with pockets on the back and a small pocket at each wrist (Figure 1b).

Eye Tracking and Defining Attention

Dyads wore wireless Pupil Labs head-mounted eye trackers connected to an Android smart phone with a USB-C cord. The smart phones were placed in a back pocket of the custom jackets and the cord was tucked behind the participant's head, out of their field of view. Parents wore the standard eye tracker like a pair of eyeglasses, while the infants wore a modified version of the eye tracker that could be attached to a hat using Velcro. The eye trackers consisted of two cameras: the scene camera was centered on the participant's forehead to record their field of view and the other was pointed back at their eye to record eye movements. Following the experiment, the eye tracking videos were calibrated to produce a fixation crosshair that indicated participant's gaze every frame of the video (30 frames/s). Participant's visual attention was coded using an in-house program, each fixation to one of the toys or their social partner's face was annotated.

JA was defined during data pre-processing. Each frame that a dyad was attending to the same object (or each other's faces) was marked as JA. In line with previous work studying coordination at a fine temporal scale, JA was defined as lasting at least 500ms and could include short looks (< 300ms) away from the attended object (Yu & Smith, 2017). As soon as one participant looked away (without returning), the bout of JA was terminated. Three instances of JA lasted longer than 15s and were excluded from the current analyses.

In addition to the eye trackers, multiple 3rd-person-view cameras recorded the experiment. These cameras were used to annotate which object(s) a participant's left hand and right hand were touching. This coding was also done frame-by-frame with an in-house program.

Motion Sensing and Defining Movements

Participants also wore three motion sensors during the experiment (Polhemus Liberty). Two sensors were placed in the small pockets on the participant's wrist (to approximate hand movements) and a third sensor was attached to the eye tracker. The head sensor was used to approximate movement of the body throughout the play space (e.g., walking around). The sensors were attached to a battery pack that was placed in the jacket's other back pocket (Figure 1b). Motion was tracked by measuring the movement of the sensors relative to two central hubs that were placed under the play area. The system captured movement in 6 degrees of freedom: positional displacement on the x-, y-, and z-axes (see Figure 1a for a visualization of the axes), as well as rotational displacement (pitch, yaw, and roll). We will discuss the positional data in this paper. Motion data was recorded at an original sampling rate of 240 Hz, but down-sampled to 30 Hz to match the eye tacking data.

For the current study, we were interested in interpersonal distance and positional speed. Interpersonal distance was defined in 2-dimensions (x and y) as the distance between the infant and parent head sensors at every frame of the experiment (in mm). Speed of each sensor was calculated in 3-dimensions for every frame and is reported in mm/s.

Analysis Plan

We first analyzed the **dynamics of JA** to see how movement before and during JA differed from times the dyad was not in JA. We conducted event-level analyses, with all instances of JA analyzed as one large corpus (N = 1431). When appropriate, we included random effects of subject contributing the data point and the object being attended to, to account for differences that might exist across participants. We then calculated a baseline value of "not JA". Not JA was defined as the periods of time between the offset of one JA bout and the onset of the next. To control for differences in movement that may arise simply from the duration of a bout, we limited our sample to bouts of not JA that were longer than 500ms and shorter than 15s (to match the definition of JA, N = 1256). We calculated the average speed of each sensor as well as the average interpersonal distance for outside of bouts of JA – these were our baseline values.

We used temporal profiles to visualize the average movement of the dyad from the 5s leading up to the onset of JA to the 5s after the onset of JA, as well as to compare these speeds to baseline. To further explore speed in JA, we looked at the mean values of speed and distance. We analyzed how movement related to the duration of JA and compared how parents and infants moved.

Our second set of analyses compared movement during **parent-led and child-led JA**. The goal of this set of analyses was to examine how dyads enter into JA – replicating previous analyses on gaze following and hand following, as well as novel analyses on how movement shapes and defines the different pathways dyads use when parent or infant leads JA. Similar to the analyses on the dynamics of JA, we conducted event-level corpus analyses. We first compared the temporal profiles of parent-led and child-led JA and then looked at the mean values of speed and distance within JA and in the 2s before each type of JA.

Results

Joint Attention and Body Movement

Table 1 reports the means and standard deviations of duration, distance, and speeds within JA and the baselines (means) of not JA. For each sensor, mean speed in JA bouts was less than the mean speed in not JA bouts (*ps* <= 0.011).

We first visualized the spatiotemporal paths of the dyads to learn about the patterns of their movement (Figure 1c). There were noticeable differences in how dyads moved, but most dyads explored the majority of the play space during the experiment. We also visualized dyads' locations during JA (Figure 1d). Even within a dyad, there was no trend in how far apart dyads were during JA. As will be discussed, there

Table 1: Descriptive statistics and baselines values

	JA	not JA
Number of events	1431	1256
Duration	2.100 (1.88)	4.494(3.46)
Distance (mm)	495.746 (256.55)	528.149
Child speed (mm/s)		
Head	78.516 (78.72)	94.448
Left hand	87.413 (93.81)	100.111
Right hand	101.53 (103.93)	112.175
Parent speed (mm/s)		
Head	69.113 (86.37)	89.351
Left hand	73.363 (126.01)	97.507
Right hand	98.068 (136.91)	123.295

was no significant difference in interpersonal distance when the dyad was in JA or not, so distance was excluded from the temporal analyses.

The Dynamics of Joint Attention

The temporal analyses took 299 "slices" of data, plotting the mean speed every 30ms from the 5s before to the 5s after the onset of a JA bout. At each slice, we performed a one-sample t-test, comparing speed at that time point to the baseline speed. Figure 2 provides a visualization of the temporal profiles and significance testing. Figures 2a-b show the temporal profiles of child and parent head speed during JA. Figure 2c shows the segments of data that were significantly above or below baseline for the left and right hands of child and parent.

For all sensors, there was a significant decrease in speed during JA. In the ~500ms leading up to JA, however, the speed of the infant's head and hands sharply increased. The infant's head then remains stilled (below baseline) through ~4s after onset of JA and their left hand for ~1s. The infant's right hand shows a different pattern – the increase in speed before JA is more pronounced and speed only falls significantly below baseline for a handful of frames within JA (although there is the trend of decreasing speed). This pattern of movement could be a result of the infant often being the one to hold objects during JA (as will be discussed in the next section). All parent sensors showed a consistent trend of stilling while in JA – head speed stilled for ~2s and hands for ~2.5s. Overall, the temporal profiles show that the body stills during JA. The stilling occurs in both infants and their parents, which is in-line with previous research on changes to movement when young children sustain attention as well as work showing dyadic sensorimotor coordination.

We then analyzed the mean values of speed and distance within instances of JA. Further demonstrating the importance of stilling the body to sustain attention, the stability of the infant's head can predict the duration of the JA event. We used linear mixed effects regressions with random effects included for subject and object (lmerTest package for R; Kuznetsova, Brockhoff, & Christensen, 2017) and then used a Chi-Square difference test to compare the full model to a null model with intercept and random effects only. Child head speed was only a weak negative predictor of the duration of JA (β = -0.003, $p < 0.001$), but including speed was an improvement from the null model ($\chi2 = 21.247$, $p < 0.001$). We saw a similar relationship for parent head speed and duration (β = -0.002, $p = 0.002$; $\chi2 = 9.311$, $p = 0.002$).

We also observed differences in how parents and infants moved. Within JA, infants had higher head and left-hand speeds than their parents ($ps < 0.001$ using paired t-tests). There was no difference in right-hand speed ($p = 0.422$), perhaps because of the tendency for right-handedness (so both parents and infants would move this hand a lot).

Lastly, although speed changes as a dyad enters and exits JA, distance does not relate to whether or not they are in JA. We compared the mean distances between dyads during each bout of JA to the mean distances of dyads during the instances of "not JA". There was no significant difference ($t(2681.9) = 0.002$, $p = 0.998$; using a Welch two sample t-test).

To summarize, dyads flexibly change their head and hand movements as they enter into and leave JA. Parents and infants differ in how quickly they move and the speed of the dyad is even weakly correlated with the duration of JA. These findings suggest that the "motor" aspect of sensorimotor behaviors plays an equally important role in shaping the multimodal pathways into visual JA. To explore these different pathways, we turned our attention to child-led and parent-led JA.

Comparing Child-led and Parent-led JA

We defined child-led JA as instances of JA in which the child was looking at the attended object for at least one frame before the parent joined. There were 752 instances of child led-JA, with a mean duration of 2.160s. Parent-led JA was

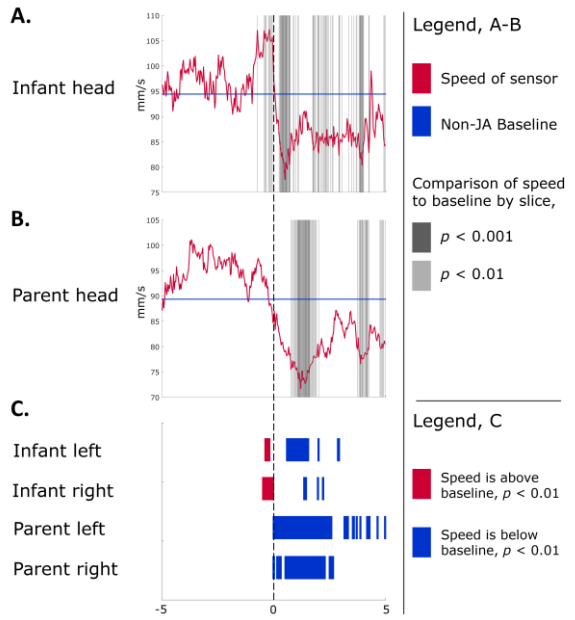

Figure 2: Temporal profiles during JA for A) infants' and B) parents' heads. C) summarizes results for all hands.

similarly defined as instances of JA in which the parent was looking at the attended object for at least one frame before the child joined. There were 512 instances of parent-led JA with a mean duration of 2.236s. There was no difference in the duration of child-led and parent-led instances (p = 0.501). There were an additional 167 instances in which no leader could be determined; these were excluded from analyses.

Replicating previous findings, gaze-following was an underused pathway for entry into JA. Gaze-following was defined as the follower looking to their social partner's face in the lag between the leader's onset of looking at the attended object and the onset of the follower's look. While parents followed their infant's gaze in about 34.4% of instances, infants only used gaze following in 3.5% of instances (only 18 times across all 31 subjects). Infant's use of gaze following in this more naturalistic lab setting is even lower than was previously reported in constrained, face-to-face table-top play (Yu & Smith, 2017). Instead, it seemed like hand-following was the predominant pathway.

The "holder" of the object during JA was then assigned by comparing the proportion of the bout the parent and child each touched the object. Whoever touched the attend object for longer was the holder. In only 7% of both child-led and parent-led instances neither member of the dyad held the attended object. Infants held the attended target in a greater proportion of instances than their parents, though this difference was more pronounced in child-led JA (62.9% instances child was holding, 27.5% parent holding) than parent-led instances of JA (48.6% child, 42.2% parent). In the remaining 2% of instances, parents and infants held the object for the same amount of time during a bout of JA.

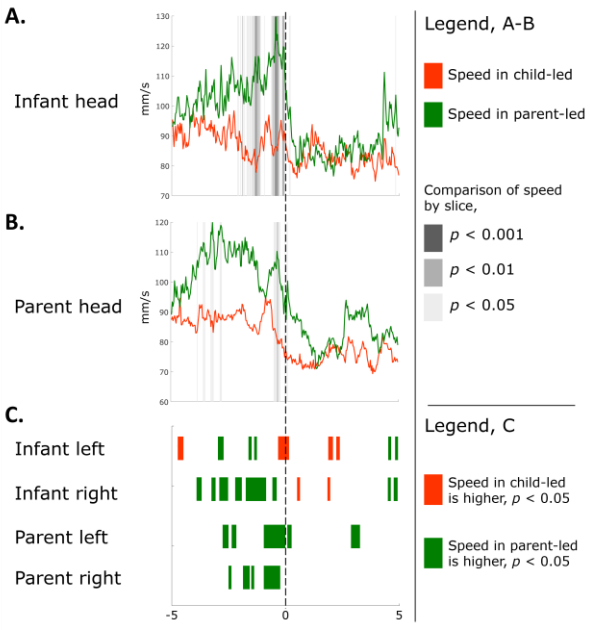

Figure 3: Temporal profiles during child-led and parent-led JA for A) the infants' and B) parents' heads. C) summarizes the results for all hands.

We first analyzed the temporal profiles of parent-led and child-led JA. As before, the temporal analyses plotted and analyzed 299 "slices" of data – the mean speed of each sensor every 30ms from the 5s before to the 5s after the onset of a JA bout. The temporal profiles of parent-led and child-led JA were compared at every slice using generalized mixed effects models to predict the type of JA by sensor speed with random effects for subject and attended object (fitglme in Matlab). Figure 3 provides a visualization of the temporal profiles and significance testing, using the same format as Figure 2.

Once the dyad was engaged in JA, there were few differences between speed in parent-led and child-led JA. Instead, the differences were mainly in the seconds before JA begins. Child head speed is slower in the 1.5s before child-led instances, presumably because their head has already stilled when they are the leader of JA. The other major difference in the temporal profiles is parents' hands. In the second before parent-led JA, parents' hands move faster (perhaps parents are bringer objects nearer to their infant).

To further explore the differences in parent-led and child-led joint attention, we then compared the mean values of speed within JA and in the 2s before JA using two-sample t-tests. We first compared the speed of parents and infants when they were the leader. Within bouts of JA, there was no significant difference in how fast the leaders moved their head or hands (ps > 0.095). Reflecting the findings in the temporal profile, we did observe differences between parents and infants in the 2s before JA. Parent leaders' heads and right hands moved faster than infant leaders' in the 2s before JA (head: p = 0.032, right hand: p < 0.001). We also saw differences when comparing parent and infant followers. When following, infants moved their head and left hand more than their parents did when following both during and in the 2s before the bout of JA (ps <= 0.003).

We then examined whether there was a decrease in speed once the dyad was in JA. We calculated the mean speed of the sensors during and in the 2s before each bout of JA and compared the mean speeds of each instance using paired t-tests. Both parent and child decreased the speed of their heads and hands once they were in a bout of JA, regardless of who led the JA bout (ps < 0.001). Despite the decrease in hand speed during JA, infant hand speeds within JA were *higher* when the infant was holding the attended object (left hand: p = 0.025; though right hand was trending, p = 0.052). Parents showed the same higher hand speed when holding the attended objects (ps <= 0.003).

To summarize, we replicated previous findings that hand-following and object manipulation is a highly used pathway into JA during free-flowing play. Although there were no differences in the temporal profiles of child-led and parent-led JA during the bouts of JA, parents and infants adjusted their speed in the seconds before JA differently depending on who was leading the bout. Additionally, parents and infants moved at similar speeds when leading JA, but following infants moved more than following parents. Finally, we observed that hands moved more within JA if the attended object was being held. Together, these results show that even

though average speed was decreasing in JA, the attended object was still being actively manipulated – meeting both criteria of sustained attention: stilling of the body, but manipulation of an object.

Discussion

We present evidence of coordinated sensorimotor behaviors when dyads engage in JA. At the onset of JA, dyads slow down. Once in JA, infants and parents will still their heads and hands, which may help to extend the bout of attention. When we compare child-led and parent-led JA, we see different sensorimotor pathways. In child-led JA, infants already begin to still their movements before their parent enters JA – perhaps signaling to their parent that they are attending to an object – and the parent then slows their movement to join. In parent-led JA, however, we see a very different pattern. Parents move a lot in the seconds before the JA bouts they lead, suggesting that parents create these moments for the infant, bringing objects into the infant's field of view. Only once inside parent-led JA does the infant begin to still. Our goal was to measure movement to explore the coordination of dyads during JA, as well as to test multiple hypotheses based on previous sustained attention research. As expected, we observed a decrease in infant's head movement during JA and a similar decrease in parent's head movement. Surprisingly, we found evidence to support both of our alternative hypotheses regarding changes in hand movement, showing that object manipulation is a key feature of JA even as the body stills.

We cannot conclude with our current dataset whether these sensorimotor behaviors are a requisite of establishing JA or simply accompany JA. Our work, however, does show a clear temporal pattern. Dyads begin changing their speed in the seconds proceeding JA, suggesting that sensorimotor coordination is at least a step in the process of jointly attending to objects. These changes in movement may signal a readiness to enter into JA or even provide scaffolding to support and focus infant attention.

Considering the role of infants' and parents' entire bodies in social interactions is a promising research direction. Infants' bodies dramatically alter their visual environment relative to an adult – shorter arms bring objects close to their face and shorter overall height limits how much of the world is present in their field of view. As a result, infants' visual scenes are less cluttered, with fewer and larger objects present, and a held object can completely dominate their field of view (Yu & Smith, 2012). These differences in view, motor abilities, and embodiment may mean that parents and infants are solving the problem of coordination in different ways. Knowing how to establish and recognize moments of shared attention and predicting the actions of a social partner could be accomplished differently by an infant and their parent. Shockley et al. (2009) speculated that "coordination reflects a functional reorganization of body segments and eye movements to support the joint goals/actions" of an interaction. How this reorganization occurs in parent-infant interactions is yet to be understood.

Nonetheless, the interconnectedness of motor and social development offers exciting insights to the motor-social relationship within interactions. Every stage of motor development is accompanied by "social revolutions". The development of postural control and reaching brings objects into view and shifts attention away from faces (Fogel et al., 1999). Infants then attend to and explore objects alone, until increasing intrapersonal coordination supports their ability to attend to their partner's object manipulations too (de Barbaro et al., 2013). The transition from crawling to walking changes the way infants can bring objects to their parents, in turn changing the way parents respond to their infants' bids (Karasik, Tamis-LeMonda, & Adolph, 2014). These changes may pave the way for the vocabulary boom that accompanies the transition to walking (He, Walle, & Campos, 2015). At the fine-motor level, parents will selectively label objects when their infants engage in more mature object manipulation (West & Iverson, 2017). How infants move within an interaction will likely affect their social partner's behavior as well. It is easy to imagine that walking, climbing, and sitting will elicit different speech from their caregiver in the same interaction, even if the infant is holding and attending to the same object in these different positions.

Our findings lay the groundwork for studying the sensorimotor foundation of joint attention – and other social behaviors. Understanding the sensory and motor bases underlying joint attention will begin to unlock the answers as to how attention develops and the types of abilities infants need in order to attend to objects successfully, as well as provide new intervention opportunities for populations with developmental disorders.

Acknowledgments

This work was supported by NIH R01HD074601 and R01HD093792 to CY. SES was supported by the NSF GRFP (DGE-1610403) and NIH T32HD007475. We thank Christian Jerry, Dian Zhi, and the members of the Computational Cognition and Learning Lab at Indiana University and the Developmental Intelligence Lab at UT Austin for their support in data collection and coding.

References

Adolph, K. E., & Berger, S. E. (2007). Motor development. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 161–213). New York: Wiley.

Byers-Heinlein, K., Tsui, R.K.Y., van Renswoude, D., Black, A. K., Barr, R., Brown, A., ... & Singh, L. (2020). The development of gaze following in monolingual and bilingual infants: A multi-laboratory study. *Infancy*.

de Barbaro, K., Johnson, C. M., & Deák, G. O. (2013). Twelve-month ''social revolution" emerges from mother-infant sensorimotor coordination: A longitudinal investigation. *Human Development*, 56(4).

Chartrand, T.L., & Bargh, J.A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, *76*(6).

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).

Fogel, A., Messinger, D.S., Dickson, K.L., & Hsu, H.C. (1999). Posture and gaze in early mother–infant communication: Synchronization of developmental trajectories. *Developmental science, 2*(3).

Gabouer, A., & Bortfeld, H. (2021). Revisiting how we operationalize joint attention. *Infant Behavior and Development*, 63, 101566.

He, M., Walle, E.A., & Campos, J.J. (2015). A cross-national investigation of the relationship between infant walking and language development. *Infancy, 20*(3).

Hoch, J.E., Ossmy, O., Cole, W.G., Hasan, S. and Adolph, K.E. (2021), "Dancing" together: Infant–mother locomotor synchrony. *Child* Development.

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).

Kuznetsova A., Brockhoff P.B., & Christensen R.H.B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software, 82*(13).

Lansink, J. M., Mintz, S., & Richards, J. E. (2000). The distribution of infant attention during object examination. *Developmental Science, 3*(2).

Monroy, C., Chen, C.H., Houston, D., & Yu, C. (2020). Action prediction during real-time parent-infant interactions. *Developmental Science*.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current directions in psychological science*, 16(5), 269-274.

Peters, R. E., Amatuni, A., Schroer, S. E., Naha, S., Crandall, D., & Yu, C. (2021). Are you with me? Modeling joint attention from egocentric vision. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*.

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

Richardson, D.C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological science, 18*(5).

Ruff, H.A., & Lawson, K.R. (1990). Development of sustained, focused attention in young children during free play. *Developmental psychology, 26*(1).

Schroer, S., Smith, L., & Yu, C. (2019). Examining the multimodal effects of parent speech in parent-infant interactions. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*

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

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. Topics in *Cognitive Science, 1*(2).

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, 189, 260-274.

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

West, K. L., & Iverson, J. M. (2017). Language learning is hands-on: Exploring links between infants' object manipulation and verbal input. *Cognitive Development*, 43.

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

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. (2016). The social origins of sustained attention in one-year-old human infants. *Current biology, 26*(9).

Yu, C., & Smith, L.B. (2017). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive science,* 41.