HHS Public Access

Author manuscript

Cogn Sci. Author manuscript; available in PMC 2018 February 01.

Published in final edited form as:

Cogn Sci. 2017 February ; 41(Suppl 1): 5–31. doi:10.1111/cogs.12366.

Multiple Sensory-Motor Pathways Lead to Coordinated Visual Attention

Chen Yu and **Linda B. Smith**

Department of Psychological and Brain Sciences, Cognitive Science Program, Indiana University, Bloomington, 1101 East 10thStreet, Bloomington, IN, 47405, USA

Abstract

Joint attention has been extensively studied in the developmental literature because of overwhelming evidence that the ability to socially coordinate visual attention to an object is essential to healthy developmental outcomes, including language learning. The goal of the present study is to understand the complex system of sensory-motor behaviors that may underlie the establishment of joint attention between parents and toddlers. In an experimental task, parents and toddlers played together with multiple toys. We objectively measured joint attention – and the sensory-motor behaviors that underlie it – using a dual head-mounted eye-tracking system and frame-by-frame coding of manual actions. By tracking the momentary visual fixations and hand actions of each participant, we precisely determined just how often they fixated on the same object at the same time, the visual behaviors that preceded joint attention, and manual behaviors that preceded and co-occurred with joint attention. We found that multiple sequential sensory-motor patterns lead to joint attention. In addition, there are developmental changes in this multi-pathway system evidenced as variations in strength among multiple routes. We propose that coordinated visual attention between parents and toddlers is primarily a sensory-motor behavior. Skill in achieving coordinated visual attention in social settings – like skills in other sensory-motor domains – emerges from multiple pathways to the same functional end.

Everyday human collaborative behavior seems so effortless that we often notice it only when it goes awry. One common psychological explanation of how we manage to (typically) work so well together is called “mind-reading” (Baron-Cohen, 1997; Wellman & Liu, 2004). The idea is that we form models of and make inferences about the internal states of others; for example, along the lines of “He is looking at the object and so must want me to look at it and/or pick it up.” However, it is not at all clear that such mental models about the states of others – and inferences from such internal representations – can explain the real-time smooth fluidity of such collaborative behaviors as everyday conversation or joint action. Instead, these behaviors seem to be composed of coordinated adjustments that happen at time scales of fractions of a second and that are highly sensitive to both different task contexts and momentary changing circumstances (Richardson, Dale, & Tomlinson, 2009). Thus, the relevant level of analysis for understanding smooth social interactions may be sensory-motor behaviors. For example, studies of adult conversations implicate a role for oscillations of bodily movement and stillness in the negotiation of speaking turns and

---

Corresponding author: Chen Yu, 1101 East 10th street, Bloomington, 47405, USA. chenyu@indiana.edu; Tel: 812-856-0838.

establishment of common ground (Riley, Richardson, Shockley, & Ramenzoni, 2011; Schmidt & Richardson, 2008; Shockley, Richardson, & Dale, 2009; Shockley, Santana, & Fowler, 2003). Given that coordinated behaviors in adults are supported by sensory-motor processes, the overarching hypothesis of the present study is that young children as a developing system are also likely to rely on external bodily actions to coordinate their behaviors with their parents, and as such will show a key hallmark of sensory-motor systems – the in-the-moment soft-assembly of a solution (Thelen & Smith, 2007).

The literature in early development provides abundant evidence that young children adapt an external bodily solution to cognitive and learning tasks. For example, long before they can sit and manipulate objects, infants actively select visual information by spatially orienting their eyes, heads and bodies (Amso & Johnson, 2006; Canfield & Kirkham, 2001; Johnson, 2010). In addition, the quantity and quality of exploratory behavior by infants determine how well they learn to perceive object completion (Soska, Adolph, & Johnson, 2010). In brief, infants actively use in-the-moment bodily actions to select perceptual information for internal cognitive processes. Moreover, recent studies on early word learning show not only that 18-month-old toddlers use their hand actions to select visual objects during free toy play, but also that parents notice and use infants' actions on objects as behavioral cues to label objects for infants (Pereira, Smith, & Yu, 2014; Yu & Smith, 2012). This sequential pattern, from infant manual handling to parent labeling, suggests an interpersonal coordination that jointly solves the referential uncertainty problem in early word learning -- finding correct word-referent mappings among many co-occurring words and objects. Building on this framework, the goal of the present study is to examine whether momentary bodily actions from social partners and their sensitivities to those behavioral cues generated by others may also play a role in coordinating visual attention between infants and their parents. By hypothesis, this coordination is at the sensory-motor level – at level of hands and eyes. Like many other skilled behaviors at the sensory-motor level, such as reaching and walking (Adolph, Bertenthal, Boker, Goldfield, & Gibson, 1997; Corbetta & Bojczyk, 2002; Thelen et al., 1993), skilled interpersonal coordination should also be the product of a complex system, with multiple degrees of freedom, and therefore rely on multiple solutions to the in-moment tasks of coordinating attention and behavior with another.

Joint attention between infants and parents has been extensively studied in the developmental literature because of overwhelming evidence that the ability to socially coordinate visual attention to an object is essential to many developmental outcomes, including language learning (Baldwin, 1993; Hoff, 2006; Tomasello, 2000; Woodward & Guajardo, 2002). At the *theoretical* level, joint attention has most typically been interpreted in terms of internal models about the mental states of others and inferences from those internal models about the object of attention and interest of one's partner. At the *experimental* level, most prior paradigms have focused on toddlers' ability to "read" the meaning of macro-level behaviors (e.g., eye gaze, head orientation or pointing) in discrete trials with few objects (Meltzoff & Brooks, 2007; Mundy & Gomes, 1998). The adult partner (usually the experimenter) is instructed to focus on the child and on effective teaching, and to provide clear and repeated signals of her attention to the object being named. In this way, the attentional task is simple, and easily described in discrete and categorical terms (the attended object vs. the distractor). Even though the experimental

paradigms offer clean ways to assess infants' social skills and their sensitivity to social cues, these contexts are not at all like the real world in which joint attention is embedded in a stream of free-flowing activity -- in which parents both react to and attempt to control toddlers' behaviors and in which toddlers react to, direct, and sometimes ignore parents as they pursue their own goals. In those naturalistic contexts, socially coordinated shifts in attention are resolved in fractions of a second (Yu & Smith, 2013). It is not at all clear that abstract logic-like inferences about the internal states of others can happen fast enough to explain the exquisite real-time "dance" of social interactions in which effective adjustments within the dyad happen in fractions of seconds.

Accordingly, the present study focuses on understanding the real-time social coordination of attention in naturalistic free-flowing interactions as they unfold in real time in dynamically complex and cluttered contexts. Toward this goal, the overarching hypothesis at the theoretical level is that joint attention as a cognitive outcome may emerge from a complex dynamical system of sensory-motor behaviors (Thelen & Smith, 1994). This complex systems framework requires abandoning the experimental approach of searching for individually necessary and sufficient behavioral cues and causes (Sporns, Chialvo, Kaiser, & Hilgetag, 2004; Thelen & Smith, 1994). Robust complex dynamical systems reliably and consistently find and stabilize an outcome across varied circumstances by multiple pathways to the same end. One of the defining features of a complex system is the dependence on initial conditions or small perturbations through which the system may evolve along multiple routes to produce the same result (Kelso, 1995; Thelen, Kelso, & Fogel, 1987). Here then is the key prediction: If socially coordinated attention is like sensory-motor coordination more generally, there should not be *one* way or *one* critical behavior that is essential to the establishment of joint attention even within a single dyad. This does not mean there are no general principles or no path to scientific understanding. By hypothesis, joint attention is a self-organizing outcome built upon the multimodal coupling of partners' individual sensory-motor behaviors, and if so, there should be a sequence of interwoven and determinable real-time behaviors by parents and toddlers that create and organize different pathways to the state of coordinated attention. If this is correct, then understanding the development of joint attention requires understanding these multiple pathways and their organization in real time.

To test this dynamic systems view of joint attention, we propose to understand the system by studying the dynamic properties of multiple continuous-in-time streams of visual and motor behaviours by the two participants, and by examining the moment-to-moment dynamics of the sensory-motor couplings that bring about joint attention. In the present study, parents and toddlers play together with multiple toys in a free-flowing interaction. Past research indicates that play with multiple toys is attentionally challenging for toddlers as they have difficulties in disengaging attention from one object to focus on a new object and also difficulties in sustaining attention on a target that has been designated by an experimenter as the target of interest (Kannass, Oakes, & Shaddy, 2006; Lansink, Mintz, & Richards, 2000; Ruff, Capozzoli, & Weissberg, 1998; Ruff & Lawson, 1990). Social play between parents and toddlers is an everyday form of collaborative interaction that involves many of the components of other kinds of social collaboration – multiple objects to perceive and act on, shifts in goals, shifts in attention, and actions among the two participants. We focus on

toddlers in two age groups, 12 and 18 months of age, for three reasons. First, children at this age range do not generally engage in collaborative activities unless scaffolded by a mature adult (Bornstein & Tamis-LeMonda, 1989; Tamis-LeMonda, Kuchirko, & Tafuro, 2013). By studying social interaction in which an asymmetry between partners may exist – and where some interactions are likely to be not smooth -- we hope to better understand the components that make for smooth interactions. Second, the developmental literature shows that between 12 and 18 months, there are dramatic changes in both infants' and parents' behaviors in social interactions (Bakeman & Adamson, 1984). This is also a period of time that toddlers acquire new skills and knowledge in various domains, such as motor, cognitive and language development. Third, those developmental changes would allow us to not only document and examine potential developmental differences in how infants and parents establish and maintain joint attention in the two age groups but also to understand the underlying mechanisms of joint attention through comparing fine-grained sensory-motor patterns extracted from the two groups.

This overall idea and experimental paradigm of free play lead to the following three testable hypotheses: First, if the processes that establish visual coordination are sensory-motor behaviors, they should be fast, accomplished through cross-partner real-time adjustments of bodily actions in fractions of a second. Hence, both infants and parents should be able to promptly react and join the social partner to start the next joint attention moment when one person switches her attention to a new target. By doing so, they should be able to spend a significant amount of time in joint attention during free play. Second, if joint attention bouts are built through different sensory-motor pathways as both partners dynamically select, in real time, locally adaptive behaviors, then there should be multiple sequential sensory-motor patterns that lead to joint attention. Third, if coordinated attention is built upon this multi-pathway solution, as infants grow and their behaviors change, there should be developmental changes in this multi-pathway system evidenced as variations in strength among multiple routes. For example, pathways built upon active behaviors from parents may play a more critical role earlier whereas pathways dependent on the child's active engagement with an object may become more important later in development.

The present study was designed to test the above hypotheses by objectively measuring joint attention– and the sensory-motor behaviors that underlie it – using a dual head-mounted eye-tracking system and frame-by-frame coding of manual actions. By tracking the momentary visual fixations and hand actions of each participant, we could precisely determine just how often they fixated on the same object at the same time, the visual behaviors that preceded joint attention, and manual behaviors that preceded and co-occurred with joint attention. The present study focused on hand actions as well as gaze because our previous research using this method suggests that 12 month olds visually follow the parent's hands and hand following plays a contributory role to the establishment of joint attention (Yu & Smith, 2013). With high-density gaze and manual action data, we perform a series of rigorous data analyses to quantity multiple sensory-motor pathways that lead to joint attention, for instance, how their relative strengths change with development, how they may be perturbed, and how they reorganize in different contexts. Our results strongly suggest that understanding joint attention at this sensory-motor level is essential to understanding the origins and nature of the smooth social interactions observed in adults and to understanding

and beneficially influencing social development in atypically developing children. We consider these broader implications in the General Discussion.

Method

Participants

The final sample consisted of 34 parent-toddler dyads. 8 additional toddlers began the study but refused to wear the measuring equipment. The 34 participants (15 male) were distributed across the two age groups, 12-month-olds (M=12.64, SD=2.45) and 18-month-olds (M=19.21, SD=2.16).

Stimuli

There were 6 unique “toys”, organized into two sets of three so that each object in the set had a unique uniform color. Each novel toy was a complex object made from multiple and often moveable parts and were of similar size, on average, 288 cm^3 (see Figure 1).

Experimental setup

Parents and toddlers sat across from each other at a small table (61cm × 91cm × 64cm). 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; also see Franchak, Kretch, Soska, & Adolph, 2011). Each eye-tracking system includes an infrared camera – mounted on the head and pointed to the right eye of the participant – that records eye images, and a scene camera (see in Figure 1) capturing the first-person view from the participant’s perspective. The scene camera’s visual field is 108 degrees, providing a broad view but one less than the full visual field (approximately 170°). Each eye tracking system recorded both the egocentric-view video and gaze direction (x and y) in that view, with a sampling rate of 30 Hz. Another high-resolution camera (recording rate 30 frames per sec) was mounted above the table and provided a bird’s eye view that was independent of participants’ movements.

Procedure

Three experimenters worked together during the experiment. One experimenter played with the toddler while another experimenter placed the eye-tracking gear low on the forehead of the toddler at a moment when the child was engaged with the toy. The third experimenter controlled the experiment computer to ensure data recording. To collect calibration points for eye tracking, the first experimenter then directed the toddler’s attention toward an attractive toy while the second experimenter recorded the attended moment that was used in later eye tracking calibration. This procedure was repeated 15 times with the toy placed in various locations on the tabletop to ensure a sufficient number of calibration points. 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. Parents were told that the goal of the experiments 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 with twice for 1.5 min, resulting in 6 minutes of play data from each dyad. Order of sets (ABAB or BABA) was counterbalanced across dyads.

Data processing

Gaze Data—Four regions-of-interest (ROIs) were defined: the three toy objects and the partner's face. These ROIs were coded manually by coders who watched the first-person view video with a cross-hair indicating gaze direction, frame-by-frame, and annotated when the cross-hairs overlapped any portion of an object or face and if so, which ROI. Thus, each dyad provided two gaze data streams containing four ROIs as shown in Figure 2. The second coder independently coded a randomly selected 10% of the frames with 95% agreement.

Hand action—Manual actions on toy objects (who and which object) from toddlers and parents were coded manually, frame-by-frame, from the images captured by the overhead camera. The second coder also independently coded a randomly selected 10% of the frames with 96% agreement.

Results

The results are organized in three parts. Part 1 reports various measures of overall joint attention, including the key measures relevant to our overarching hypotheses about the speed with which infants follow parents' lead to an object and the speed with which parents follow infants' lead to an object. We show that the coordination is rapid, smooth and consistent. In Part 2, we examine two major classes of sensory-motor bases to joint attention attention – hand following and gaze following. We show that these two major pathways that are differentially used by toddlers and parents when they are the partner that follows the attentional lead of the other. In Part 3, we focus on hand following and report the results from the sequential analyses of sensory-motor patterns that reveal mulitiple in-the-moment and local pathways to the rapid coordination of attention with a social partner.

Part 1 Joint Attention

Part 1.1 Coordinated Visual Attention—Many definitions of joint attention require that two individuals (X and Y) attend to the same object Z, based on X using the attention cues signaled by Y such that X switches attentional focus to join Y to attend to Z (Emery, 2000). Our operational definition of joint attention builds on this definition. We first aligned the gaze streams from each parent and toddler in a dyad yielding a series of frame-by-frame events in which the two partners were (or were not) fixated on the same ROI (same object). Because meaningful shared attention should last some amount of time longer than a frame (33msec) but might also include very brief looks to elsewhere, we defined a joint attention bout as a continuous alignment of parent and toddler fixation to the same ROI that lasted longer than 500 msec and that included brief looks elsewhere if those brief looks were each shorter than 300 msec.

For each joint attention bout, either parent or child needs to be the initiator, fixating on the object ahead of the other person who is thus the follower responding to the behavior of the

initiator to create joint attention. This characterization yields three distinct components to a joint attention bout: the target object, the initiator, and the follower (Bayliss et al., 2013; Emery, 2000). Accordingly, for each episode of sustained joint attention, we determined which partner was first in time to enter our definition of sustained joint attention and categorized the joint attention bout as either child-led or parent-led. As shown in Figure 3(a), the infants in the 12-mo group were the initiator of more than 4 JA bouts per minute (with parents as the follower), which made up 46.51% of all JA episodes, and thus the parents were initiators (with infants as the follower) of more than 4 JA bouts per minute, which created 53.49% of all the JA episodes. These percentages of child-led and parent-led JA episodes did not differ from each other ($t(16)=0.64$, n.s.), indicating that parents and infants were equal contributors to establishing joint attention bouts. Infants in the 18-mo group were initiators on 45.30% of the JA episodes and thus followers on 54.70% of the JA episodes; again the proportions of child-led and parent-led episodes did not differ ($t(16)=0.47$, n.s.). The overall frequency of JA bouts in the two age groups was also not different ($t(32) = 0.58$, n.s.). Across all dyads, both infants and parents were leading and following the other to create joint attention bouts, that is, both were sending behavioral signals to their partner and adjusting their own looking behavior in response to their partner's behavior.

In Figure 3(b) and (c), **proportion of joint attention** measures the overall proportion of time that participants were in a defined bout of joint attention, and **mean duration** captures the average duration of a joint attention bout. For each measure, the results are further divided by two factors, two age groups and the types of JA, either child-led or parent-led. A mixed 2 × 2 ANOVA was performed with age (12-mo vs. 18-mo) as the between-subjects factor and JA type (child-led vs. parent-led) as the within-subjects factor. The results revealed no significant main effect of two age groups ($F(1, 16)= 1.25$, n.s.), of JA types($F(1,32)=0.55$, n.s.), nor a significant age × type interaction ($F(1, 32)=1.32$, n.s.). A similar ANOVA was applied to mean duration measures with the same results. In summary, when playing with their parents, both 12-mo and 18-mo toddlers spent approximately 40% of toy-play time jointly looking at the same targets. They did so by creating more than 4 child-led and 4 parent-led joint attention bouts within a minute (also shown in Figure 3(a)), with each bout lasting about 2 seconds. Given 5 possible attentional states (attending to one of the 4 ROIs plus attending to somewhere else) from each partner in a dyad, a theoretical baseline of joint attention time by chance would be 12% (3/25 -- jointly attending to one of the 3 ROIs divided by 5×5 possible states). Clearly, infants and parents managed to coordinate their visual attention during the interaction.

Part 1.2 Time lag in establishing joint attention—Figure 3(d) shows the time difference (lag, etc.) between when the initiator who first fixated on an object and the follower who subsequently fixated on the same object to form a joint attention bout. The first notable fact is that both parents and infants promptly followed the other's attentional lead. The second notable fact is that toddlers were faster to join the parent, than the parent was to join the toddler. When the child led, it took parents in the 12-mo group 947ms(SD=223ms), and parents in the 18-mo group 867ms (SD=279ms), to join the child's attention to objects. In the child-led (parent following) case, it took children in the 12-mo group 698ms (SD=154ms), and children in the 18-mo group 664ms(SD=187ms), to join the

parent. A two-way mixed ANOVA testing the effects of age group (between subjects) and JA types (child-led vs. parent-led; within subjects) found no main effect of age ($F(1,16)=2.42$, n.s.), and no interaction effect ($F(1,32)=0.06$, $p=0.798$, n.s.), but a main effect of child following or parent following ($F(1,32)=38.41$,$p<0.001$, $\eta_p^2 = 0.52$). The finding that children were faster in following parents' attention than parents were in following toddlers' attention may seem surprising under a "mental model" or "mind reading" approach, because generally adults are more proficient in making inferences from cognitive models than toddlers. However, we found the very same pattern in our previous study (Yu & Smith, 2013) using dual head-mounted eye tracking. In that study, we discovered the source of toddlers' faster speed to be hand following. More specifically, infants in that study rarely looked at the parent's face and therefore they were more likely to follow the parent's attention based on what objects parents were handling which, because the spatial precision of hands is greater than the spatial precision of gaze direction, and hand following is a faster way to discern the partner's object of interest than is gaze following. As we show next, in contrast to their infants, parents often took the slower approach of looking to the toddlers' eyes when joining their child's attention to an object.

Part 2 Gaze following and hand following

Gaze following and hand following may be distinct routes to joint attention that require different sequences of behaviors by the follower. Gaze following has at least three steps: 1) looking at the initiator's face; 2) computing the gaze direction; and 3) switching 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, there is no need to compute and infer the target object; instead, the in-hand object may be directly perceived with minimal uncertainty; further, because there is no need to go back and forth between the initiator's face and the target object, there is just one required attentional shift – from where one was originally looking to the handled object. In contrast, gaze following would seem to require at least two: (1) from the previous attended object to the partner's face, and (2) from the face to the jointly attended object. These considerations lead to two testable hypotheses about how these two pathways to joint attention may be distinguished: 1) In JA bouts created through hand following, the follower – both children in parent-led JA and parents in child-led JA -- should be faster to join the initiator compared with JA bouts through gaze following; 2) Children in parent-led JA bouts are faster to follow the parent because those instances are more likely to be created through hand following than through gaze following, while child-led JA bouts may be most often created through gaze following (by the parent) than through hand following.

To test the above hypotheses, we first need to categorize individual JA bouts as gaze following or hand following. To do so, we measured whether a face look was generated within the temporal window between the initiator's look and the follower's look. We operationally define a JA bout as due to gaze following if there was a face look by the follower before the follower joined the initiator. This is a "liberal" way to count gaze following since looks to the face need not be seeking gaze information but serve some other social functions (Argyle, 1988). On the other hand, *without face looks* from the follower, the

JA bout can *not* be due to gaze following. In those cases, the JA bout was considered to be due to hand following if the initiator of JA manually handled a target object for more than 50% of time within the initiator's looking onset and the follower's onset, thus providing a clear manual cue to the attended object, and if there was no look to the face. With such definitions of gaze and hand following, we let "face looks" trump hand activity as a cue and so errors in classification are most likely to over-estimate the role of gaze following. Based on these definitions, we found that in only a small number of parent-led JA bouts did the toddlers join the parent-attended object by following eye gaze. Toddlers in the 12-mo group looked to their parent's face prior to joining them in attending to an object on only 9.31% of the JA bouts and children in the 18-mo group did so only on 10.54%, ($t(32) = 0.903$, n.s.). When toddlers followed the attentional lead of their parent, they did so primarily when the object was being handled by parents and thus by following the hand, 43.36% for the 12 month olds and 38.15% for the 18 month olds. This leaves a proportion (roughly 30%) of parent-led bouts unaccounted for, a point to which we will return when we consider finer-grained analyses of the pathways in time. In brief, when toddlers followed parent attention, they rarely did so by gaze following but instead typically followed parent hands to the object. In contrast, when parents followed their children's attentional lead, gaze following by the parent was often involved. In child-led JA bouts, parents looked to the child's face prior to joining more than half the time ($M_{12\text{-mo}}$=70.34%, $M_{18\text{-mo}}$=69.13%). Most of the remaining child-led parent-follow instances fit the definition for hand following, 21.68% for the 12-month-olds and 22.75% for the 18-month-olds.[1] To summarize, both child-led and parent-led JA bouts may be created through the gaze following and hand following pathways. However, the frequency of these pathways differ as a function of who is leading and who is following, with hand following being much more likely by the toddler than the parent overall.

Hand-following – looking at the object being handled – is spatially and attentionally simpler than gaze following. Therefore, followers should be faster to attend to the shared target through this route. Accordingly, we calculated the lag between the initiator's fixation on the object and the follower's fixation to the same object for parent- versus child-led, and as a function of whether the bout had been identified as due to gaze following or hand following. When the child was the initiator, it took parents in the 12-mo group 634ms, and parents in the 18-mo group 618ms, to follow through hand following but 1250ms for the 12-mo group and 1367ms for the 18-mo group to do so through gaze following. When the parent led and the child used gaze following to join the parent's attentional focus, it took children in the 12-month group 1226ms, and children in the 18-mo group 1282ms, to join the parent. But when they followed the hand (the more frequent route for toddlers), it took 12-mo olds 654ms and 18-mo olds 609ms to join the initiator. A direct comparison using a 2×2 mixed ANOVA between hand following and gaze following shows a faster response through the hand following pathway ($F(1,32)=54.97$, $p<0.001$, $\eta_p^2 = 0.63$). This difference explains the lag difference between child-led and parent-led JA bouts as parent-led JA bouts include much more instances through hand following than child-led JA bouts. It is not that children are

[1]In child-led case, there were few instances (<10%) that were not categorized as neither hand following nor gaze following based on our definition. Those instances are likely to happen either by chance or through other behavioral cues not considered in the present study, such as vocal cues.

faster than parents, nor that parents are faster gaze followers than the children, but children use the faster hand-following route more often.

In summary, in both child-led and parent-led JA bouts, gaze following and hand following are used as pathways to create joint attention. Hand following seems to play a more important role than gaze following when the child is the follower. When the child leads, parents use both hand following and gaze following to join their toddler in attending to an object. Finally, hand following is a faster solution to follow the initiator's attention than gaze following for both toddlers and parents.

Part 3. Sequential analysis of sensory-motor pathways

Social interactions are continuous streams of behavior and thus joint attention does not begin when the first partner, the initiator, first looks at an object. The initiator's look to the object also emerges from the just preceding behavior of both partners. In the analyses in Parts 1 and 2, we identified joint attention bouts as beginning with both partners fixating on the same object, and then we looked back in time to examine which partner was there first (the initiator), and what was happening in the interval between the initiator's look and when the other partner joined in to the object. But to understand how joint attention emerges in a stream of back-and-forth social behavior, we need go back at least one step further, to the period just before the initiator looks at the object that will become the focus of joint attention. As illustrated in Figure 4, there are now three sequential periods: **Pre-first look** -- The 1 sec period prior to the initiators look to the object; **Before JA** – the lag period between the initiator's look to the object and the follower's joining to form a JA bout; and **During JA** – the period in which the two partners' visual attention is coordinated to the same object. These three intervals are thus defined by the *looking* behavior of the participants. To understand how looking relates to the ongoing manual actions of the social partners, we examined the object handling behaviors of the two participants for each of the three periods: was either partner handling the to-be-target object before the first look, during the lag (after the first look but before JA), or during JA? Object handling is a dynamic event. Accordingly, from our frame-by-frame coding, we categorized one partner as handling the object if her hand was in contact with the object for more than 50% of the frames during the segment. We defined the target object as the object that will become the focus of joint attention. More precisely, for each of the three temporal segments – Pre-first look, Before JA, During JA — we defined three states: 1) child handling: if the child's hand was in contact with the target object over 50% of time within the window; 2) parent handling: if the parent's hand was in contact with object over 50% of time in the window; and 3) no handling: if neither partner's hand was in contact with the object over 50% of time. There is a possible fourth state in which both participants' hands were in contact with the same target object more than 50% of time; this rarely happened (less than 2% of JA instances) and thus is not included in the following analyses.

Hand sequences—Given the three hand activity states at each of three temporal windows, there are, in total, 27 (3×3×3) possible sequential patterns of hand activity alone. For example, one sequence (labeled as pathway (a) in Figure 4) starts with child handling, and the child continues handling the target object through the lag and through the joint

attention segment. Another sequence (pathway (b) in Figure 4) starts with no one handling the target object but then the object is handled by child through the lag and also during the period of shared attention. Given the 27 possible hand-activity sequences, the first set of questions concerns how many of them characterized the interaction, whether individual dyads primarily exhibited many or only a few, and whether different hand activity sequences were differentially associated with child-led and parent-led JA episodes. Accordingly, we calculated the number of pathways that each dyad used to create and maintain joint attention episodes, and found that there were indeed multiple sequences of handling activities as shown in Figure 5(a). Further, for both age groups, hand activity sequences in child-led JA bouts were more variable than those in parent-led JA bouts. A 2×2 ANOVA indicated a main effect of JA types (parent-led or child-led) ($F(1,32)=65.23$, $p<0.001$, $\eta_p^2 = 0.52$). Neither age group ($F(1,16) = 3.87$, $p=0.18$, n.s.) nor the interaction ($F(1,32) = 2.14$, $p=0.08$, n.s.) was reliable. If we collapse across child-led and parent-led episodes, within a single session of 6-minute toy play, every dyad exhibited more than 15 hand activity sequences on the route to joint attention. Moreover, dyads did not appear to individually rely on just a few sequences. As shown in Figure 5(b), the most frequent hand activity sequence accounted for – at best – less than 40% of the joint attention episodes for a dyad.

We also calculated the entropy of the pathway distribution for each dyad as a way to quantify to what degree these pathways were evenly used. Entropy is a measure of uncertainty given a distribution. In the present context, entropy can be viewed as measuring, given that a joint attention episode is achieved, how uncertain/certain a particular pathway is used among several possibilities. A higher entropy value means that multiple pathways are more evenly used and a lower value means one or a few pathways are used more frequently than others. Figure 6 shows the results with several baseline entropy measures. Given a fixed number of pathways, if a dyad uses all of the possible pathways equally frequently, that would maximize the utility of those pathways and therefore also have a highest degree of uncertainty in terms of which pathway may be used case by case. Hence, the entropy of an even distribution serves as the ceiling. More specially, the baseline of 2 even pathways is calculated by assuming 2 pathways in total with each being used 50% of time. Similarly, the entropy of 3 even pathways is calculated for 3 pathways with 33% for each. Since the average number of pathways is close to 10 in child-led cases and 7 in parent-led cases, we also calculated the entropies of those two cases. The results in Figures 5 and 6 clearly show that dyads employed multiple pathways. In particular, a comparison between the entropies from our data with the ceiling values derived from an even distribution with the same total number indicates that dyads use many -- if not all – pathways frequently.

In our view, these findings about multiple hand sequences are fundamentally important on several grounds. First, they make clear the diversity of real-time behaviors that can lead to the same functional outcome. Social competence requires being able to coordinate attention in real time across many different circumstances, and these patterns show that toddlers and parents successfully negotiate joint attention during toy play through multiple means. Second, as the evidence in Part 1 made clear (see also Yu & Smith, 2013), hand activity and visual following of that hand activity is the predominant route through which toddlers follow their parent’s visual attention. Thus, these varied patterns of hand sequences may be critical

to understanding the development of socially coordinated visual attention, and perhaps in particular, how toddlers become adept at "reading" the behavioral cues of others.

Accordingly, we next examined how the frequency of hand sequences differed across the two age groups. We visualized hand activity data using a specific type of flow diagram called a Sankey diagram (Tufte & Graves-Morris, 1983) as shown in Figure 7. A *sankey diagram* is a visualization used to show many-to-many mappings or multiple paths, in which the width of the "rivers" and nodes is proportionally to the flow quantity. The sum of the incoming "rivers" for each node is equal to its outgoing "rivers". In the present context, the width of each "river" shows the probability that a particular hand pathway is employed in the end of a joint attention bout. Each pathway goes through three temporal stages (nodes in Figure 8), from the first pre-look segment, to the before-JA segment, to the JA segment. For example, in the case of child-led JA for the 12 months olds, the top bar shows that when parents were holding the object during the pre-look stage, they continued to do so. Many bouts also began with child holding throughout all segments. Much more rarely, during any of these segments for 12 month olds, did parents holding yield to child manual activities on that object. Overall, the visualization illustrates quantitative information about different types of pathways, their relationships and their dynamic transitions within a system. For each of the four cases (age by who leads), there are multiple ways to end up in the same state, which provides an overall picture of how multiple pathways jointly lead to joint attention. Moreover, the Sankey diagrams suggest two general patterns that we will examine next: 1) there are different flow patterns in child-led and parent-led JA bouts and 2) there are different flow patterns between dyads with 12-month-olds and dyads with 18-month-olds.

Among the 27 possible pathways defined by hand activities across three temporal windows, we computed the probability that each pathway was used in child-led and parent-led JA bouts respectively. Figure 8 shows the top 7 pathways that are used most frequently. Overall, the hand activities of the initiator of the JA bout appear most important. In child-led cases, three pathways involving child's manual activities (nnc, ncc, ccc) have higher probabilities ($M_{nnc+ncc+ccc}$=35.11%) than the three pathways involving parent's hand activities ($M_{nnp+npp+ppp}$=22.11%, $F(1,32)=83.25$, $p<0.001$, $\eta_p^2 = 0.53$). In contrast, the three parent pathways are used more in parent-led cases than the three child hand pathways ($M_{nnp+npp+ppp}$=44.72%, $M_{nnc+ncc+ccc}$=19.03%, $F(1,32)=184.23$, $p<0.001$, $\eta_p^2 = 0.62$). However, for parent-led cases, the parent was more likely to hold the object prior to JA than were the toddlers likely to hold the object prior to child-led JA ($M_{npp+ppp}$=37.45%, $M_{ncc+ccc}$=26.64%, $F(1,32)=35.62$, $p<0.005$, $\eta_p^2 = 0.43$), which fits the importance of the parent's hand activity for toddlers who primarily follow the parent's interests by watching her hand actions on objects. Finally, there was an increased likelihood of object handling with the progression toward JA. Across all cases**,** there were only fewer than 20% of bouts in which the object jointly attended was *not* in either child's or parent's hands during the JA segment. The apparent goal of joint attention in toy play is not to simply look at the same object but for someone to *do* something with the object.

There are two key conclusions from this last set of analyses: First, despite all this variation, and different pathways, parents and toddlers end up in the same state –shared attention to an object – a state that is known to have important consequences for social learning. Second,

both hands and eyes –within and across the social partners –matter in organizing bouts of coordinated visual attention. The patterns shown in the Sankey diagrams also suggest that the likelihoods of specific changing patterns in pathways with development. In both child-led and parent-led bouts, 12-month old toddlers and their parents employed more parent handling pathways. In contrast, 18-month old toddlers employed more child handling pathways in not only child-led but also parent-led JA bouts, a result that suggests the increasing autonomy and "equal partnership" of the developing child in joint play. The key comparisons are shown in Figure 8 and Table 1.

General Discussion

There is no single recipe for effective and smooth social interactions. The moment-to-moment interests, goals, and behaviors of partners in a social interaction are variable and open-ended. Instead, smooth social interactions require rapid reading and adjustment to the behavioral signals of one's partner. These rely on the very properties of skilled sensory-motor co-ordinations evident in such joint actions as dancing or basketball (Sebanz, Bekkering, & Knoblich, 2006; Thelen & Smith, 1994). Because the moment-to-moment contingencies for action are always changing, behaviors with the same goal cannot be executed in the same way if they are to fulfill that same function. Consider, for example, the greatness of Michael Jordan in basketball: fluid, inventive, flexible, and perfectly fit-to-the moment. His greatness in putting the ball in basket was not from doing so in one optimal way, but from the potential to do so in many ways, making real-time decisions and adjustments on which way to score play by play. Similarly, coordinated visual attention between parents and toddlers can be viewed primarily as a sensory-motor behavior. Hence, skill in achieving coordinated visual attention in social settings, just like other sensory-motor skills, emerges from the multiple pathways to the same functional end. Within such system, each individual joint attention bout may be created by a particular pathway, dependent on a particular behavioral cue in a particular context. Therefore, one can create experimental conditions to test each pathway individually to examine what situation is sufficient and necessary to trigger social partners to follow this particular pathway. However, if the robust flexibility that characterizes skilled human social interactions across a variety of social contexts lie in a multi-pathway solution as suggested in the present study, the critical research questions on joint attention should not be just about whether and how each individual pathway works in well-controlled experimental conditions, but on how social partners negotiate moment by moment which pathway they should go through to achieve the same functional end, how different contexts may influence the real-time decision and self-organization of joint attention, and how multiple pathways are utilized together in the same coordination system. Toward this goal, the results in Part 1 show how shared attention between parents and toddlers is common and how both partners rapidly coordinate attention with the other. The results in Parts 2 show that this is a cross-person sensory-motor coordination, dependent on eyes and hands. The results in Part 3 show that this coordination emerges from multiple pathways, and also that the likelihood of different real-time pathways to joint attention changes with development. The following discussion considers the implications of conceptualizing joint attention as a sensory-motor system, and of multiple

pathways for understanding both typical and atypical developmental patterns, and the role of manual activities in parent-infant joint play.

Joint Attention as a Sensory-Motor System

Long before they can sit and manipulate objects, infants actively select visual information by spatially orienting their eyes, heads and bodies (Amso & Johnson, 2006; Canfield & Kirkham, 2001; Johnson, 2010). With the advent of reaching and stable sitting, they use hands to bring objects close to the body and to the eyes for visual exploration (Corbetta, Thelen, & Johnson, 2000; Iverson, 2010; James & Swain, 2011; Rochat & Goubet, 1995). Research in toddlers shows that these activities support sustained and focused visual attention (Ruff & Rothbart, 2001; Smith, Yu, & Pereira, 2011; Yu, Smith, Shen, Pereira, & Smith, 2009), predict word learning in the toy-play task (Yu & Smith, 2012), and predict later language and cognitive development (Ruff, Lawson, Parrinello, & Weissberg, 1990; Tamis-LeMonda & Bornstein, 1990). In these ways, infant visual attention is not just about where the eye goes, but a whole-body matter. In active contexts (when doing and not just watching screens), visual attention is also a whole-body affair in adults (Hayhoe & Ballard, 2005). Heads, hands, shoulders as well as eyes are spatially directed to the target objects in such actions as making sandwiches or building a pattern with blocks. In these goal-directed actions, hands and eyes are tightly coordinated in that they are directed to the same spatial location and show similar spatial precision, latency and velocity profiles in adults (Bekkering, Adam, Kingma, Huson, & Whiting, 1994; Pelz, Hayhoe, & Loeber, 2001; Song & Nakayama, 2006), and as well as in toddlers (Smith, Thelen, Titzer, & McLin, 1999; Von Hofsten, 1982; Yu & Smith, 2013).

The tight coordination of hands and eyes in goal-directed actions is a core fact about the human sensory-motor system and one that creates a pervasive and useable statistical regularity in the visual world. By knowing the direction of gaze of a social partner, one can predict the likely direction of actions; by knowing the direction of hand movements and their contact with objects, one can predict where the eyes of one's social partner are directed. In their social interactions, people clearly make predictions in both directions. However, predicting from hands to eyes has less uncertainty than predicting from eyes to hands because eye-gaze direction is spatially much less precise than hand contact with an object. There is a quite large literature documenting the spatial imprecision of eye-gaze direction. For example, when two-year-olds are asked to determine the target of another's eye gaze, given multiple potential targets and when eye direction is the only cue, they completely fail; 3- and 4-year olds succeed but only if the spatial distances between potential targets are large. The spatial precision of gaze following does not approach adult levels until children are 6 or perhaps even 10 years of age (Doherty, Anderson, & Howieson, 2009; Lee, Eskritt, Symons, & Muir, 1998; Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Vida & Maurer, 2012). The spatial precision of gaze following is especially poor – for adults as well as toddlers – given any head position other than a full-frontal view, or when head and eye are discordant (Corkum & Moore, 1998; Langton, Watt, & Bruce, 2000; Loomis, Kelly, Pusch, Bailenson, & Beall, 2008). Yet in everyday interactions, heads move a lot; faces are most often not in frontal views; and there must often be many potential targets near each other. Perceiving a hand in contact with an object seems unlikely to be influenced by any of these

factors. All this suggests that the hand-following pathway may be *the developmentally early way* into joint attention and the developmentally early way to make predictions about the intentions and interests of one's social partners.

The role of hand actions on objects in organizing joint attention episodes is evident throughout the present results: Toddlers primarily join their parent's visual attention to objects by visually following the parent's hand actions on objects; parents often also look to their toddler's hand actions; and finally, despite the multiple pathways, all JA bouts ended up with both parents and toddlers jointly fixated on a single object being held by one of them. This makes sense because parents do *play* with toys and "play" implies action, not just looking. Action – doing, not just watching – is the core aspect of everyday life. Precisely because the direction of eye gaze in these contexts is just one redundant source of information, people – and perhaps especially very young people -- do not need to "read" eye gaze to stay attuned to their partner's momentary interests.

The role of hands -- and their coordination with eyes -- in coupling the attention of parents and toddlers fits the view of socially coordinated attention (Marsh, Richardson, & Schmidt, 2009; Shockley et al., 2003) as a whole body affair with head turns, body posture, mouth openings, and hand and eye directions, all of which contribute in the moment to the rapid local sensory-motor decisions that keep the two partners eyes – and minds – on the same topic. The present results suggest further that these whole body sensory-motor co-ordinations characterize the developmental origins of socially shared visual attention. Toddlers coordinate attention with their parents – joining and leading and doing so with minimal delay – for substantial portions of time without looking to parents' face nor eyes and do so flexibly in the stream of ongoing behavior. Social partners may not have time to think about it. Thinking, in the usual sense of the word, would seem to have little to do with the behavioral coordination. This does not mean that older preschoolers and adults do not have mental models about the intentions and meanings of others' behaviors, but it may mean that those mental models are "after the fact" thoughts that have little causal efficacy in real-time coordination of behavior with a social partner.

Multiple Behavioral Pathways to Joint Attention

The fact of multiple pathways to joint attention is important in its own right as these multiple routes may be key to the ability to rapidly adjust behaviors on-line to the behavioral cues of many different partners in many different contexts. Multimodal systems, including the human brain, often show this property of a single function that emerges from more than one configuration of component elements (Edelman, 1987; Sporns, 2011). Within the theory of complex systems, these solutions are understood as "softly assembled," as they are locally assembled in the context of the current task, out of a multiple, largely independent components that become inter-dependent in the context. These are solutions that emerge in ways that fit the idiosyncrasies of context, yet satisfy a common function (see Thelen & Smith, 1994). Such systems are often robust – achieving functionality under unusual circumstances and even with the loss of some components. Thus, the early development of multiple behavioral routes to joint attention may be evidence not of the immaturity of early social systems but of its strength. Toddlers who can organize attention and coordinate their

activity with that of a social partner in multiple but effective ways are likely to have more broadly successful social interactions and also to learn more from and about those interactions.

The present results suggest changes of the likelihoods of different pathways between 12- and 18-months, with the role for parent handling of the object decreasing and the role for child handling of the object increasing, as well as differences in the likely major pathway (hand following versus gaze following) in how toddlers and parents follow their partners' attentional leads. In the present study, none of these differences were all-or-none – not between younger and older children, and not between children and parents. Instead, multiple pathways were part of the repertoire for younger and older toddlers and for adults but they changed in likelihood, a pattern that has also been reported in other developmental domains (Adolph & Berger, 2007; Siegler, 1987). One research question for future work is to examine individual differences in the multi-pathway solution. The present results show that each dyad used multiple pathways to achieve joint attention. What we don't know is whether different dyads used the same set of pathways, or alternatively, whether each dyad selected a different subset, with some dyads using fewer or more of those possible pathways. These questions are highly relevant to the source of individual differences in social skills and whether they come from differences in pathway selection and/or differences in the flexibility of these pathways. More generally, examining individual differences at the sensory-motor level has both theoretical importance and applied utilities.

In the context of the present study, a theoretical framework that embraces multiple pathways rather than privileging just one has consequences for understanding developmental changes in joint attention and the relation of joint attention to other developing abilities. Because multiple pathways share components, they may interact developmentally, training and tuning each other. For example, visually following hand actions may support the development of more spatially precise gaze following (Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; Ullman, Harari, & Dorfman, 2012; Yu & Smith, 2013) by providing a clear spatial signal as to the target. In the present task, toddlers rarely looked at the parent's face; but these rare looks to eyes over time coupled with looks to hands could play a role in the development of gaze following pathways. Understanding the relation of gaze following to other pathways is important because experimental tests of toddlers' ability to follow gaze – typically in simple contexts with two spatially separated targets – are strongly predictive of later social and language outcomes (Brooks & Meltzoff, 2005) and are part of the diagnostic battery for identifying autism spectrum disorders (ASD) in young children. Understanding the developmental origins of gaze following and its role in flexible multiple routes to socially coordinated attention is also essential for effective intervention. Because gaze following has been a useful diagnostic behavior and because of the theoretical focus on gaze following as the essential route to effective social interactions, many intervention programs for children and adults with ASD have centered on training to look to faces and to eyes. As several researchers have noted, gaze following abilities may not be sufficient for successful language learning. For example, even children with a family history of autism followed an adult's gaze to the target object but they didn't learn the word associated with that gazed object (Gliga, Elsabbagh, Hudry, Charman, & Johnson, 2012). For another example, the training regimens based on gaze following have been successful in increasing face looks but

may not be successful in creating skilled social behavior (Matson & Konst, 2013; Meindl & Cannella-Malone, 2011; White et al., 2011). However, if the robust flexibility that characterizes skilled human social interactions depends not on a single route, but on the development of a network of overlapping and partially redundant pathways, then interventions focused on one solution may not be ideal. Indeed, the key question for understanding and training skilled social interactions may not be determining the "best pathway" among several options for specific contexts, but instead in fostering sensitivity to multiple cues and the rapid whole body adjustment to those cues that enables multiple pathways to effectively emerge in real time and to work together to provide robust social coordination.

Manual Actions on Objects

A large literature, under the rubric of "active vision" has shown how many problems in vision are different and sometimes simpler when understood in the context of an active moving body (Ballard, 1991; Findlay, 1998). Studies of the development of visual attention indicate strong roles for the body, and particularly object manipulation, in supporting the development of sustained visual attention (Bakeman & Adamson, 1984; Ruff & Capozzoli, 2003) and in visual development (James & Swain, 2011; Needham, Barrett, & Peterman, 2002). Action may be critical because its effectiveness relies on the local here-and-now conditions and so actions can never quite repeat themselves exactly. Instead, as Piaget (1952) put it, they must accommodate. For example, when manually handling objects, infants must adjust their actions to the specific materials and features of the object in the moment (Bourgeois, Khawar, Neal, & Lockman, 2005), adjustments that foster learning not only about the specific object in the here-and-now but generalizable skills in object segregation (Needham, 2000), multi-modal representations (Oakes & Baumgartner, 2012), and visual object recognition (James & Swain, 2011).

For young children and their parents, joint attention is a form of "active vision". The solution to the problem of how children achieve proficient social skills – and what drives development – may best emerge from this perspective. Here, we have shown that hand actions on objects are an important component in the coordination of visual attention between parents and toddlers. Active manual exploration of an object is an easy indicator of sustained attention and sustained interest. We suggest that these *overt* aspects of active vision and visual attention provide information and experiences that support fluid and robust real-time coordination of visual attention with a partner and may also nourish the development of more advanced social cognition. Finally, because real-time actions – including the orienting of one's gaze and body to the object of interest to another – are inherently variable and must be so to be effective, understanding social interactions as a sensory-motor system in addition or rather than as a social-cognitive system may be essential to developing the skills that characterize healthy mature social interactions.

Acknowledgments

We thank Melissa Elston, Steven Elmlinger, Charlotte Wozniak, Melissa Hall, Charlene Tay and Seth Foster for collection of the data, Tian (Linger) Xu, Seth Foster and Thomas Smith for developing data management and processing software. Sumarga (Umay) Suanda, Sven Bambach, and John Franchak for fruitful discussions. This

work was funded by National Institutes of Health Grant R01HD074601 and R21 EY017843, and National Science Foundation Grant BCS 0924248

References

Adolph, KE., Berger, SE. Handbook of child psychology. John Wiley & Sons, Inc; 2007. Motor Development.

Adolph KE, Bertenthal BI, Boker SM, Goldfield EC, Gibson EJ. Learning in the development of infant locomotion. Monographs of the society for research in child development. 1997:i–162.

Amso D, Johnson SP. Learning by selection: visual search and object perception in young infants. Developmental psychology. 2006; 42(6):1236. [PubMed: 17087555]

Argyle, M. Bodily communication. 2. New York, NY: Methuen; 1988.

Bakeman R, Adamson LB. Coordinating Attention to People and Objects in Mother-Infant and Peer-Infant Interaction. Child development. 1984; 55(4):1278–1289. DOI: 10.2307/1129997 [PubMed: 6488956]

Baldwin D. Early referential understanding: Infants' ability to recognize referential acts for what they are. Developmental psychology. 1993; 29(5):832–843.

Ballard DH. Animate vision. Artificial intelligence. 1991; 48(1):57–86.

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

Bayliss AP, Murphy E, Naughtin CK, Kritikos A, Schilbach L, Becker SI. "Gaze leading": Initiating simulated joint attention influences eye movements and choice behavior. Journal of Experimental Psychology: General. 2013; 142(1):76–92. DOI: 10.1037/a0029286 [PubMed: 22800442]

Bekkering H, Adam JJ, Kingma H, Huson A, Whiting H. Reaction time latencies of eye and hand movements in single-and dual-task conditions. Experimental Brain Research. 1994; 97(3):471–476. [PubMed: 8187858]

Bornstein MH, Tamis-LeMonda CS. Maternal responsiveness and cognitive development in children. New Directions for Child and Adolescent Development. 1989; (43):49–61.

Bourgeois KS, Khawar AW, Neal SA, Lockman JJ. Infant manual exploration of objects, surfaces, and their interrelations. Infancy. 2005; 8(3):233–252.

Brooks R, Meltzoff AN. The development of gaze following and its relation to language. Developmental Science. 2005; 8(6):535–543. [PubMed: 16246245]

Canfield RL, Kirkham NZ. Infant cortical development and the prospective control of saccadic eye movements. Infancy. 2001; 2(2):197–211.

Corbetta D, Bojczyk KE. Infants return to two-handed reaching when they are learning to walk. Journal of motor behavior. 2002; 34(1):83–95. [PubMed: 11880252]

Corbetta D, Thelen E, Johnson K. Motor constraints on the development of perception-action matching in infant reaching. Infant behavior and development. 2000; 23(3):351–374.

Corkum V, Moore C. The Origins of joint visual attention in infants. Developmental psychology. 1998; 34(1):28–38. [PubMed: 9471002]

Deák GO, Krasno AM, Triesch J, Lewis J, Sepeta L. Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. Developmental Science. 2014; 17(2):270–281. DOI: 10.1111/desc.12122 [PubMed: 24387193]

Doherty MJ, Anderson JR, Howieson L. The rapid development of explicit gaze judgment ability at 3 years. Journal of experimental child psychology. 2009; 104(3):296–312. [PubMed: 19640550]

Edelman, GM. Neural Darwinism: The theory of neuronal group selection. Basic Books; 1987.

Emery N. The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience & Biobehavioral Reviews. 2000; 24(6):581–604. [PubMed: 10940436]

Findlay J. Active vision: Visual activity in everyday life. Current Biology. 1998; 8(18):R640–R642. [PubMed: 9740792]

Franchak JM, Kretch KS, Soska KC, Adolph KE. Head-Mounted Eye Tracking: A New Method to Describe Infant Looking. Child development. 2011; 82(6):1738–1750. [PubMed: 22023310]

Gliga T, Elsabbagh M, Hudry K, Charman T, Johnson MH. Gaze following, gaze reading, and word learning in children at risk for autism. Child development. 2012; 83(3):926–938. [PubMed: 22462503]

Hayhoe M, Ballard D. Eye movements in natural behavior. Trends in cognitive sciences. 2005; 9(4): 188–194. [PubMed: 15808501]

Hoff E. How social contexts support and shape language development. Developmental Review. 2006; 26(1):55–88.

Iverson JM. Developing language in a developing body: The relationship between motor development and language development. Journal of child language. 2010; 37(02):229–261. [PubMed: 20096145]

James KH, Swain SN. Only self-generated actions create sensori-motor systems in the developing brain. Developmental Science. 2011; 14(4):673–678. DOI: 10.1111/j.1467-7687.2010.01011.x [PubMed: 21676088]

Johnson SP. How infants learn about the visual world. Cognitive Science. 2010; 34(7):1158–1184. [PubMed: 21116440]

Kannass KN, Oakes LM, Shaddy DJ. A longitudinal investigation of the development of attention and distractibility. Journal of Cognition and Development. 2006; 7(3):381–409.

Kelso, J. Dynamic Patterns: The Self Organization of Brain and Behaviour. The MIT Press; 1995.

Langton SRH, Watt RJ, Bruce V. Do the eyes have it? Cues to the direction of social attention. Trends in cognitive sciences. 2000; 4(2):50–59. [PubMed: 10652522]

Lansink JM, Mintz S, Richards JE. The distribution of infant attention during object examination. Developmental Science. 2000; 3(2):163–170.

Lee K, Eskritt M, Symons LA, Muir D. Children's use of triadic eye gaze information for "mind reading". Developmental psychology. 1998; 34(3):525. [PubMed: 9597362]

Leekam S, Baron-Cohen S, Perrett D, Milders M, Brown S. Eye-direction detection: a dissociation between geometric and joint attention skills in autism. British Journal of Developmental Psychology. 1997; 15(1):77–95.

Loomis JM, Kelly JW, Pusch M, Bailenson JN, Beall AC. Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. PERCEPTION. 2008; 37(9):1443–1457. [PubMed: 18986070]

Marsh KL, Richardson MJ, Schmidt R. Social connection through joint action and interpersonal coordination. Topics in Cognitive Science. 2009; 1(2):320–339. [PubMed: 25164936]

Matson JL, Konst MJ. What is the evidence for long term effects of early autism interventions? Research in Autism Spectrum Disorders. 2013; 7(3):475–479.

Meindl JN, Cannella-Malone HI. Initiating and responding to joint attention bids in children with autism: A review of the literature. Research in developmental disabilities. 2011; 32(5):1441–1454. [PubMed: 21450441]

Meltzoff, AN., Brooks, R. Eyes wide shut: The importance of eyes in infant gaze following and understanding other minds. In: Lee, K., Muir, D., editors. Gaze following: Its development and significance. Mahwah, NJ: Erlbaum; 2007.

Mundy P, Gomes A. Individual differences in joint attention skill development in the second year. Infant behavior and development. 1998; 21(3):469–482.

Needham A. Improvements in object exploration skills may facilitate the development of object segregation in early infancy. Journal of Cognition and Development. 2000; 1(2):131–156.

Needham A, Barrett T, Peterman K. 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. 2002; 25(3):279–295.

Oakes, LM., Baumgartner, HA. Manual object exploration and learning about object features in human infants. Paper presented at the Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on; 2012.

Pelz J, Hayhoe M, Loeber R. The coordination of eye, head, and hand movements in a natural task. Experimental Brain Research. 2001; 139(3):266–277. [PubMed: 11545465]

Pereira AF, Smith LB, Yu C. A Bottom-up View of Toddler Word Learning. Psychological Bulletin & Review. 2014:1–8.

Piaget, J., Cook, M., Norton, W. The origins of intelligence in children. Vol. 8. International Universities Press; New York: 1952.

Richardson DC, Dale R, Tomlinson JM. Conversation, gaze coordination, and beliefs about visual context. Cognitive Science. 2009; 33(8):1468–1482. [PubMed: 21585512]

Riley MA, Richardson MJ, Shockley K, Ramenzoni VC. Interpersonal synergies. Frontiers in psychology. 2011; 2

Rochat P, Goubet N. Development of sitting and reaching in 5-to 6-month-old infants. Infant behavior and development. 1995; 18(1):53–68.

Ruff HA, Capozzoli M, Weissberg R. Age, individuality, and context as factors in sustained visual attention during the preschool years. Developmental psychology. 1998; 34(3):454–464. [PubMed: 9597356]

Ruff HA, Capozzoli MC. Development of attention and distractibility in the first 4 years of life. Developmental psychology. 2003; 39(5):877. [PubMed: 12952400]

Ruff HA, Lawson KR. Development of sustained, focused attention in young children during free play. Developmental psychology. 1990; 26(1):85–93.

Ruff HA, Lawson KR, Parrinello R, Weissberg R. Long-term stability of individual differences in sustained attention in the early years. Child development. 1990; 61(1):60–75. [PubMed: 2307047]

Ruff, HA., Rothbart, MK. Attention in early development: Themes and variations. Oxford University Press; USA: 2001.

Schmidt, RC., Richardson, MJ. Coordination: Neural, behavioral and social dynamics. Springer; 2008. Dynamics of interpersonal coordination; p. 281-308.

Sebanz N, Bekkering H, Knoblich G. Joint action: bodies and minds moving together. Trends in cognitive sciences. 2006; 10(2):70–76. [PubMed: 16406326]

Shockley K, Richardson DC, Dale R. Conversation and coordinative structures. Topics in Cognitive Science. 2009; 1(2):305–319. [PubMed: 25164935]

Shockley K, Santana M, Fowler C. Mutual interpersonal postural constraints are involved in cooperative conversation. Journal of Experimental Psychology: Human Perception and Performance. 2003; 29(2):326–332. [PubMed: 12760618]

Siegler RS. The perils of averaging data over strategies: An example from children's addition. Journal of Experimental Psychology: General. 1987; 116(3):250.

Smith LB, Thelen E, Titzer R, McLin D. Knowing in the context of acting: the task dynamics of the A-not-B error. Psychological Review; Psychological Review. 1999; 106(2):235–260. [PubMed: 10378013]

Smith LB, Yu C, Pereira AF. Not your mother's view: The dynamics of toddler visual experience. Developmental Science. 2011; 14(1):9–17. [PubMed: 21159083]

Song JH, Nakayama K. Role of focal attention on latencies and trajectories of visually guided manual pointing. Journal of Vision. 2006; 6(9):11.

Soska K, Adolph K, Johnson S. Systems in Development: Motor Skill Acquisition Facilitates Three-Dimensional Object Completion. Developmental psychology. 2010; 46(1):129–138. [PubMed: 20053012]

Sporns, O. Networks of the Brain. MIT press; 2011.

Sporns O, Chialvo DR, Kaiser M, Hilgetag CC. Organization, development and function of complex brain networks. Trends in cognitive sciences. 2004; 8(9):418–425. [PubMed: 15350243]

Tamis-LeMonda CS, Bornstein MH. Language, play, and attention at one year. Infant behavior and development. 1990; 13(1):85–98.

Tamis-LeMonda CS, Kuchirko Y, Tafuro L. From Action to Interaction: Infant Object Exploration and Mothers' Contingent Responsiveness (June 2013). IEEE Tranactions on Autonomous Mental Development. 2013; 5(3):202–209.

Thelen E, Corbetta D, Kamm K, Spencer JP, Schneider K, Zernicke RF. The transition to reaching: Mapping intention and intrinsic dynamics. Child development. 1993; 64(4):1058–1098. [PubMed: 8404257]

Thelen E, Kelso J, Fogel A. Self-organizing systems and infant motor development. Developmental Review. 1987; 7(1):39–65.

Thelen, E., Smith, LB. A dynamic systems approach to the development of cognition and action. Vol. 10. MIT Press; 1994.

Thelen, E., Smith, LB. Handbook of child psychology. John Wiley & Sons, Inc; 2007. Dynamic Systems Theories.

Tomasello M. The social-pragmatic theory of word learning. Pragmatics. 2000; 10(4):401–413.

Tufte, ER., Graves-Morris, P. The visual display of quantitative information. Vol. 2. Graphics press; Cheshire, CT: 1983.

Ullman S, Harari D, Dorfman N. From simple innate biases to complex visual concepts. Proceedings of the National Academy of Sciences. 2012; 109(44):18215–18220.

Vida MD, Maurer D. Fine-grained sensitivity to vertical differences in triadic gaze is slow to develop. Journal of Vision. 2012; 12(9):634–634.

Von Hofsten C. Eye–hand coordination in the newborn. Developmental psychology. 1982; 18(3):450–461.

Wellman HM, Liu D. Scaling of Theory-of-Mind Tasks. Child development. 2004; 75(2):523–541. [PubMed: 15056204]

White PJ, O'Reilly M, Streusand W, Levine A, Sigafoos J, Lancioni G, … Aguilar J. Best practices for teaching joint attention: A systematic review of the intervention literature. Research in Autism Spectrum Disorders. 2011; 5(4):1283–1295.

Woodward AL, Guajardo JJ. Infants' understanding of the point gesture as an object-directed action. Cognitive Development. 2002; 17(1):1061–1084.

Yu C, Smith LB. Embodied Attention and Word Learning by Toddlers. Cognition. 2012; 125(2):244–262. [PubMed: 22878116]

Yu C, Smith LB. Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. PLOS ONE. 2013; 8(11)

Yu C, Smith LB, Shen H, Pereira A, Smith T. Active Information Selection: Visual Attention Through the Hands. IEEE Transactions on Autonomous Mental Development. 2009; 2:141–151.

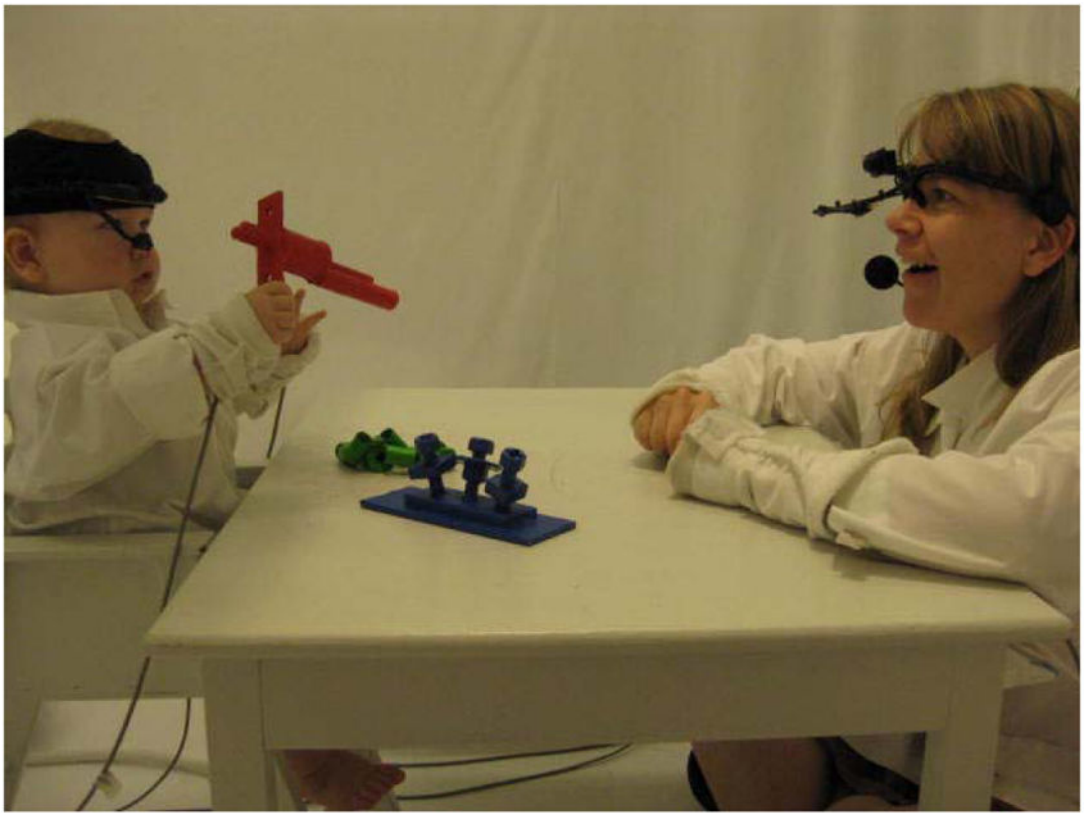

Figure 1.
A dual eye tracking experimental paradigm wherein toddlers and parents played with a set of toys on a tabletop in a free-flowing way. Both participants wore a head-mounted eye tracker that recorded their moment-to-moment gaze direction from their egocentric views.

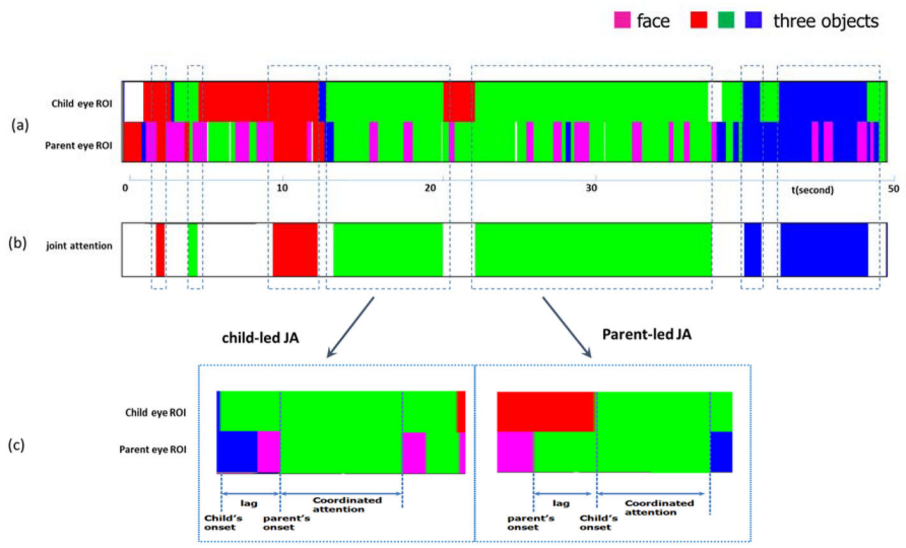

Figure 2.

Joint attention measures are derived with two steps. First, sustained joint attention is calculated by integrating child and parent eye ROI data streams to find the shared moments that children and parents looked at the same object at the same time. Next, each joint attention instance is categorized as child-led or parent-led based on who was gazing at the target object first.

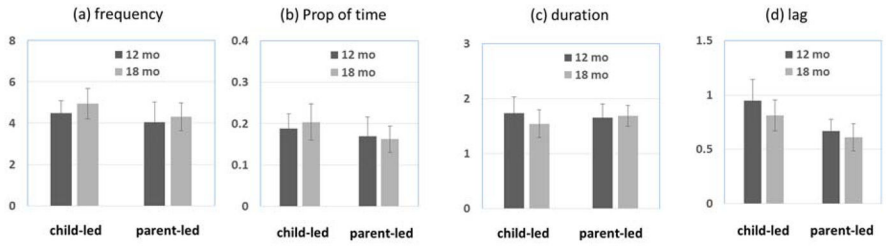

Figure 3.

A comparison of child-led and parent-led JA bouts between 12-month and 18-month olds with four behavioral measures: (a) frequency: how many JA bouts per minute; (b) proportion of time: the overall proportion of time that participants were in a defined bout; (c) duration: how long a bout lasted; and (4) lag: how fast a follower joined an initiator to establish a new JA bout.

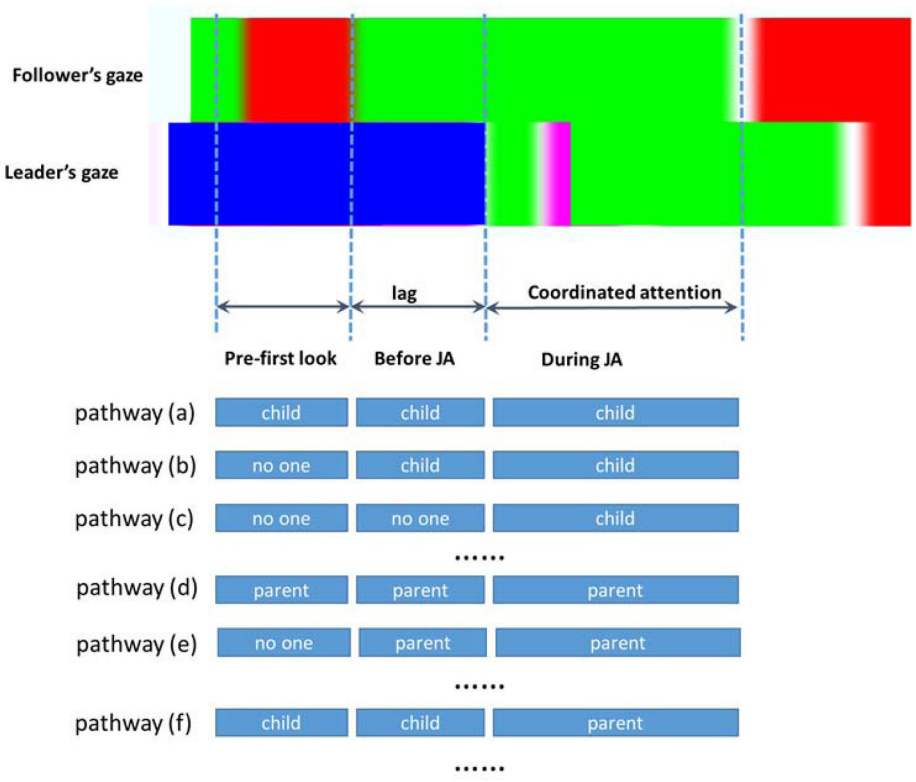

Figure 4.

Three temporal windows are defined for each JA bout: pre-first look, before JA and during JA. Three hand states (either child handling, or parent handling, or no handling) are extracted from each window. Taken together, three hand states in three windows form a hand activity pathway that leads to JA. As listed, there are many such possible hand pathways that are defined by parents' and infants' manual activities on objects.

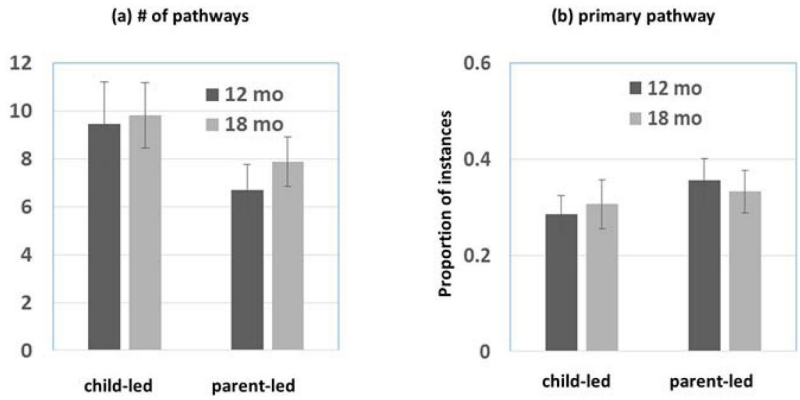

Figure 5.

Multiple hand activity sequences to establish and maintain joint attention bouts. (a) the number of hand pathways revealed in child-led and parent-led JA bouts. (b) the proportion of the most frequent hand activity pathway within individual dyads.

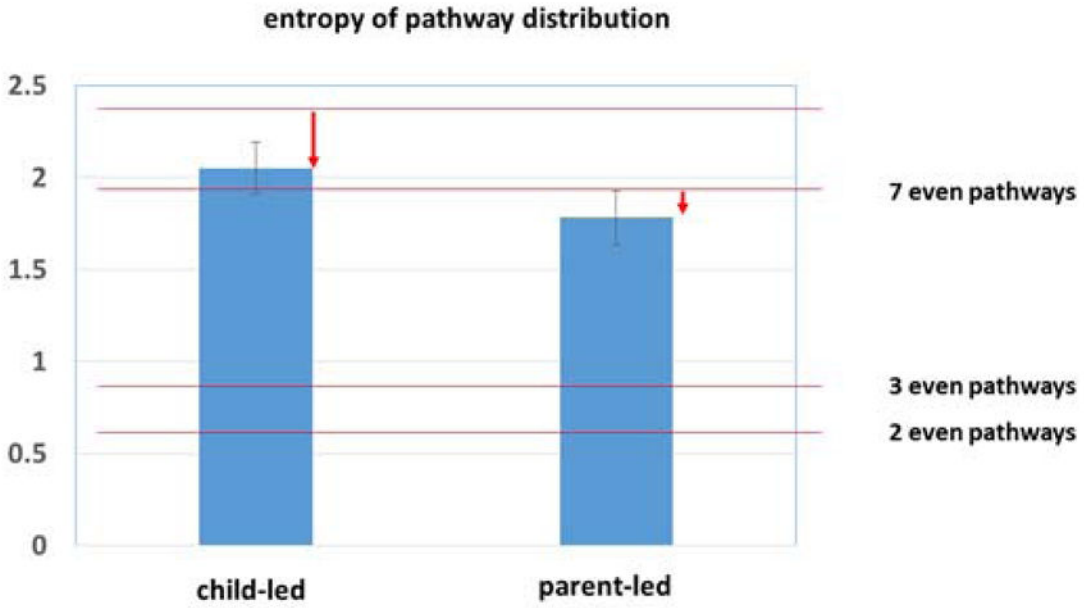

Figure 6.
Entropy of the distribution of multiple hand activity pathways.

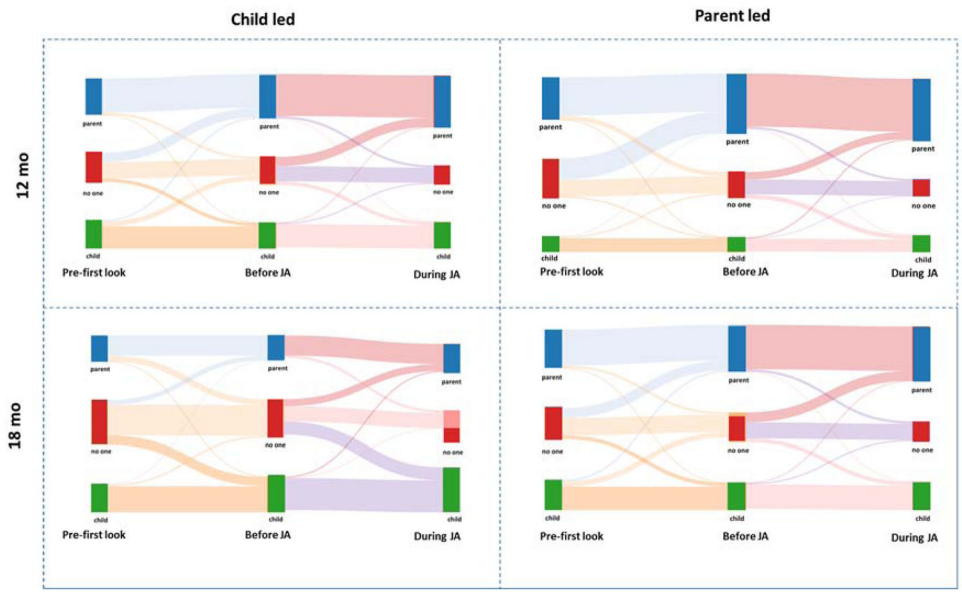

Figure 7.

Four Sankey diagrams are used to illustrate dynamics of hand activities in three temporal windows. In each sub-figure, a Sankey diagram is used to show hand sequences when parents and children jointly went through the three defined windows ending in a JA bout. The dark blue, dark red, and dark green boxes indicate who was handling in the three defined states. The lighter “flows” between these states indicate the transitions. For example, a flow from dark blue to dark blue indicates a transition from parent handling in the earlier window to parent handling in the subsequent window. A flow from a red state to a green state, indicates a transition from no one handling to child handling. For example, starting with parent handling during the pre-look period (the upper left box in each figure), there are three flows that go to the next stage “before JA”: parent handling, no one handling and child handling. These three flows indicate the transitions of hand activity states, from parent handling to parent handling, from parent handling to no one handling, and from parent handling to child handling.

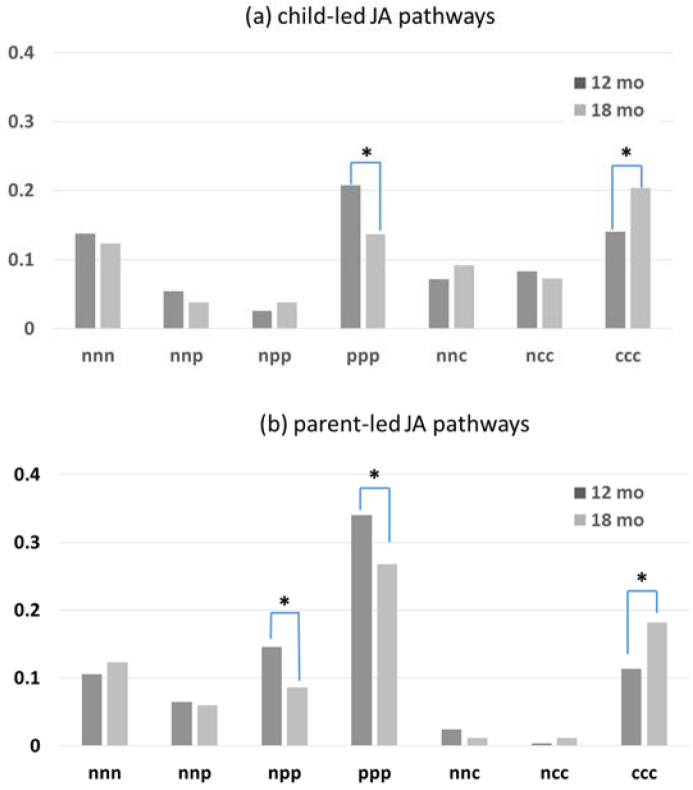

Figure 8.

Probabilities of hand pathways in child-led (a) and parent-led (b) JA bouts. Each pathway is represented by a three-letter string corresponding to the hand state within one of the three temporal windows. There are three possible states: n -- no handling, p – parent handling, c – child handling. For example, “nnn” means no handling across all of the three temporal window; “nnp” means no handling in the first two windows and parent handling during JA; “ppp” means parent handling across the three temporal windows.

Table 1

A comparison of child handling and parent handling activities in two age groups.

Proportion of time	12 mo	18 mo	comparison
ccc			
child-led	0.1408	0.2032	$F(1,32)=65.21$ $p<0.001$;
parent-led	0.1134	0.1822	$\eta_p^2 = 0.45$
ppp			
child-led	0.2069	0.1365	$F(1,32)=64.73$ $p<0.001$
parent-led	0.3401	0.2677	$\eta_p^2 = 0.42$