CHILD DEVELOPMENT

Child Development, xxxx 2017, Volume 00, Number 0, Pages 1–19

Hand–Eye Coordination Predicts Joint Attention

Chen Yu and Linda B. Smith
Indiana University

The present article shows that infant and dyad differences in hand–eye coordination predict dyad differences in joint attention (JA). In the study reported here, 51 toddlers ranging in age from 11 to 24 months and their parents wore head-mounted eye trackers as they played with objects together. We found that physically active toddlers aligned their looking behavior with their parent and achieved a substantial proportion of time spent jointly attending to the same object. However, JA did not arise through gaze following but rather through the coordination of gaze with manual actions on objects as both infants and parents attended to their partner's object manipulations. Moreover, dyad differences in JA were associated with dyad differences in hand following.

Momentary looking behavior is tightly tied to one's internal attentional state (Baron-Cohen, 1997; Brooks & Meltzoff, 2005; Frischen, Bayliss, & Tipper, 2007). It is for this reason that eye-tracking measures are widely used in behavioral research (Aslin & McMurray, 2004; Hayhoe & Ballard, 2005; Johnson, Amso, & Slemmer, 2003; Kingstone, Smilek, Ristic, Friesen, & Eastwood, 2003; Knoeferle & Crocker, 2006; Richardson & Dale, 2005; Yu & Smith, 2011). It is for this same reason that people attend closely to the eyes of their social partner and use the partner's gaze direction to establish the common ground necessary for smooth social engagements (Argyle, 2007; Bayliss et al., 2013; Corkum & Moore, 1995; Johnson, Slaughter, & Carey, 1998; Mundy & Newell, 2007). These smooth social interactions and coordinated visual attention, they require, are also central to healthy development in many domains. Individual differences in infants' and children's ability to coordinate visual attention with a social partner strongly predict individual differences in language, social, and cognitive development (Brooks & Meltzoff, 2005; Mundy & Gomes, 1998).

The traditional laboratory studies in which these predictive individual differences have been documented typically measure infants' *responses* to joint attention (JA) bids, that is, their ability to follow gaze shifts, head turns, and sometimes manual points so as to jointly attend to the same object with their social partner (Baldwin & Moses, 1996; Brooks & Meltzoff, 2005). In many of these experiments, the signals indicating the direction of attention of the mature partner were designed to be unambiguous (e.g., concurrent gaze and head shifts) and were repeated to ensure that the infant attended to them. Moreover, the spatial tasks in these laboratory settings were purposely simple with an experimenter directly facing the infant and with just two potential targets on opposite sides of midline. This task structure fits the goal of measuring and accessing the infant's ability to interpret gaze direction as a meaningful social cue, unhindered by potential limitations of the infant's spatial precision in interpreting gaze direction (Butterworth & Cochran, 1980; Corkum & Moore, 1995; Scaife & Bruner, 1975). Although individual differences in these tasks are related to individual differences in social and language skills (Brooks & Meltzoff, 2005; Mundy & Newell, 2007; Wellman, Phillips, Dunphy-Lelii, & LaLonde, 2004), the developmental origins of these individual differences are not known.

Everyday parent–infant interactions such as joint toy play are much messier than the "clean" and diagnostic laboratory tasks described above (also see Kingstone, Smilek, & Eastwood, 2008). The spatial context is often crowded with multiple potential targets close to each other (Deák, Walden, Yale

We thank Melissa Elston, Steven Elmlinger, Charlotte Wozniak, Melissa Hall, Charlene Tay, and Seth Foster for collection of the data, and Tian (Linger) Xu, Seth Foster, and Thomas Smith for developing data management and processing software. This work was funded by National Institutes of Health Grant R01HD074601 and R21 EY017843.

Correspondence concerning this article should be addressed to Chen Yu, Department of Psychological and Brain Sciences, Cognitive Science Program, Indiana University, 1101 East 10th Street, Bloomington, IN 47405. Electronic mail may be sent to chenyu@indiana.edu.

 0009-3920/2017/xxxx-xxxx
DOI: 10.1111/cdev.12730

Kaiser, & Lewis, 2008). Therefore, the spatial precision of head and eye direction may not be a sufficient cue to differentiate given multiple spatially close objects (Langton, 2000; Loomis, Kelly, Pusch, Bailenson, & Beall, 2008; Vida & Maurer, 2012). In everyday social interactions, infants do not just respond to JA bids but also initiate them (see Mundy & Newell, 2007). Thus, in everyday free-flowing interactions, the two partners may sometimes have competing attentional goals that need to be resolved if they are to share attention to the same object. In brief, everyday social interactions are spatially and dynamically complex and challenging, and thus likely to increase individual differences relative to laboratory tasks. The everyday interactions of parents and infants are also the likely training ground in which infants first learn to read social cues and to coordinate attention with partners (Bakeman & Adamson, 1984; Gredeback, Fikke, & Melinder, 2010; Triesch, Teuscher, Deák, & Carlson, 2006; Ullman, Harari, & Dorfman, 2012; Yu & Smith, 2013). Differences in these interactions then may be the source of individual differences in infants' developing abilities to read and send social cues. This is the overarching hypothesis that motivates the present study.

Within this larger framework, we focus on one key factor: manual actions on objects. Our hypothesis that actions on objects are critical to the establishment of JA between parents and infants was suggested by a prior study in which parents and their 12-month-old infants played together with toys (Rader & Zukow-Goldring, 2010; Yu & Smith, 2013). In that study, parents and infants both wore head-mounted eye-tracking systems that measured the momentary gaze direction of each partner and provided a precise measure of the coordination of visual attention to the same object. Consistent with findings from a growing number of studies (Aslin, 2009; Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; Franchak, Kretch, Soska, & Adolph, 2011; Rader & Zukow-Goldring, 2010; Yoshida & Smith, 2008), the gaze data indicated that infants rarely looked to their parent's face, a fact that precludes gaze following by the infant as a contributing factor to JA. The gaze data also indicated that the dynamics of infant visual attention were very different from those of parents, whereas parents rapidly shifted eye gaze among many visual targets generating a series of brief fixations, and infants generated long looks and showed sustained attention to an object, a property of infant attention during toy play that has been noted by other researchers (Ruff & Lawson, 1990). Nonetheless, parents' and infants' visual attention were often coordinated. Indeed, parents and infants not only often fixated the same object at the same time, but they often jointly shifted attention from one object to another in near unison, at the time scale of adult–adult interpersonal coordination (Marsh, Richardson, & Schmidt, 2009; Shockley, Santana, & Fowler, 2003).

How did they achieve this smooth coordination despite the different dynamics of parent and infant visual attention and despite the fact that the infants rarely looked to their parent's face? The eye-tracking data indicated a strong role of hand actions on objects: When infants manually interacted with an object, they looked at the object in contact with their own hands and parents also looked at those infant-handled objects. When parents manually contacted an object, they looked at the object in contact with their own hands, and infants also looked at the object being manipulated by the parent. In brief, because one's own eye gaze and one's own hand actions are spatially coordinated in goal-driven actions, directing visual attention to the object being manipulated by one's social partner will results in JA between the two partners to the same object without gaze following (Yu & Smith, 2013). Figure 1 shows two paths through which hand following may yield JA between infants and parents: (a) the infant handles an object and parent gaze follows the infant's hands to the object, and (b) the parent handles an object and the infant gaze follows the parent's hands to the object. Gaze following and hand following may be distinct routes to JA that require different sequences of behaviors by the follower. Gaze following requires the follower to look at the initiator's face and then switch attention to the spatial location to which the initiator's gaze is directed. In contrast, the hand-following pathway would seem to have just one step: looking at the object in contact with the partner's hand. Because hands and the handled objects are spatially close to each other, this makes hands a much more salient and robust cue to attention direction. Figure 1c also shows the gaze-following pathway to JA that has been the sole focus of previous research on JA. The present article focuses on the "hand-following" pathways to JA that are potential sources of individual differences in how well parents and infants can coordinate their visual attention to the same object. If hand following is the principle path to JA for parents and toddlers in joint object play, then dyad differences in JA should be associated with individual differences in these manual activity and hand–eye coordination components.

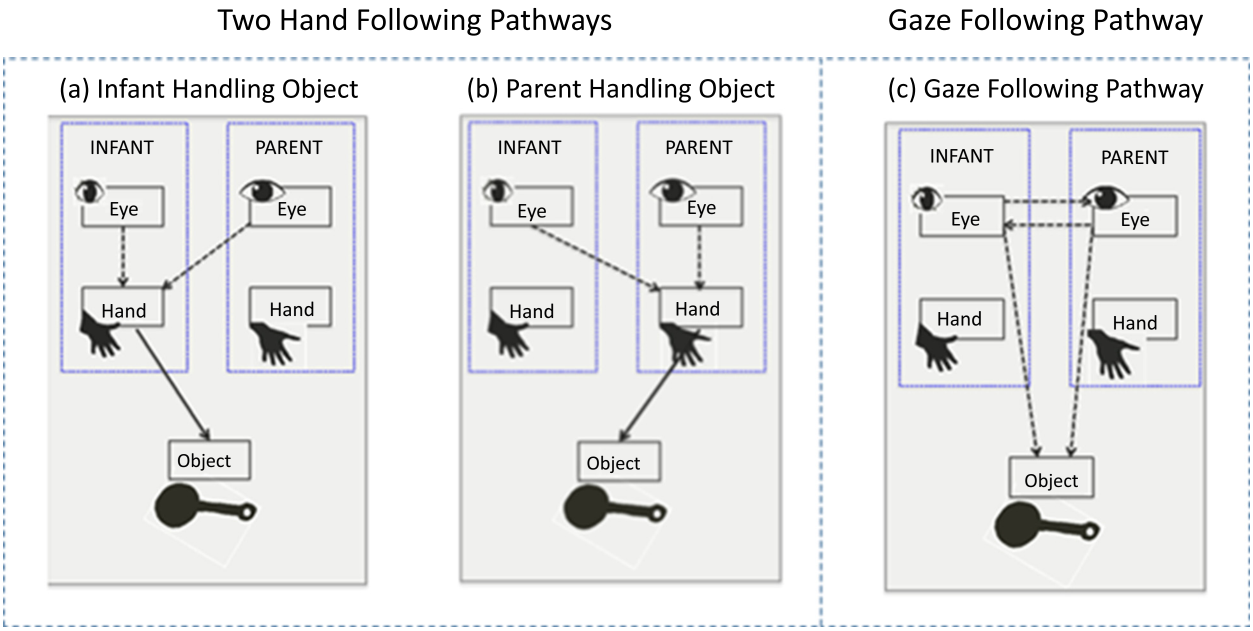

Figure 1. Hand-following and gaze-following pathways to joint attention (JA). (a) The infant holds an object and visually attends to his own hands as they handle the object and the parent attends to the infant's hand actions, leading to both parent gaze and infant gaze directed to the same object. (b) The parent holds the object and attends to her own hands as does the infant. The four dashed lined in (a) and (b) show the four hand–eye links that are the focus of the present study—hand–eye coordination within infant, hand–eye coordination within parent, parent eye to infant hand, and infant eye to parent hand. The traditional gaze-following pathway to JA to an object is shown in (c). [Color figure can be viewed at wileyonlinelibrary.com]

To test this hypothesis, we used a method similar to that in the previous dual eye-tracking study of 12-month-olds and their parents (Yu & Smith, 2013). The task context was free-flowing parent–infant play with multiple toys. Head-mounted eye-tracking systems were worn by both participants, allowing us to record eye in head position from both infants and parents during play. Gaze to and hand actions on objects by both parents and infants were recorded and coded. The infants participating in this study ranged in age from 11 to 24 months in an effort to capture a broader range of individual and dyad differences. We focused on the 2nd year of life because that is when infants become increasingly active and autonomous during this age range (Eckerman & Didow, 1989), because individual differences in both motor behavior and JA are noticeable during this period (Landa, Gross, Stuart, & Faherty, 2013), and because individual differences in manual actions on objects have been linked to differences in sustained attention (Ruff, 1986; Ruff, Capozzoli, & Weissberg, 1998), to parent talk about objects (Karasik, Tamis-LeMonda, & Adolph, 2014), and to language development (Iverson, 2010).

Our measure of JA was straightforward and transparent based on calculating, frame by frame, the moments that children and parents looked to the same object at the same time. Our measure of hand–eye coordination was one that is taken for granted in adult research: the systematicity with which hands and eyes are directed to the same object. More advanced measures of hand–eye coordination assume the spatial correspondence of hands and eyes and focus on precise timing and velocity profiles (Flanders, Daghestani, & Berthoz, 1999; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Johansson, Westling, Bäckström, & Flanagan, 2001; Sailer, Flanagan, & Johansson, 2005). However, the direction of hands and eyes to the same object is not certain in toddlers (Bertenthal & Von Hofsten, 1998; Bushnell & Boudreau, 1993; Eppler, 1995; Iverson, 2010; Lockman & McHale, 1989; Soska, Adolph, & Johnson, 2010). Accordingly, three sets of analyses were conducted: The first set focused on gaze patterns and JA. In preview, the frequency of these JA bouts was only weakly related to infant age, and dyad differences were much larger than differences related to infant age. The second set of analyses partitioned the dyads into low and high JA groups based on the frequency of their JA bouts and examined between-group differences in the components of the hand-following pathway—manual activity, within-individual hand–eye coordination, and between-partner hand–eye coordination.

The third set of analyses used regression to examine the associations among the components of hand–eye coordination as well as age as predictors of the frequency of JA during a dyad's toy play.

Method

Participants

For the main experiment, the final sample consisted of 51 (24 male toddlers) parent–toddlers dyads with the toddlers ranging in age from 11 to 24 months ($M = 17.92$, $SD = 4.15$); 14 additional dyads began the study, but the toddlers refused to wear the measuring equipment throughout the entire procedure. Data were collected between June 2013 and April 2014. Because the eye-tracking equipment on the parent could alter toddler gaze to the parent or the social interaction in some way, we also tested five toddlers (2 boys, between 16- and 20-month-old) in a version in which *only the toddler* wore the head-tracking gear to ensure that toddler gaze in the main experiment was not altered by the head gear worn by the parent. The entire sample of toddlers was broadly representative of Monroe County, Indiana (84% European American, 5% African American, 5% Asian American, 2% Latino, 4% Other) consisting of predominantly working- and middle-class families. Toddlers were recruited through birth records and community organizations (e.g., museums, children's outreach events, boys and girls clubs) that serve a diverse population.

Stimuli

There were six unique novel "toys" constructed in the laboratory and pilot tested to be interesting and engaging to infants. Each novel toy was a complex object made from multiple and often moveable parts, and ranged in size from 5 to 8 cm in length, 6 to 12 cm in width, and 4 to 6 cm in depth when measured from their gravitational upright (flat bottom of object placed on a surface). These were organized into two sets of three so that each object in the set had a unique uniform color.

Experimental Setup

As shown in Figure 2, parents and toddlers sat across from each other at a small table (61 cm × 91 cm × 64 cm). Parents sat on the floor such that their eyes and heads were at approximately the same distance from the tabletop as those of the toddlers, a posture that parents reported to be natural and comfortable. Both participants wore head-mounted eye trackers (Positive Science, LLC, http://www.positivescience.com/; also see Franchak et al., 2011). The Positive Science eye tracker was designed for use with infants, and was designed to be attached to the head so as to be stable on the head (even in self-locomoting infants and toddlers, see Franchak & Adolph, 2010; Franchak, Kretch, Soska, Babcock, & Adolph, 2010). The tracking system has been widely and successfully used in both infant and adult research (Baschnagel, 2013; Evans, Jacobs, Tarduno, & Pelz, 2012; Franchak et al., 2011; Kretch, Franchak, & Adolph, 2014; Macdonald & Tatler, 2013; Maldarelli, Kahrs, Hunt, & Lockman, 2015). Both parent and infant eye-tracking systems include an infrared camera—mounted on the head and pointed to the right eye of the participant—that records eye images, and a scene camera that captures the events from the participant's perspective. The scene camera's visual field is 108°, providing a broad view but one less than the full visual field —approximately 170° (Smith, Yu, Yoshida, & Fausey, 2015). Each eye-tracking system recorded both the egocentric view video and eye in head position (x and y) in the captured scene at a sampling rate of 30 Hz.

Placing the Head Gear and Eye-Tracker Calibration

Prior to entering the testing room, in the waiting area, the first experimenter desensitized the toddler to touches to the head and hair by lightly touching the hair several times when the attention and interest of the toddler were directed to a toy. Both the parent and the toddler entered the experimental room, and a second experimenter and the parent engaged the toddler with an enticing toy with buttons to push that make animals pop up. The toddler's head gear was placed while the toddler was engaged with the toy. This was done in one movement and care was taken by the experimenter to ensure that the toddler remained engaged with the toy and that the toddler's hands did not go to the head gear. The first experimenter then adjusted the scene camera to ensure that the button being pushed by the toddler was in the center of the scene camera. We have used this procedure in multiple head-camera and head-mounted eye-tracking experiments (Pereira, Smith, & Yu, 2014; Smith, Yu, & Pereira, 2011; Yu & Smith, 2012, 2013; Yu, Smith, Shen, Pereira, & Smith, 2009) with an overall 70% success rate. Detailed information can be found in Appendix A.

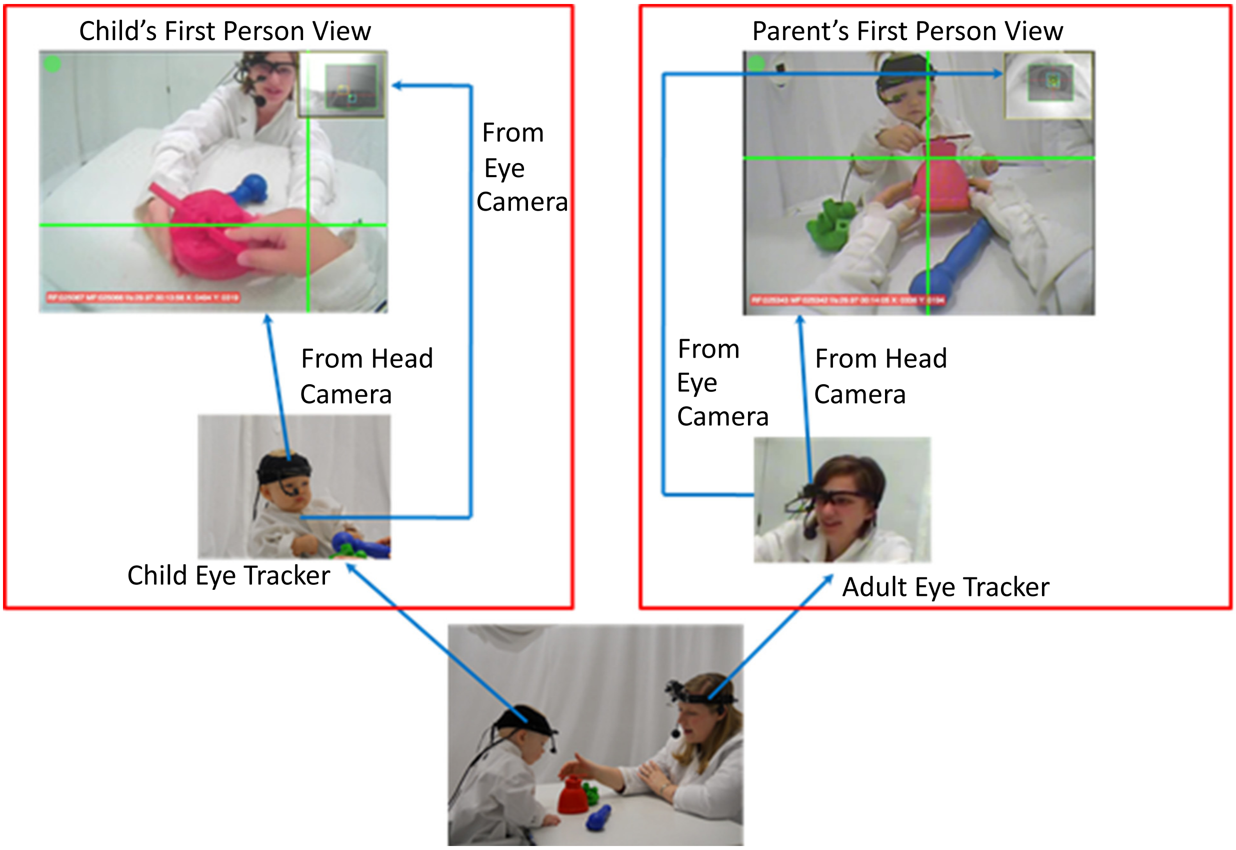

Figure 2. The dual eye-tracking experimental paradigm wherein toddlers and parents played with a set of toys on a tabletop in a free-flowing way. Both participants wore a head-mounted eye tracker that recorded their moment-to-moment gaze direction from their egocentric views. Also shown are three of the laboratory-made toys with their multiple moveable parts and uniform colors. [Color figure can be viewed at wileyonlinelibrary.com]

Instructions and Procedure

Parents were told that the goal of the experiment was to study how parents and toddlers interacted with objects during play, and therefore they were asked to engage their toddlers with the toys and to do so as naturally as possible. Each of the two sets of toys was played twice for 1.5 min, resulting in 6 min of play data from each dyad. Order of sets (ABAB or BABA) was counterbalanced across dyads.

Data Processing

During postprocessing and before coding, the quality of the eye-tracking videos (with eye images superimposed) for each toddler and parent was checked as described earlier to ensure the quality of calibration at the end as well as the beginning of the session. Recalibration would be conducted if necessary.

The eye tracker collected at a rate of 30 frames per second for approximately 360 s (four trials with 1.5 min per trial) of interaction, yielding potentially 10,800 data points per measure for each participant. Not all participants provided eye-tracking data for the entire session, the mean number of good eye-tracking frames was 8,125 (*SD* = 984) for toddlers and 8,356 (*SD* = 825) for adults. Roughly 25% of frames from toddlers that were not codable with respect to regions of interest (ROIs, defined in the next); this was due to 10% eye-tracking failure and the rest due to the toddler's being off task (looking elsewhere than defined ROIs). The main data for analyses were gaze data directed to four ROIs (described next). All results are reported in terms of percentage of the total *interaction time*—as if the number of recorded frames equaled the total number of possible (i.e., 10,800) frames. Therefore, estimates of percentage time on ROIs are underestimates because they include both off-task time and eye-tracking failures.

ROI coding was done by human coders. These coders were highly trained and code these variables for many different experiments and projects. They were naïve to the specific hypotheses and experimental questions of this study. The four ROIs were defined in the head-camera videos: the three toy objects and the partner's face. From gaze ROI coding, each dyad provided two gaze data streams containing the four ROIs as shown in Figure 3a. A second coder independently coded a randomly selected 10% of the frames with the intercoder reliability ranging from 82% to 95% (Cohen's kappa = .81). Detailed information about coding and reliability is provided in Appendix B.

Manual contact with an object (who and which object) was also coded frame by frame from the images captured by the overhead camera and the other two third-person cameras. We developed a custom-coding program that allowed coders to access three views simultaneously to determine which object was manually handled frame by frame. In practice, coders most often relied on the view of the overhead camera, but in case of uncertainty, they would consult with the other two views to make a decision. Similar to gaze ROI coding, each video was coded through three rounds wherein one object was focused in each round and the coder made the yes/no decision that a hand—and whose hand—was in contact. This coding scheme increases the total coding time compared with coding all three objects at once but reduces error. The second coder also independently coded a randomly selected 10% of the frames with the intercoder reliability ranged from 91% to 100% (Cohen's kappa = .94).

Figure 3c shows frame-by-frame measures of four types of hand–eye coordination. For example, at Moment a, both parent's and toddler's eyes were on the object manually handled by toddler; at Moment b, both attended to the object manually handled by parent.

Results

Individual Gaze Patterns and Joint Attention

Figure 3a shows a representative example of the raw gaze data streams for one dyad. Table 1 provides the summary statistics of several measures of infant and parent looks within ROIs for the entire sample. For each type of looking behavior, we report three measures: (a) percentage of total looking time within ROIs, (b) frequency with which these looks occurred (in rate/min), and (c) mean duration of within-ROI looks (in seconds). For all measures, correlations with age of the infant were small and not significant, with one exception (proportion of time infants looked within the ROIs vs. "off task"), suggesting that neither infants' nor parents' looking behaviors change systematically as a function of infant age. However, consistent with previous findings (Smith et al., 2011; Yu & Smith, 2012, 2013), infants and parents differed considerably and reliably on all measures: Infants and parents spent a high proportion of time fixating the ROIs, but parents spent more total time overall than infants (M_{parent} = 82.58%, M_{infant} = 75.76%) and exhibited more attentional switches between objects and faces (M_{parent} = 61.29 switches per minute) than infants (M_{infant} = 25.46). Infants, in contrast, had longer unbroken looks within the same ROIs than did the parents (M_{parent} = 806 ms, M_{infant} = 1,825 ms), showing the "stability" often observed in infant and toddler attention during object play (Kannass, Oakes, & Shaddy, 2006; Yu & Smith, 2013). The different dynamics of infants' and parents' visual attention (e.g., see an example shown in Figure 3a) suggest two different attentional systems with two different rhythms that could be challenging for the coordinating of the two partners' visual attention. Moreover, infants fixated the objects more than their parents (M_{parent} = 64.80%, M_{infant} = 55.33%), whereas parents fixated the faces of their infants much more often than infants looked to their parents' face (M_{parent} = 34.03%, M_{infant} = 11.61%).

These patterns—and perhaps particularly infant looks to parent face—do not appear to depend on the fact the parent head gear in some way altered infant gaze patterns. The results from the additional five infants who interacted with a parent not wearing any head gear were similar to infants in the main study (overall duration: M_{infant} = 1,809 ms; overall frequency: M_{infant} = 29.67, face look duration: M_{infant} = 1,302 ms, face look frequency: M_{infant} = 4.65). Within this sample, and consistent with past findings in which both participants wore head gear (Yu & Smith, 2013) and in which only the infant did (Yoshida & Smith, 2008), infants rarely looked to their parents' face during the play session. This fact precludes infant gaze following as a route to JA.

To find JA episodes in the gaze stream data, we applied the same method used in the previous dual head-mounted eye-tracking study (Yu & Smith, 2013). We first determined—frame by frame—the frames in which parents and infants looks were within the same ROI (on the same object or on each

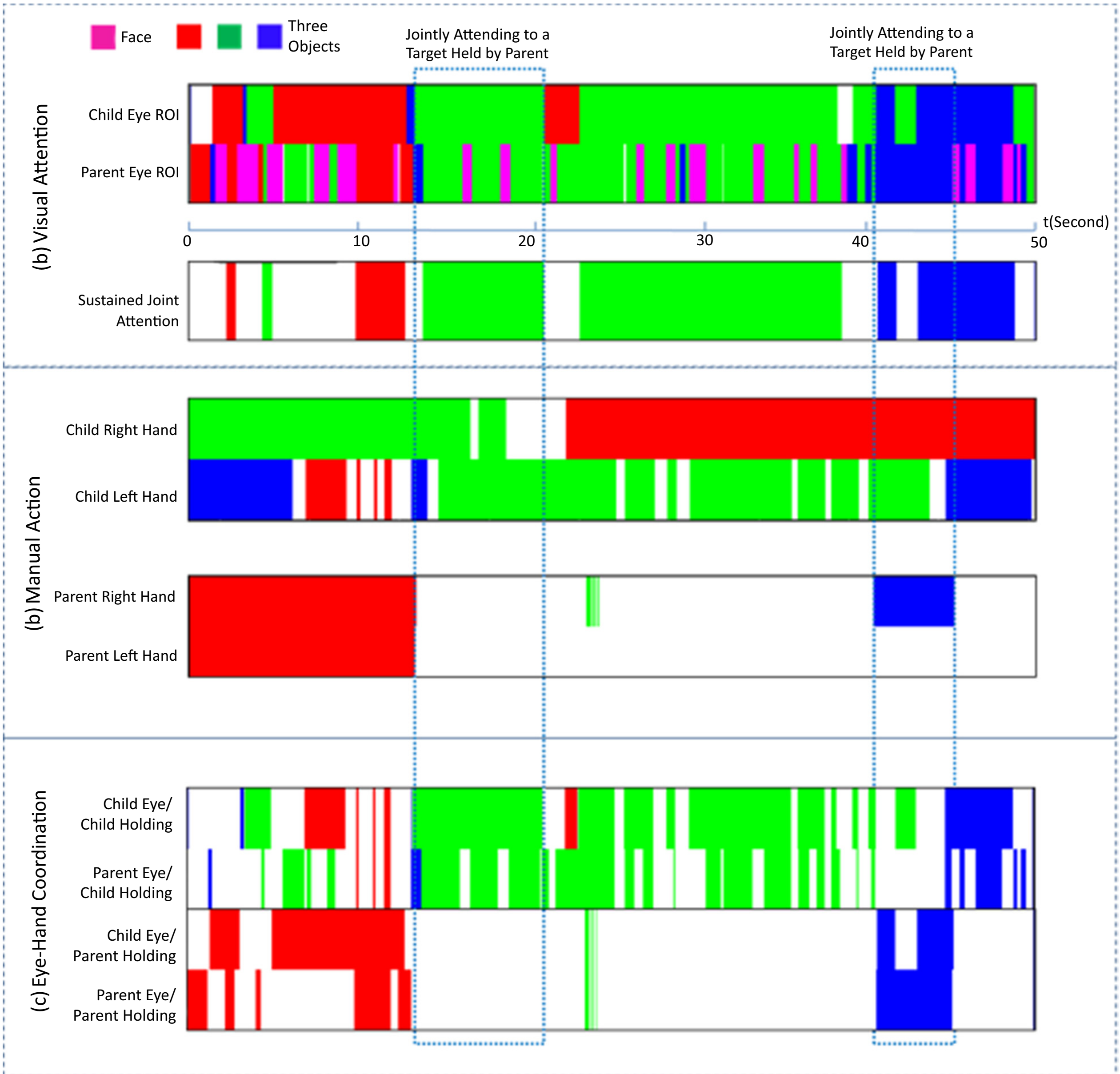

Figure 3. An overview of raw and derived data. (a) Visual attention: two gaze data streams from child and parent include four regions of interest (ROIs)—three toy objects and the partner's face. Sustained visual attention (the third row) is derived based on child's and parent's gaze data. (b) Manual activities on objects from child and parent. (c) Hand–eye coordination: four types of hand–eye coordination within each person and across two partners. At some moments, both child's and parent's eyes are "caught" by what the child was manually handling. At other moments, they jointly attended to objects in parent's hands. [Color figure can be viewed at wileyonlinelibrary.com]

other's face). Meaningful shared attention should last some amount of time longer than a frame (33 ms) but might also include very brief looks elsewhere. Therefore, a JA bout was defined as a continuous alignment of parent and toddler looks within the same ROI that lasted longer than 500 ms and included segments of these looks that were to the same object but separated by brief looks away by one partner of no longer than 300 ms. We have tested slightly different defining windows for JA measures, using 400 and 600 ms for the minimal duration of JA, and 200 and 300 ms for in-between brief looks. Repetition of the analyses with those thresholds yielded the same patterns of results. Examples of the so-defined JA bouts from one dyad's gaze streams are shown in Figure 3a.

Table 1
Parent and Infant Differences in Fixations on the Defined Regions of Interest (ROIs): The Three Objects and Each Other's Face

	Infant M (SD)	Correlation with age	Parent M (SD)	Correlation with age	Infant–parent comparison
Fixations to ROIs					
% of time	75.76 (12.65)	.256*	82.58 (12.23)	.016	$t(50) = 2.77$ $p < .001$, $d = 0.783$
Frequency (rate/min)	25.46 (8.33)	.145	61.29 (13.71)	.032	$t(50) = 17.366$ $p < .001$, $d = 4.910$
Duration (ms)	1,825 (745)	.103	806 (282)	.021	$t(50) = 8.79$ $p < .001$, $d = 2.486$
Looks to faces					
% of time	11.61 (7.14)	−.04	34.03 (13.87)	−.203	$t(50) = 8.038$ $p < .001$, $d = 2.273$
Frequency (rate/min)	4.74 (2.28)	.099	21.52 (7.22)	−.155	$t(50) = 15.805$ $p < .001$, $d = 4.265$
Duration (ms)	1,364 (530)	.085	791 (338)	−.010	$t(50) = 6.514$ $p < .001$, $d = 1.841$
Looks to objects					
% of time	64.8 (13.71)	−.127	55.33 (11.63)	−.235	$t(50) = 5.573$ $p < .001$, $d = 1.576$
Frequency (rate/min)	20.71 (7.96)	.123	39.77 (10.80)	.145	$t(50) = 10.13$ $p < .001$, $d = 2.865$
Duration (ms)	2,093 (885)	.139	826 (243)	.049	$t(50) = 9.86$ $p < .001$, $d = 2.789$

*$p < .05$.

Column 1 of Table 2 summarizes a set of statistics on JA measures across the whole sample: the percentage of overall time in JA within any ROI, the frequency with which JA bouts were formed (in rate/min), and the mean duration of these bouts (in seconds). These same statistics are provided for the two subcomponents of overall JA—mutual gaze and JA to an object. Overall, the results in Table 2 indicate many moments of visual gaze coordination, especially for looks to objects: Parents and toddlers looked at the same object at the same time over 32% of the play session; in contrast, they looked at each other's faces at the same time (mutual gaze) only 5% of the time. The overall time in coordinated attention to objects consisted of multiple bouts of JA, on average, over eight such bouts per minute. As shown in Column 2 of Table 2, these measures were generally not reliably associated with age, with the only statistically reliable correlation being between percentage of overall JA time and age. However, there were substantial dyad differences—JA episodes varied from near 14% to over 63% of the play session.

The main hypothesis motivating this study is that toddlers and parents create JA moments by jointly attending to the objects being manually handled and therefore that dyad differences in hand–eye coordination are critical predictors of dyad differences in the frequency of JA. Consistent with this hypothesis, across dyads, for 82.34% of JA moments on a visual object, the jointly attended object was being manually contacted by at least one partner. For non-JA moments, only 43.67% of the time was one of the partners hands in contact with an object, $t(100) = 29.40$, $p < .001$, $d = 6.07$. Infants held the jointly attended object 45.23% of time ($SD = 5.63\%$); parents held the jointly attended object 37.72% of time ($SD = 5.81\%$), reliably less often than did their infants, $t(100) = 5.85$, $p < .001$, $d = 1.17$.

Low and High JA Groups

We partitioned dyads into those with high and low incidence of JA bouts using a median split of the overall percentage of JA time. Columns 3 and 4 of Table 2 provide the statistics for the two defined groups for the measures of percentage of JA time, frequency of JA bouts and duration of JA bouts. Because the two groups were defined by the overall percentage of time in JA, the expectation is that they would differ on all the components

Table 2
Measures of Low and High Joint Attention (JA) Dyads

	Whole sample	Correlation with age	Low JA	High JA	Low–high comparison
Overall					
% of time	39.24 (12.07)	.295*	28.08 (8.06)	48.41 (4.85)	$t(49) = 10.61$ $p < .001$, $d = 3.03$
Frequency (rate/min)	9.44 (2.19)	.246	7.82 (1.55)	10.76 (1.70)	$t(49) = 6.45$ $p < .001$, $d = 1.84$
Duration (ms)	2.38 (0.53)	.199	2.02 (0.434)	2.68 (0.44)	$t(49) = 5.49$ $p < .001$, $d = 1.56$
Mutual gaze					
% of time	5.74 (4.56)	.093	4.27 (4.05)	4.92 (6.62)	$t(49) = 0.72$, *n*s.
Frequency (rate/min)	1.78 (1.28)	.06	1.37 (0.96)	1.81 (1.43)	$t(49) = 0.93$, *n*s.
Duration (ms)	2.44 (0.68)	−.02	1.54 (0.69)	1.65 (0.73)	$t(49) = 0.78$, *n*s.
JA to object					
% of time	34.72 (4.56)	−.092	23.82 (7.68)	41.76 (6.90)	$t(49) = 8.94$ $p < .001$, $d = 2.55$
Frequency (rate/min)	7.66 (1.65)	.212	6.45 (1.41)	8.65 (1.07)	$t(49) = 5.09$ $p < .001$, $d = 1.45$
Duration (ms)	1.68 (0.71)	.42	1.99 (0.46)	2.82 (0.61)	$t(49) = 5.77$ $p < .001$, $d = 1.65$

$^{*}p < .05$.

contributing to this overall measure. As shown in Column 5 of Table 2, this is generally true with the exception of measures of mutual gaze, a low-frequency behavior in the present study, and one that at least in the context of active toy play may not be linked to the likelihood of JA (see also Yu & Smith, 2013). High and low JA dyads also did not differ in the frequency with which parents looked to infant faces, $M_{\text{high}} = 22.77$, $M_{\text{low}} = 20.23$, $t(49) < 1.00$, nor in the frequency with which infants looked to parent faces, $M_{\text{high}} = 4.56$, $M_{\text{low}} = 4.94$, $t(49) < 1.00$. The high JA infants were younger than the low JA infants, but the difference was not reliable; the mean age of the high JA infants was 18.78 months ($SD = 4.24$), and the mean of the low JA infants was 19.8 months ($SD = 3.91$), $t(49) = 1.61$, $p = .113$. Thus, neither age of the infant nor visual attention to the partner's face seems to be a determining factor of dyad differences in JA.

For "hand following" to play a role in JA, the participants first need to handle the objects, the solid black arrows in the two hand-following pathways in Figures 1a and 1b. Table 3 shows the summary statistics for the percentage play time that infants and parents were in manual contact with an object across all dyads and also for high and low JA dyads. Infants only were handling an object more than a quarter of the time, parents only were handling an object also about a quarter of the time, and the two partners were both handling objects 36% of the time. Only the frequency of infant handling differed between high and low JA groups; infant handling of an object was also reliably (albeit modestly) correlated with age.

Recent studies suggest that handling objects matters to JA because partners look to their own and to their partner's hand actions on objects (Rader & Zukow-Goldring, 2010; Yu & Smith, 2013, 2016). Figure 4 shows the proportion of total time that either the infant's or parent's gaze was fixated on a hand-handled object in both high JA and low JA groups for the three kinds of handling moments: infant (only) handling an object, parent (only) handling an object, and both handling an object.

Figure 4a shows the frequency with which gaze was directed to the infant-handled object at the moments that infants (only) were manually in contact an object. As predicted, low JA infants showed less hand–eye coordination, looking to their own manual actions on objects less frequently than high JA infants. High and low JA parents also differed: High JA parents attended to the object being handled by their infant more than low JA parents. These conclusions were confirmed via a 2 (JA group) × 2 (parent gaze vs. infant gaze) analysis of looking behavior for the cases when the (only) infant handled an object. Besides the two main factors, the interaction between the two would

Table 3
Percentage Total Play Time in Which Hand Was in Contact With an Object for Low and High Joint Attention (JA) Dyads (Standard Deviation [SD] in Parentheses)

	Overall	Correlation with age	Low JA	High JA	Low–high comparison
Infant	29.73 (16.45)	.31* p = .02	24.07 (13.72)	35.62 (17.24)	$t(49) = 2.65$ $p < .01$, $d = 0.75$
Parent	26.17 (15.38)	−.12 p = .16	29.31 (16.14)	22.91 (14.13)	$t(49) = 1.50$, *ns*.
Both	35.96 (16.91)	−.22 p = .12	38.15 (18.68)	33.7 (14.90)	$t(49) = 0.94$, *ns*.
Neither	8.13 (6.56)	.23 p = .10	8.48 (7.74)	7.76 (5.19)	$t(49) = 0.04$, *ns*.

*$p < .05$.

indicate that infants across the two groups might look equally long at the objects they handled, but their parents differed in their attention on the infant-handled objects, or that parents across the two groups might look equally long at the objects handled by infants, but infants differed in their attention on self-handled objects. The results revealed only two main effects—low versus high JA, $F(1, 98) = 12.11$, $p < .001$, $\eta_p^2 = .08$, and parent gaze versus infant gaze, $F(1, 98) = 18.37$, $p < .001$, $\eta_p^2 = .16$. Across both groups, when infants were handling an object, the infant was more likely to be looking at that object than the parent, but high JA infants and high JA parents looked more at the object handled by the infant than did low JA infants and low JA parents. These findings provide support for the main hypotheses from the infant side of manual actions: Infants who were more likely to achieve JA bouts with their parents not only manually act on objects more, but they also looked more to the object when they were handling it. Moreover, they had parents who visually followed their hand actions to objects more frequently than did low JA infants. Put in other words, when the infant was handling a potential target object for JA, the infants in low JA dyads showed less within-self hand–eye coordination and their parents showed less between-self and infant hand–eye coordination.

Figure 4b shows the frequency with which gaze was directed to the parent-handled object (when infants were not manually in contact with object). Here, we see no group differences in gaze directed at the handled object. A 2 (high vs. low JA group) × 2 (parent gaze vs. child gaze) analysis of variance yielded no significant main effects nor interactions, $F_{\text{group}}(1, 98) = 3.56$, $p = .06$; $F_{\text{agent}}(1, 98) = 0.17$, $p = .67$, *ns*; $F_{\text{interaction}}(1, 98) = 0.11$, $p = .73$, *ns*. The lack of differences in hand–eye coordination across the two groups in these cases suggests that dyad differences may lie primarily in the pathway shown in Figure 1a, infant handling an object, than in the pathway shown in Figure 1b, parent handling an object.

Figures 4c and 4d show the findings from the more complicated cases in which the infant and parent were each holding different objects wherein they could attend to either the object handled by the infant or the one handled by the parent. To which object did the partners jointly look? For the objects held by the infant, a 2 (JA group) × 2 (parent gaze vs. infant gaze) analysis of variance indicated a main effect of JA group, $F(1, 98) = 6.87$, $p < .01$, $\eta_p^2 = .08$, and parent gaze versus infant gaze, $F(1, 98) = 15.74$, $p < .001$, $\eta_p^2 = .14$, but no interaction, $F(1, 98) = 0.09$, $p = .75$, *ns*. High JA parents and infants paid more attention to the objects being handled by the infant than did low JA parents and infants. The same analyses with respect to the object handled by the parent revealed only a significant effect of JA group, $F(1, 98) = 10.48$, $p < .005$, $\eta_p^2 = .11$, with parents and infants in the high JA dyads attending more to the objects handled by the parent than did low JA infants and parents. No other effects approached significance. In a context in which there were two potential targets for shared attention, one in the parent's hands and one in the infant's hands, high JA infants and parents managed to find a joint solution more frequently than did low JA infants and parents.

These results provide clear support for the hand-following pathway in parent–infant JA and suggest that the origins of individual differences may be located in infant manual activity: High and low JA

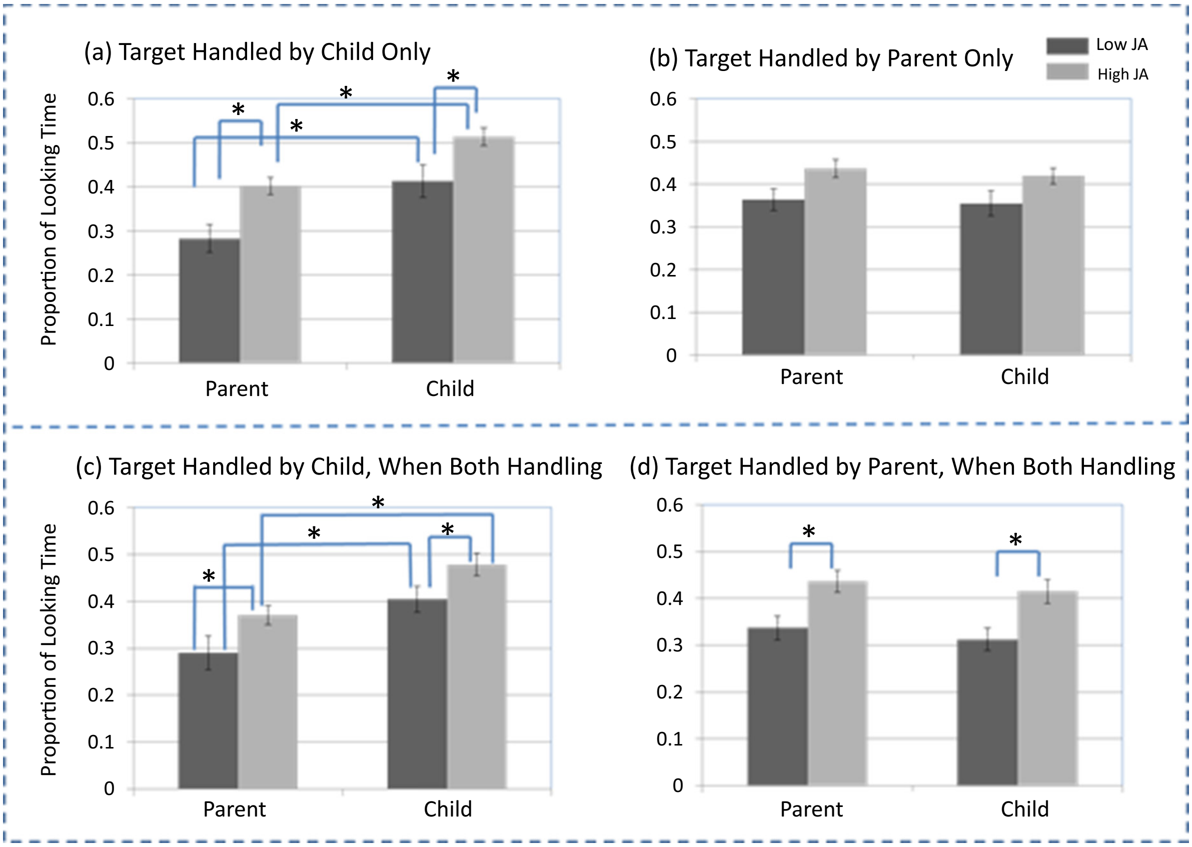

Figure 4. The proportion of total time child and parent visually fixate the target object, when the child is handling the target and the parent is not manually in contact with an object (a), when the parent is handling the target object the child is not handling any object (b), when the child is handling the target and the parent is handling another object (c), and when the parent is handling the target and the child is handling another object (d). [Color figure can be viewed at wileyonlinelibrary.com]

dyads are distinguished by the frequency of infant manual activity on objects, by infant attention to their own object manipulations, by parent attention to infant object manipulations, and by the joint resolution of competing targets when the two partners are holding different objects.

Correlational Analyses

Table 4 shows the bivariate correlations among JA (percentage time) and the three components of the hand-following path: (a) manual activity, (b) within-partner hand–eye coordination, and (c) between-partner hand–eye coordination. All measures except parent handling of objects correlated reliably with JA to an object. These correlations are consistent with the categorical group analyses and provide support that parents and infants use both of the hand-following pathways in Figures 1a and 1b. The new information concerns the dependencies among components of hand-following pathways. Infant object manipulation correlated strongly and negatively with parent object manipulation. Given that parents and infants often acted simultaneously on separate objects, this is not a necessary dependency but suggests that parents are more active when their infants are less active (or vice versa). There was a strong association (.56) between the likelihood that the infant looked to their own object manipulations and the likelihood that the parent looked to infant object manipulations, the two hand–eye coordination components critical to the infant-handling object pathway shown in Figure 1a. If infants do not systematically look to their own hand actions on objects, parents may be less likely to follow those hand actions with their own gaze because hand actions are unreliable cues as to the direction of their infant's visual attention. There was not a strong correlation between parent hand–eye coordination when the parent was handling the object and infant gaze to the parent-handled object, the two hand–eye coordination components of the parent-handling object pathway in Figure 1b. Apparently, parent hand–eye coordination is not a

Table 4
Bivariate Correlations Among the Hand, Eye, Object Links in Figures 1a and 1b

	Manual activity		Within hand–eye		Between hand–eye		
	Infant actions	Parent actions	Infant gaze when infant holding	Parent gaze when parent holding	Infant gaze when parent holding	Parent gaze when infant holding	JA
Manual activity							
Infant actions		.63***	.23	−.03	.03	.24	.33*
Parent actions			.01	0	.06	−.16	−.11
Within hand–eye							
Infant gaze when infant holding				.23	.04	.56***	.45***
Parent gaze when parent holding					.21	.37**	.41**
Between hand–eye							
Infant gaze when parent holding						−.02	.34*
Parent gaze when infant holding							.44**

*$p < .05$. **$p < .01$. ***$p < .005$.

factor in infants' visually following of the parent's hand movements. The correlation between the parent's own hand–eye coordination and parent looks to infant object handling was reliable; the more parents paid attention to the self-handled object, the more likely they were to pay attention to the object handled by the infant.

The overall pattern of correlations suggests the hypothesis shown in Figure 5: The causal pathways through which within- and between-partner hand–eye coordination contributes to JA during active play with toys may be primarily from infant's own hand–eye coordination to parent attention to infant hands. A confirmatory path analysis was conducted as a hierarchical sequential analysis as recommended by Pedhazur (1997). Parents' gaze to infant hand and JA were the endogenous variables because their variance is hypothesized to be explained by other variables in the mode. A multiple regression was conducted for each endogenous variable in which all variables hypothesized to have direct effects on the endogenous variable were included. The beta weights for these multiple regressions are the path weights in the model. In the confirmatory model, the two between-partner components of hand–eye coordination were treated as independent contributors to parents' attention to objects handled by their infant. Because infant attention to the object that is the target of parent actions and parent attention to the object that is the target of infant actions were uncorrelated in their bivariate correlations, they were treated as independent contributors to JA. The beta weights for the paths in this model are given in Figure 5, and indicate that the strongest predictive path to JA was from infant hand–eye coordination through parent visual attention to the targets of infant manual actions.

Discussion

The traditional pathway to JA to an object is through gaze following as shown in Figure 1c. However, gaze is a spatially imprecise and difficult cue for infants, children, and even adults to read in contexts in which there are multiple, spatially near, and moving visual targets. Because people coordinate their attention to objects in these more complex contexts, there must be other routes than gaze following. The present results provide evidence for a hand-following path to coordinated attention to an object as in Figures 1a and 1b. By hypothesis, these hand-following routes characterize parent–infant everyday interactions are thus possible sources of individual differences in the development of socially coordinated attention. Consistent with this larger idea, the present results show that dyad differences in JA resided principally in components of the hand-following pathway shown in Figure 1a, the path in which infants look at their own object manipulations and parents also look at the object handled by their infant. In the following discussion, we consider the implications of hand following for individual differences in the development of socially coordinated visual attention and how the three pathways in Figure 1 may be developmentally related to each other.

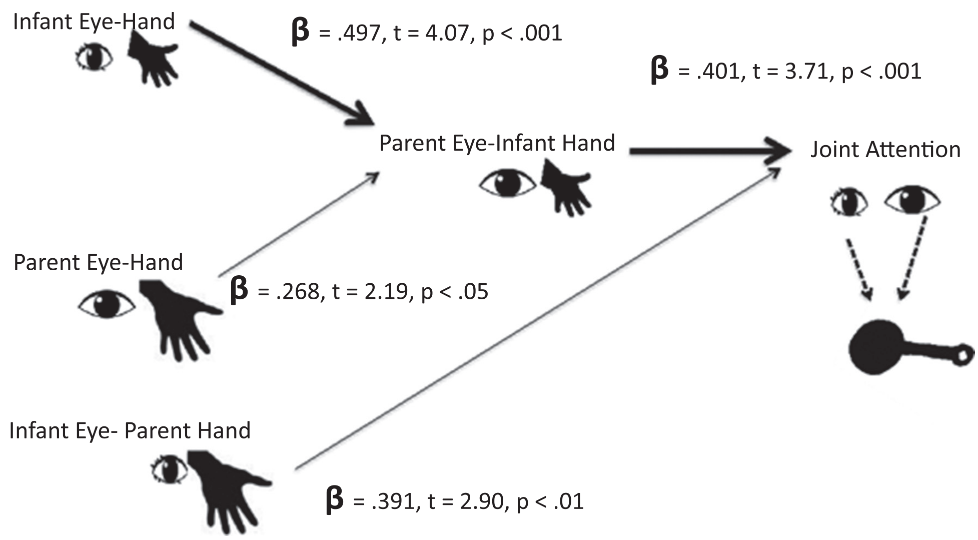

Figure 5. Beta weights and component t values ($df = 50$) for a confirmatory path analysis of the proposed relations among the four measures of hand–eye coordination (within each partner and between the partners) and percentage time of the dyad in joint attention (JA).

The Hand-Following Paths to Joint Attention

Our overarching hypothesis is that the sensorimotor coordination of parents and infants as they jointly interact with objects teaches infants how to rapidly read and respond appropriately to social signals, and how to use their own behavior to send signals to their parent. Because hand actions on objects provide precise and readily perceived cues as to the target of interest, hand actions—and attentional responses to the objects on which hands act—may play a critical role in training more precise gaze following (Ullman et al., 2012). By hypothesis, parents who effectively scaffold JA with their infants during object play provide the kind of coherent context in which the relevant signals and behavioral responses to those signals are discovered. Thus, parent–infant dyads who for whatever reason have difficulty coordinating attention in object play may put the infant at risk for poorer developmental outcomes. If, as the present results imply, weaker hand–eye coordination on the part of the infant limits the parent's ability to effectively scaffold JA, then hand–eye discoordination could cascade into longer term consequences in social development and language learning. These proposals highlight the importance of infant object manipulation to the development of JA and add to the now growing list of domains in which object manipulation appears to be an important component of the developmental pathway (Iverson, 2010; Libertus & Needham, 2011), a list that includes visual object learning (Needham, 2000; Needham, Barrett, & Peterman, 2002; Soska et al., 2010) and understanding others' intentions (Woodward, 2009).

The present proposal about the role of object manipulation and hand–eye coordination in JA is also relevant to the well-documented but not well-understood link between atypical sensorimotor development and atypical social and language development. More specifically, infants at risk for significant delays in social and language development have been reported to show atypical patterns of early sensorimotor development, including delayed and unusual manual interactions with objects (Baranek, 1999; Koterba, Leezenbaum, & Iverson, 2014; Provost, Lopez, & Heimerl, 2007), limited fine motor skills (e.g., Libertus et al., 2014), discoordination of hands and eyes in prospective reaching (Ekberg, Falck-Ytter, Bölte, & Gredebäck, 2015), and perhaps related to the present findings, the exploration of objects with one modality at a time (Kawa & Pisula, 2010). Because social behavior depends on the signals we send through bodily actions (Wolpert, Doya, & Kawato, 2003), atypical sensorimotor behaviors may cause a problematic developmental cascade for optimal social development (Ekberg et al., 2015; Thelen, 2004).

The present results implicate the systematicity with which infants look at their own hand actions on objects as a limiting factor in establishing JA. Why do some infants show less coordination between hands and eyes in this context than others at the same age? Motor development is known to show wide variation in the timing of specific achievements (Adolph & Berger, 2007), and thus the observed differences in the typically developing children in present sample could reflect differences in motor development. If this is correct, hand–eye coordination in social tasks should be related to hand–eye coordination in nonsocial tasks, for example, to performance in insertion tasks that are also known to develop markedly during this period (Smith, 2009). Alternatively, or in addition, the observed individual differences in JA and hand–eye coordination may be linked to the development of sustained attention in nonsocial settings (Richards, 1989; Richards & Casey, 1992), which is also related to the infant's handling of objects (Pereira et al., 2014; Ruff & Capozzoli, 2003; Yu & Smith, 2012). A child who is less distractible and plays longer and more coherently with objects may provide better cues to their social partner. This suggests a possible developmental relation between sustained attention and socially coordinated attention. Moreover, an additional possibility is suggested by two components of the present findings on the parent side: First, parents' own hand–eye coordination predicts the proportion of time in JA with their infant; and second, high JA dyads were better at resolving the competition between the two hand-following pathways in the case when each partner held an object as they jointly attended more to the objects either held by the parent or by the child. Less coordinated parents (through either genetic or experiential history) are likely to have less coordinated infants, and together such parents and infants may have difficulty in providing the sensorimotor cues needed to resolve the competition among the targets. In sum, the present findings by locating one source of dyad differences in JA—in infant object manipulation—offer new and testable hypotheses about how the development of socially coordinated attention is supported by—as well as supports—other developmental achievements.

Multiple Interacting Pathways to Joint Attention

Newborns have been shown to shift their own gaze to match the direction of an eye movement in the context of a still frontal face (Farroni, Massaccesi, Pividori, & Johnson, 2004). In laboratory experiments with well-separated targets and clear social signals, infants as young as 8-month-old follow another person's gaze to an object (Brooks & Meltzoff, 2005). The ability of toddlers to follow the gaze of a partner in laboratory tests strongly predicts developmental outcomes in language learning (Brooks & Meltzoff, 2005; Markus, Mundy, Morales, Delgado, & Yale, 2000; Mundy, Sigman, & Kasari, 1990). Results such as these implicate gaze following as a core ability in the social coordination of attention. However, gaze following may be hard to use in spatially complex contexts in which toddlers rarely look at parent faces and but rather look to their hands (Deák et al., 2014; Yoshida & Smith, 2008; Yu & Smith, 2013). In the present study, both high and low JA infants showed this pattern. What then are the relations between hand following and gaze following?

One possibility is that the social coordination of visual attention begins with gaze following in simple contexts, and meanwhile, in more complex situations, hand following is used. This indirect path to shared gaze may scaffold and train JA enabling infants over time to better follow both hands and eyes in spatially complex spatial situations. By this line of reasoning, the dyad differences observed in the present study should predict infants' developing abilities in laboratory tasks to respond appropriately to an experimenter's signals. Another possibility is that the three pathways shown in Figure 1 (along with potentially other pathways not shown) are not really separable but form a complex system of social coordination in which all the elements codevelop. Face-to-face play in early infancy may set the stage for later hand following (Libertus & Needham, 2011). Hand following may help tune inferred gaze direction (Frischen et al., 2007) through hand-action cues to turn taking and gaze shifting (Nyström, Ljunghammar, Rosander, & von Hofsten, 2011; Pereira, Smith, & Yu, 2008). These ideas suggest perhaps systematic but interdependent developmental changes in the prevalent pathways to coordinated social interactions. This idea of multiple but interrelated paths may help explain the not well-understood shift from so-called dyadic to triadic (or object-centered) social interactions that occurs between 9 and 12 months (Adamson & Bakeman, 1991; De Barbaro, Johnson, & Deák, 2013), a shift that has been linked to the initial *decoupling* of infants' own hands and eyes with respect to objects (De Barbaro et al., 2013). Perhaps the low JA infants in the present study—those who also showed lower hand–eye coordination when engaging objects—are showing the developmentally earlier pattern. That is, the initial decoupling of hands and eyes in those infants is a transition point, enabling the infant to shift attention between objects and social partner, but then followed by hand–eye recoordination in support of organized object play and shared attention to objects within a context of joint action on objects.

In conclusion, using head-mounted eye tracking to record and analyze high-density gaze data during parent–child toy play, we found that JA to an object emerged through the coordination of gaze with manual actions on an object. Hand movements to an object if coordinated with eye movements provide redundant and easy to read information about the object of interest. Dyad differences in JA are associated with dyad differences in hand following, with parents' and infants' manual activities on objects and with within- and between-partner coordination of hands and eyes. Infants who systematically coordinate their own gaze and hand actions on objects are likely to experience more bouts of joint of attention with their parents, a potentially consequential fact if these infant–parent interactions are the training ground for learning the cues that support smooth social interactions.

References

Adamson, L. B., & Bakeman, R. (1991). The development of shared attention during infancy. In L. B. Adamson & R. Bakeman, (Eds.). *Annals of child development* Vol. 8, (pp. 1–41). London, UK: Jessica Kingsley Publishers.

Adolph, K. E., & Berger, S. E. (2006). Motor development. In W. Damon & R. Lerner (Series Eds.) D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol 2: Cognition, perception, and language 6th ed* (pp. 161–213). New York, NY: Wiley.

Argyle, M. (2007). *Social interaction*. Piscataway, NJ: Aldine Transaction.

Aslin, R. N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and Vision Science*, *86*, 561–565. doi:10.1097/OPX.0b013e3181a76e96

Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. *Infancy*, *6*, 155–163. doi:10.1207/s15327078in0602_1

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother–infant and peer-infant interaction. *Child Development*, *55*, 1278–1289. Retrieved from http://www.jstor.org/stable/1129997

Baldwin, D., & Moses, L. (1996). The ontogeny of social information gathering. *Child Development*, *67*, 1915–1939. Retrieved from http://www.jstor.org/stable/1131601

Baranek, G. T. (1999). Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age. *Journal of Autism and Developmental Disorders*, *29*, 213–224. doi:10.1023/A:1023080005650

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: The MIT Press.

Baschnagel, J. S. (2013). Using mobile eye-tracking to assess attention to smoking cues in a naturalized environment. *Addictive Behaviors*, *38*, 2837–2840. doi:10.1016/j.addbeh.2013.08.005

Bayliss, A. P., Murphy, E., Naughtin, C. K., Kritikos, A., Schilbach, L., & Becker, S. I. (2013). "Gaze leading": Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General*, *142*, 76–92. doi:10.1037/a0029286

Bertenthal, B., & Von Hofsten, C. (1998). Eye, head and trunk control: The foundation for manual development. *Neuroscience & Biobehavioral Reviews*, *22*, 515–520.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*, 535–543. doi:10.1111/j.1467-7687.2005.00445.x

Bushnell, E. W., & Boudreau, J. P. (1993). Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. *Child Development*, *64*, 1005–1021. doi:10.1111/j.1467-8624.1993.tb04184.x

Butterworth, G., & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioral Development*, *3*, 253–272.

Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Morre & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 61–83). Hillsdale, NJ: Erlbaum.

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*, 223–248. doi:10.1159/000351313

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, *17*, 270–281. doi:10.1111/desc.12122

Deák, G. O., Walden, T. A., Yale Kaiser, M., & Lewis, A. (2008). Driven from distraction: How infants respond to parents' attempts to elicit and re-direct their attention. *Infant Behavior and Development*, *31*, 34–50. doi:10.1016/j.infbeh.2007.06.004

Eckerman, C. O., & Didow, S. M. (1989). Toddlers' social coordinations: Changing responses to another's invitation to play. *Developmental Psychology*, *25*, 794–804.

Ekberg, T. L., Falck-Ytter, T., Bölte, S., Gredebäck, G., & the EASE Team. (2015). Reduced prospective motor control in 10-month-olds at risk for autism spectrum disorder. *Clinical Psychological Science*, *4*, 129–135. doi:10.1177/2167702615576697

Eppler, M. A. (1995). Development of manipulatory skills and the deployment of attention. *Infant Behavior and Development*, *18*, 391–405.

Evans, K. M., Jacobs, R. A., Tarduno, J. A., & Pelz, J. B. (2012). Collecting and analyzing eye tracking data in outdoor environments. *Journal of Eye Movement Research*, *5*(2), 6. doi:10.16910/jemr.5.2.6

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy*, *5*, 39–60. doi:10.1207/s15327078in0501_2

Flanders, M., Daghestani, L., & Berthoz, A. (1999). Reaching beyond reach. *Experimental Brain Research*, *126*, 19–30.
Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, *50*(24), 2766–2774. doi:10.1016/j.visres.2010.09.024
Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*, 1738–1750. doi:10.1111/j.1467-8624.2011.01670.x
Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*, 694–724. doi:10.1037/0033-2909.133.4.694
Gredeback, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science*, *13*, 839–848. doi:10.1111/j.1467-7687.2009.00945.x
Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*, 188–194. doi:10.1016/j.tics.2005.02.009
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. doi:10.1167/3.1.6
Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*, 229–261. doi:10.1017/S0305000909990432
Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *The Journal of Neuroscience*, *21*, 6917–6932.
Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 10568–10573. doi:10.1073/pnas.1630655100
Johnson, S., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Developmental Science*, *1*, 233–238. doi:10.1111/1467-7687.00036
Kannass, K. N., Oakes, L. M., & Shaddy, D. J. (2006). A longitudinal investigation of the development of attention and distractibility. *Journal of Cognition and Development*, *7*, 381–409. doi:10.1207/s15327647jcd0703_8
Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science*, *17*, 388–395. doi:10.1111/desc.12129
Kawa, R., & Pisula, E. (2010). Locomotor activity, object exploration and space preference in children with autism and Down syndrome. *Acta Neurobiologiae Experimentalis*, *70*, 131–140.
Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*, 317–340. doi:10.1348/000712607X251243
Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science*, *12*, 176–180. doi:10.1111/1467-8721.01255
Knoeferle, P., & Crocker, M. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science: A Multidisciplinary Journal*, *30*, 481–529. doi:10.1207/s15516709cog0000_65
Koterba, E., Leezenbaum, N. B., & Iverson, J. M. (2014). Object exploration at 6 and 9 months in infants with and without risk for autism. *Autism: the International Journal of Research and Practice*, *18*, 97–105. doi:10.1177/1362361312464826
Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*, 1503–1518. doi:10.1111/cdev.12206
Landa, R. J., Gross, A. L., Stuart, E. A., & Faherty, A. (2013). Developmental trajectories in children with and without autism spectrum disorders: The first 3 years. *Child Development*, *84*, 429–442. doi:10.1111/j.1467-8624.2012.01870.x
Langton, S. R. H. (2000). The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology: Section A*, *53*, 825–845. doi:10.1080/713755908
Libertus, K., & Needham, A. (2011). Reaching experience increases face preference in 3-month-old infants. *Developmental Science*, *14*, 1355–1364. doi:10.1111/j.1467-7687.2011.01084.x
Lockman, J. J., & McHale, J. P. (1989). Object manipulation in infancy: Developmental and contextual determinants. In J. J. Lockman & N. L. Hazen (Eds.), *Action in social context: Perspectives on early development* (pp. 129–167). New York, NY: Plenum.
Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, *37*, 1443–1457.
Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4), 6. doi:10.1167/13.4.6
Maldarelli, J. E., Kahrs, B. A., Hunt, S. C., & Lockman, J. J. (2015). Development of early handwriting: Visual-motor control during letter copying. *Developmental Psychology*, *51*, 879–888. doi:10.1037/a0039424
Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development*, *9*, 302–315. doi:10.1111/1467-9507.00127
Marsh, K. L., Richardson, M. J., & Schmidt, R. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*, 320–339. doi:10.1111/j.1756-8765.2009.01022.x

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, *21*, 469–482. doi:10.1016/S0163-6383(98)90020-0

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*, 269–274. doi:10.1111/j.1467-8721.2007.00518.x

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*, 115–128.

Needham, A. (2000). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development*, *1*, 131–156. doi:10.1207/S15327647JCD010201

Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using "sticky mittens" enhances young infants' object exploration skills. *Infant Behavior and Development*, *25*, 279–295. doi:10.1016/S0163-6383(02)00097-8

Nyström, P., Ljunghammar, T., Rosander, K., & von Hofsten, C. (2011). Using mu rhythm desynchronization to measure mirror neuron activity in infants. *Developmental Science*, *14*, 327–335.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction*. Belmont, CA: Wadsworth.

Pereira, A. F., Smith, L. B., & Yu, C. (2008). Social coordination in toddler's word learning: Interacting systems of perception and action. *Connection Science*, *20*, 73–89. doi:10.1080/09540090802091891

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychological Bulletin & Review*, *21*, 1–8. doi:10.3758/s13423-013-0466-4

Provost, B., Lopez, B. R., & Heimerl, S. (2007). A comparison of motor delays in young children: Autism spectrum disorder, developmental delay, and developmental concerns. *Journal of Autism and Developmental Disorders*, *37*, 321–328. doi:10.1007/s10803-006-0170-6

Rader, N. V., & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture*, *2*, 202–221. doi:10.1075/gest.10.2-3.05rad

Richards, J. E. (1989). Development and stability in visual sustained attention in 14, 20, and 26 week old infants. *Psychophysiology*, *26*, 422–430. doi:10.1111/j.1469-8986.1989.tb01944.x

Richards, J. E., & Casey, B. J. (1992). Development of sustained visual attention in the human infant. In B.A. Campbell, H. Hayne, R. Richardson, (Eds.), *Attention and Information Processing in Infants and Adults*. (pp. 30–60). Hillsdale, NJ: Erlbaum.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*, 1045–1060. doi:10.1207/s15516709cog0000_29

Ruff, H. A. (1986). Components of attention during infants' manipulative exploration. *Child Development*, *57*, 105–114. doi:10.1111/j.1467-8624.1986.tb00011.x

Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology*, *39*, 877. doi:10.1037/0012-1649.39.5.877

Ruff, H. A., Capozzoli, M., & Weissberg, R. (1998). Age, individuality, and context as factors in sustained visual attention during the preschool years. *Developmental Psychology*, *34*, 454–464. doi:10.1037/0012-1649.34.3.454

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology*, *26*, 85–93. doi:10.1037/0012-1649.26.1.85

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye–hand coordination during learning of a novel visuomotor task. *The Journal of Neuroscience*, *25*, 8833–8842. doi:10.1523/JNEUROSCI.2658-05.2005

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*, 265–266. doi:10.1038/253265a0

Shockley, K., Santana, M., & Fowler, C. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 326–332. doi:10.1037/0096-1523.29.2.326

Smith, L. B. (2009). From fragments to geometric shape changes in visual object recognition between 18 and 24 months. *Current Directions in Psychological Science*, *18*, 290–294.

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*, 9–17. doi:10.1111/j.1467-7687.2009.00947.x

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*, 407–419. doi:10.1080/15248372.2014.933430

Soska, K., Adolph, K., & Johnson, S. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*, *46*, 129–138. doi:10.1037/a0014618

Thelen, E. (2004). The central role of action in typical and atypical development: A dynamic systems perspective. In IJ. Stockman (Ed.), *Movement and action in learning and development: Clinical implications for pervasive developmental disorders* (pp. 49–73). San Diego, CA: Elsevier Academic Press.

Triesch, J., Teuscher, C., Deák, G., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science*, *9*, 125–147. doi:10.1111/j.1467-7687.2006.00470.x

Ullman, S., Harari, D., & Dorfman, N. (2012). From simple innate biases to complex visual concepts. *Proceedings of the National Academy of Sciences of the United*

States of America, *109*, 18215–18220. doi:10.1073/pnas.1207690109

Vida, M. D., & Maurer, D. (2012). Gradual improvement in fine-grained sensitivity to triadic gaze after 6 years of age. *Journal of Experimental Child Psychology*, *111*, 299–318. doi:10.1016/j.jecp.2011.08.009

Wellman, H. M., Phillips, A. T., Dunphy-Lelii, S., & LaLonde, N. (2004). Infant social attention predicts preschool social cognition. *Developmental Science*, *7*, 283–288. doi:10.1111/j.1467-7687.2004.00347.x

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*, 593–602. doi:10.1098/rstb.2002.1238

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, *18*, 53–57. doi:10.1111/j.1467-8721.2009.01605.x

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*, 229–248. doi:10.1080/15250000802004437

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *16*, 165–180.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244–262. doi:10.1016/j.cognition.2012.06.016

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*, e79659. doi:10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2016). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science*. Advance online publication. doi:10.1111/cogs.12366

Yu, C., Smith, L. B., Shen, H., Pereira, A., & Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE Transactions on Autonomous Mental Development*, *2*, 141–151. doi:10.1109/TAMD.2009.2031513

Appendix A

The first step in the calibration procedure for eye tracking is to adjust the infrared eye camera to capture the child's eye image (the first image in the second row in Figure 4). For head-mounted eye tracking, it is critical to position the eye camera in a good spot to capture a clean image of the eye because the eye tracking software relies on processing eye images to detect the pupil and corneal reflection in order to project where participants look in their first-person view. Next, a target object is systematically moved around on the play surface by an experimenter, the moments that the child attends to the target object in different locations will be used for the off-line eye tracking calibration in later data preprocessing. To collect these pre-experiment calibration points, the experimenter directed the infant's attention toward a high-salience small toy, placing it in a location on the table and ensuring that the infant looked at the placed object. This procedure was repeated 15 times with the toy placed in various locations on the tabletop to ensure a sufficient number of calibration points. A video of the calibration procedure is available in on-line supplementary information for Yu and Smith (2013). After infant calibration was completed, an experimenter played with the infant while the parent's head gear was placed and calibrated. To calibrate the parent's eye tracker, the experimenter asked the parent to look at one of the objects on the table, placed close to the toddler, and then repeated the same procedure to obtain at least 15 calibration points from the parent. Those points were also used for off-line calibration after the experiment was complete. In practice, we found that as many as 9 points are sufficient to achieve high quality of eye tracking. Collecting more data points gave us the flexibility to select different sets of calibration points to feed into the calibration process if the results could not be confirmed by the follow-up validation. For each calibration, following current best practices in research with head-mounted eye tracking (Hayhoe, McKinney, Chajka, & Pelz, 2012; Hayhoe & Rothkopf, 2011), we validated calibration (and re-calibrated using different calibration points if so indicated) by examining moments in which the participant picked up an object at distributed points through out the video and using these moments as both validating and additional calibration points, thereby ensuring the quality of calibration at the end as well as the beginning of the session. During the experiment, one experimenter monitored the eye images on LCD monitors, if the eye tracker was touched or moved causing low quality in eye images, the experimenter entered the room and adjusted the eye tracker, and then the dyad resumed the study. At the end of the experiment, the experimenters repeated the calibration procedure again. In eye tracking calibration, the whole video was divided into two segments and each segment was calibrated separately using different sets of calibration points.

In addition to head-mounted eye tracking, three additional video cameras were used to record the interaction from three different viewpoints that was independent of participants' movements: a bird's-eye camera mounted on the top of the interaction tabletop, a camera pointing to the infant, and a camera pointing to the parent. In total, 7 video streams were recorded in a geovision video capture card (Model 1480B) which automatically synchronized multiple video streams. Synchronization was verified by using a standard camera flash procedure. An experimenter triggered a camera flash both at the beginning and end of each interaction which was captured in one frame in all cameras. Before data processing, coders compared the frame numbers across all the video streams to confirm synchronization. The flash-marked frames were used for resynchronization, if necessary.

Appendix B

To determine gaze that fell within these ROIs, coders watched the first-person view video with a cross-hair indicating gaze direction, frame-by-frame, and annotated when the cross-hair fell on a pixel identified as any part of the four ROIs. Because the experimental room is white and all participants wore white clothing that covers all but faces and hands, and because the three toys in play were three different primary colors that are different from skin tones, it was straightforward for coders to identify the three object and face regions in view. In addition, using the eye tracking software, we rendered eye images via picture-in-picture superimposed at the upper-right corner of a scene frame (see Figure 4), which allowed coders to constantly use them as a reference to verify reliability of cross-hair indicating gaze direction in view. If coders detected from an eye image that the eye tracking software failed to detect the pupil correctly due to image quality or eye blinks, coders disregarded that frame for any ROIs because the cross-hair was incorrect. Thus, we measured infants' and parents' visual attention in terms of gaze directed at any of the three objects or the partner's face. In implementation, coders went through each video four times wherein one of the four ROIs was focused in each round and they needed to make a yes/no decision (whether the cross-hair was on the ROI) based on the overlap of the cross-hair with the ROI. In previous studies with the same setup, we've also developed an image processing algorithm to automatically separate the three objects in play from each other and the background (see Yu, et al. 2009 and Yu & Smith, 2012, for details). We applied automatic object detection to this dataset and calculated object sizes in view. We found that on average, each object took 3.25% of the scene image in the infant's view and 1.82% in the parent's view. Thus, relatively large objects in view with the clean background made ROI coding highly reliable when compared with coding ROIs from more naturalistic and complex visual scenes. Because our research questions were about how well and in what ways infants and parents coordinated their visual attention on the same object, data coding and data analyses were conducted at the object level. That is, as long as a participant was fixated on an object – even with many switches among different parts of the same object – we counted the whole time as one *look* to the object. This is different from the more usual and spatially precise measures of eye fixations used in many studies of visual processing (Wass, Smith, & Johnson, 2013).

Because eye-position does not precisely indicate gaze direction in 3-dimensional space, the ROI coding of overlapping cross-hairs and object-face pixels could be problematic if two ROIs were close in an image but separated in depth in the world (Jovancevic-Misic & Hayhoe, 2009). However, in the context of the present study, given the controlled environment (white room), and the constrained geometry (object play on a table top with both participants sitting on the floor), the central interest of infants in one object at time (Pereira et al., 2014; Smith et al., 2011; Yu & Smith, 2012), this is unlikely to be a significant factor in the observed pattern of results.