Multimodal Parent behaviors within Joint Attention support Sustained Attention in infants

Catalina Suarez-Rivera, **Linda B. Smith**, and **Chen Yu**

Indiana University

Abstract

Parents support and scaffold more mature behaviors in their infants. Recent research suggests that parent-infant joint visual attention may scaffold the development of sustained attention by extending the duration of an infant's attention to an object. The open question concerns the parent behaviors that occur within joint-attention episodes and support infant sustained attention to an object. In the study, parent-infant dyads played with objects on a tabletop while their eye-gaze was recorded with head-mounted eye-trackers. Parent hand contact with the objects as well as speech were coded and analyzed to identify the presence of parent touch and talk during bouts of infant visual attention. This study, consistent with prior research, showed that joint attention is associated with longer infant visual attention. The relevant parent behaviors considered, parent talk and touch, not only were highly likely to occur when both the parent and infant visually attended to the same object, but were also associated with infant attention to an object that was longer than infant attention that did not include these parent behaviors. Parent talk was the most potent behavior that coincided with longer infant looks. In sum, joint attention extends infant attention and joint attention involves more than mutual coordination of eye-gaze, it involves multimodal parent behaviors coordinated with the infant's visual attention.

Keywords

Perceivers sometimes glance at objects but other times they visually examine a single object with a sustained look. These longer looks are strongly related to visual learning about the attended target, in infants (Lansik & Richards, 1997; Richards & Casey, 1992; Ruff, 1986), young children (Ruff & Lawson, 1990) and adults (Steinmayr, Ziegler & Träuble, 2010; Wei, Wang & Klausner, 2012). Sustained visual attention develops incrementally from late infancy through early childhood; for example, the average time that a one-year-old infant attends to a single toy during active play is about 3 seconds whereas the average time for a 3-year-old child approaches 9 seconds (Ruff & Lawson, 1990). Early individual differences in sustained attention to objects predict later individual differences in inhibitory control and self-regulation (Kochanska, Murray, & Harlan, 2000; Reck & Hund, 2011; Ruff, 1986), language (Welsh, Nix, Blair, Bierman & Nelson, 2010), and school achievement (Duncan, Dowsett, Claessens, Magnuson, Huston, Klebanov, Pagani, Feinstein, Engel, Brooks-Gunn, Sexton, Duckworth & Japel, 2007; McClelland, Acock & Morrison, 2006; McClelland & Cameron, 2012). The factors that drive the development of sustained attention and their role

in the observed individual differences in sustained attention have not been identified. The ability to sustain attention is sometimes treated as an intrinsic and stable child attribute related to temperament (Colombo, 2001; Posner & Rothbart, 2000). However, developmental theorists have also suggested that parents and social context play a role (e.g., Kopp, 1982; Miller, Ables, King & West, 2009; Parrinello & Ruff, 1988; Sigel, 2002; Spruijt, Dekker, Ziermans, & Swaab, 2018; Vygotsky, 1978). The potential role of parent behaviors in infant sustained attention is the focus of this research.

Yu and Smith (2016) recently showed a direct social effect on the duration of infant attention to a single object during toy play with a parent. Their study used dual head-mounted eye trackers to measure both parent and infant (12 month olds) gaze during play. They defined *infant sustained attention* as an unbroken look to an object that was longer than 3 sec. They defined joint attention objectively, as moments in which the two participants' gaze was directed to the same object, without considering how that joint gaze was achieved and without regard to what one might infer about the knowledge states of the participants, a definition of joint attention distinct from that used in the past (Baron-Cohen & Cross, 1992; Mundy, Sullivan & Mastergeorge, 2009; Tomasello, 1995).. The main result was that toddler *sustained* attention occurred most frequently when the period of infant attention to an object overlapped with a *joint* attention episode. Yu and Smith suggested that parent attention to an infant-attended object may extend the infant's own interest, causing the infant to visually attend to the object longer than the infant would otherwise. Yu and Smith further suggested that this in-the-moment extension of the duration of infant attention by parent shared interest –when repeated day-in and day-out – may tune and train the internal mechanisms that support the development of the self-regulation of attention. They offered the following analogy from how children learn to ride two-wheel bikes to explain the idea: parents often hold onto and balance a two-wheel bike for the young rider at the beginning, so that the learner can get the feel of balancing a bike. After repeated episodes of such parent support, the child becomes able to balance a two-wheeler without help. In the same way, Yu and Smith proposed, parent joint attention to an object may help infants stay attending to that object and through these socially-guided moments of sustained attention events, infants may develop the means to sustain focused attention on their own.

There are many untested predictions that follow from this hypothesis. Here we focus on one open question relevant to the future testing of those hypotheses: What is happening inside joint attention episodes that supports the infant's longer looks to an object? This question is critical because the only parent behavior measured by Yu and Smith was parent gaze in coordination with infant gaze. In principle, the child's extended attention to the object – given joint parent attention to that object– could result from the infant perceiving the direction of parent gaze and inferring parent shared interest in the object from gaze direction alone (e.g., Baron-Cohen & Cross, 1992; Brooks & Meltzoff, 2005). However, Yu and Smith also reported that during play with toys, infants almost never looked to their parents' face and thus couldn't use parent eye-gaze to infer that the parent was also looking at the same object. The finding that toddlers do not often look to the parent's face during joint object play, has been previously reported by multiple laboratories (Bakeman & Adamson, 1984; Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Franchak, Kretch, Soska & Adolph, 2011; Yoshida & Smith, 2008; Yu & Smith, 2013).. Accordingly, the present study was designed to

address two empirical questions: (1) what additional parent behaviors are part of joint parent-infant attention to an object? (2) Are these additional parent behaviors associated with longer visual attention to the object attended by the infant?

When parents are playing with their infants in free-flow interactions, they may do more than just look at objects and at their infants. Parents generate multimodal behaviors and their own attention and interest in the object is potentially signaled through multiple modalities including handling objects, talking about the objects as well as looking at those objects (Bakeman & Adamson, 1984; Tomasello & Farrar, 1986; Yu & Smith, 2012; Yu & Smith, 2013). Several researchers have specifically observed that parent talk increases infant visual attention to objects (Baldwin and Markman, 1989; Belsky, Goode and Most, 1980; see also Parrinello & Ruff, 1988). Other evidence suggests that parent hand actions also play a role in organizing infant visual attention to objects (Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Yu & Smith, 2013, 2017). Accordingly, we tested the hypothesis that multimodal parent behaviors, such as parent talk and parent handling of objects, are often components of a joint attention episode that are associated with longer lasting visual attention to the object by the infant. Because parent looks, utterances and manual activities during toy play are real-time behaviors happening at the time scales of seconds and fractions of seconds, our analyses also focus on the finer temporal details of parent behaviors and their real-time effects on infant gaze.

Method

Participants

The participants were 40 infants with a mean age of 13.97 months (range 12.1 to 16.1 months, 19 females) and their parents. There was a failure of auditory recording on one infant whose data were included in analyses of infant gaze distributions and the relation of infant gaze to parent gaze and touch, but were excluded in analyses involving parent talk. Eighteen additional infants were recruited but did not contribute data because of refusal to wear the head-mounted device or other technical failure. Families were recruited from a population of working and middle class families. Participants were given a small toy as compensation for their participation in the study. This research project was approved by the Research Subjects Review Board at Indiana University (protocol number 0808000094) and titled "Multimodal word learning".

Stimuli

A pool of 15 novel objects (on average, about $9.50 \times 6.5 \times 5.0$ cm) was created in the lab with unique shapes and objects were combined in sets of 6 in three different ways (each set defined a unique experiment because the selected participants for the analyses were pooled from these three experiments). Through piloting, objects were designed to be visually and manipulatively engaging. All children were given objects in sets of three in which one was painted blue, one red and one green. Each child played with two unique sets of objects and the criterion for selecting a set of toys for an infant was that the toys were novel for that infant. Figure 1 shows a participating dyad as well as one of the sets of objects used for the study.

Experimental Room

Parent and infant sat across from each other at a white table 61cm × 91cm × 64cm (see in Figure 1). The infant sat on a chair and the parent sat on the floor such that the tabletop was approximately 46 cm from the center of the table to the child's eye and to the parent's eye. Both participants wore head-mounted eye trackers (Positive Science LLC, http://www.positivescience.com; also see Franchak & Adolph, 2010; Franchak et al., 2011). Both parent and infant eye-tracking systems included an infrared camera –mounted on the head and pointed to the right eye of the participant that recorded eye images– and a scene camera that captured the events from the participant's perspective. The scene camera's visual field was 108 degrees. 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. In order to eliminate distractors in the environment and encourage infants to focus on object play, everything in the room –other than the objects and the hands and faces of the participants– was white. Three additional cameras recorded the interaction from third-person views.

Procedure

Prior to entering the testing room, an experimenter desensitized the infant to touches to the head and hair by lightly touching the hair several times when the attention and interest of the infant was directed to a toy. Upon entering the experimental room, a second experimenter and the parent engaged the infant with a toy with buttons to push that made animals pop up as the first experimenter placed the head gear on the infant. This was done in one movement and care was taken to ensure that the infant remained engaged with the toy and that the infant'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 infant was in the center of the scene camera. The experimenter then directed the child's attention toward an attractive toy on the table ensuring the child's eyes were following the toy. This procedure was repeated in 15 different locations on the tabletop to ensure a sufficient number of calibration points for the infant's eye-tracking. After placing the parent's head gear the experimenter asked the parent to look at one of the objects on the table in various locations. This procedure was repeated 15 times in order to obtain at a sufficient number of calibration points for the parent's eye-tracker. Parents were instructed to play with their child with 3 toys at a time as they would normally do at home and asked if they named the object to use the objects' labels provided. Parent-infant dyads played in four 1.5-minutes-long trials, using two different sets of 3 toys in an alternating fashion across the 4 trials. The duration of the trials was chosen so that infants remained engaged in play with limited off-task behavior during the entire experiment. This duration is also consistent with previous research (Yu & Smith, 2013; Yu & Smith, 2016; Yu & Smith, 2017).

Coding and Definitions

The quality of the eye tracking videos (with eye images superimposed) for each infant and parent was checked (using centered hand actions on an object as described above) to ensure the quality of calibration throughout the session, at the end as well as at the beginning of the session. The eye-tracker collected data at a rate of 30 frames per second for approximately 360 seconds (four trials with 1.5 minutes per trial) of interaction, yielding potentially 10,800

data points per measure for each participant. Of this total possible, the number of analyzed frames –frames in which infant gaze was directed to one of the regions of interest—was 7,816 (SD = 1,893) and for parents, it was 8,697 (SD= 1,970). The "missing" frames include eye-blinks and periods when the infant was off-task (e.g. looking around the room rather than at the objects or parent).

Looks.—The main data for analyses were eye gaze data directed to four regions of interest (ROIs): the three objects in play at any time and the partner's face. ROI coding was done by highly trained human coders who continuously code these variables for multiple experiments without knowledge of or regard to the hypotheses under test. The ROI coding for this experiment was done as part of that workflow. Each ROI was strictly defined in terms of the in-view pixels belonging to the object. The coders annotated gaze direction – frame by frame – judging whether the cross hairs fell on the pixels of the ROI. Thus, an *unbroken look* to an object might have multiple fixations on the object as long as all gaze fell within the ROI. Reliability was computed between the coding of two independent coders on eleven dyads that were randomly selected. Coders coded 25% of each dyad's frames making judgments on 2,790 frames per dyad on average. The inter-coder reliability of eye-gaze coding performed by these highly trained coders ranged from 82% to 95% assessed by Cohen's kappa of 0.75 (ranging from 0.57 to 0.91). This level of reliability is consistent with the reliability reported by previous published research (see Yu & Smith, 2016; Yu & Smith, 2017).

The main dependent variable is the duration of infant unbroken looks within an ROI. Infants may generate multiple fixations on the same object which is counted here as one unbroken look. For the reported analyses, the duration of an infant's continuous gaze within an ROI needs to be longer than *500 milliseconds (msec) for that infant's continuous gaze* to be counted as a look. We did this because our interest is in sustained attention and because the dynamics of infant looking behavior are much slower than adults such that meaningful looks are at least this long (Yu & Smith, 2013; Yu & Smith, 2017). This approach allowed us to measure the parent behaviors that overlapped with unbroken and meaningful infant looks that were longer than this minimal duration. In order to ensure that this imposed floor on infant unbroken looks did not determine the results, we repeated the analyses using all infant unbroken looks, including those as brief as one frame (33 msec). The patterns of results and conclusions remained the same.

The main context of interest to examine infant looks is whether an infant look did or did not include joint attention with the parent, or moments in which the infant and parent looked at the same object. We defined a joint attention objectively in terms of parent looks to an object that temporally overlapped with unbroken infant look to that same object. We counted all cases in which a parent look to the object (regardless of the duration and thus in principle as brief as a single frame or 33.3msec) overlapped with an infant look to that same object (defined as lasting at least 500msec). In this way, we divided all infant looks into two categories – those that overlapped with a parent look to the same object or those that did not. We took this approach so as to cleanly capture all looks by the infant of which the parent might be aware of the child's interest and thus behave in some way that encouraged that interest. Adults (unlike infants, Yu & Smith, 2013; Yu & Smith, 2017) rapidly shift gaze and

can pick up and process useable information in very brief glances (Carpenter, 1988; Land & Hayhoe, 2001; Land, Mennie & Rusted, 1999).

Parent hand contact.—Parent manual contact with an object was coded frame-by-frame from images captured by the overhead camera and the other two third-person cameras. Although parent touch was coded in all the frames, the only relevant coded frames for parent touch used for the analyses were frames that occurred when the infant looked at the objects. Parent touch was counted only when the parent touched the object attended by the infant. We used 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 overhead camera, but in cases of uncertainty could consult the other two views. Coders made frame-by-frame yes/no decisions that a parent hand was in contact with an object. A second coder also independently coded a randomly selected 25% of the frames of five parents and obtained inter-coder reliability assessed by Cohen's kappa of 0.90 (ranged from 0.76 to 0.96).

Parent talk.—Parent speech was objectively coded at the utterance level, starting a new utterance after 400 milliseconds of silence (Suanda, Smith & Yu, 2016; Pereira, Smith & Yu, 2014; Yu & Smith, 2012). We included as speech, all sounds (words and word-like sounds) that included a vowel. This criteria excludes sounds such as coughs, raspberries or sighs and does not consider the content of the talk, treating naming ("it looks like a helicopter"), pointing out attributes ("that can spin") and general comments ("cool" or "that's fun") as all the same. Our assumption was that if parent talk had an effect on the duration of infant attention, it would be discernible from this coarse coding and thus our approach would be the right first step prior to a closer examination of effects of kinds of content, as well as prosody, which may be influential factors.

Statistical Analyses

The main empirical question concerns different kinds of joint attention experiences based on their multimodal components (parent talk and touch) and their effects on infants sustained attention to the jointly attended object. To this end, the results consist of three parts, with the first two being preliminary to the main question. First, we measure the duration of visual attention in infants. Second, we compare infant looks that did or did not include joint attention by the parent. By doing so, we replicate Yu & Smith (2016), testing the contribution of joint attention –that is parent look to the infant attended object– to the duration of the infant's unbroken visual attention to that object. Third, we turn to the main question, the multimodal nature of these joint attention episodes —defining four different categories — based on the combinations of parent talk and touching of the jointly attended object and their association with different durations of infant look to the object.

The main dependent variables of interest for all analyses are the durations of infant looks under different conditions. The distribution of infant look durations is extremely skewed (see Figure 2) as is true of many human behaviors (Clauset, Shalizi & Newman, 2009; Clerkin, Hart, Rehg, Yu & Smith, 2017; Kello, Brown, Ferrer-i-Cancho, Holden, Linkenkaer-Hansen, Rhodes & Van Orden, 2010; Piantadosi, 2014). Accordingly, and as is appropriate for these

right skewed distributions, we categorized infant looks based on their durations into bins (e.g., brief, long and very long as defined below) and counted the number of looks for each participant in each bin (normalized as a proportion of all looks by that infant that were in each bin since different children had different numbers of looks) using both parametric and nonparametric statistics. The main analysis, however, consists of a linear mixed effects model with fixed and random effects (R Development Core Team, 2006) on the logs of the look durations.

Results

I. The Distribution of Infant Attention

Table 1 provides a summary of all infant unbroken looks to an object or parent face without regard to parent looks or other behaviors. Infants looked at the objects much more than they looked at the parent's face, a result that has been reported by many other investigators in a variety of social contexts for infants this age (Bakeman & Adamson, 1984; Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Franchak, Kretch, Soska & Adolph, 2011; Yoshida & Smith, 2008; Yu & Smith, 2013). Figure 2 shows the histogram of the durations of infant unbroken looks to the objects, the main dependent measure in subsequent analyses. Infant look durations varied from 500 msec (the imposed floor) to nearly 31 sec. Most looks to objects were very brief but the tail of the distribution is quite long. By hypothesis, these very long but relatively infrequent looks that comprise the tail of the distribution are most relevant to assessing the role of parent behavior in sustaining infant attention. Accordingly, for the categorical analyses of look durations as a function of parent behavior, we used two main categories of durations: Brief looks, less than 3 sec (the threshold for sustained attention used in previous studies, Yu & Smith, 2016; Ruff & Lawson, 1990), accounted for about 75% of all infants looks and Long looks, 3 sec and longer, and typically considered sustained attention, accounted for about 25% of all infant looks. Within the Long looks –and *included in all Long Look analyses* -- we provide additional information about what we call Very Long looks. These are looks that are 10 sec and longer. They are not common, accounting for just 2% of all infant looks (Figure 2). Whereas all infants had at least some Long looks, not all of them had Very Long (10 sec or greater) looks. Nonetheless, we include results of Very Long looks because 65% of all infants had at least one Very Long Look, and because, as we report subsequently, these Very Long infant looks were associated exclusively with joint attention or moments in which the parent also looked at the infant-attended object.

II. Joint Visual Attention

Each infant look was classified in one of two mutually exclusive categories–including or not including joint attention with the parent. If during any period of the infant's continuous gaze to the object, the parent also directed gaze to same object (no matter how briefly), *the entire infant look* was categorized as including joint attention, as illustrated in Figure 3. The durations of parent overlapping looks could be short or long, and the parent look to the object could follow the infant's look (infant-led) or could precede it (parent-led). By these definitions, parent and infants jointly attended to objects during play 49% (SE=2%) of the play session whereas infants looked at objects without an overlapping parent look 12%

(SE=1%) of the time. The remaining times consists of looks shorter than 500 msec, looks elsewhere in the room, or to the partner's face. Overall, the results show, as has been reported before (Yu & Smith, 2013) that parents and infants consistently coordinate their gaze to the same object during free-flowing play.

Figure 4A shows the histogram of the durations of the parents' overlapping looks to the same object to which the infant attended. Most parent looks were very brief and overall much shorter than infant looks (parent looks were on average 1.28-seconds-long (SD=0.53, Median=1.16), infant looks were on average 2.26-seconds-long (SD=0.51, Median=2.26)). Figure 5A shows the frequency of parent-led and child-led looks to the same object and the relative timing of the onset of parent looks to the infant-attended object. We used very small bin sizes incrementing at a tenth of a second in order to show the tight temporal coordination of parent-and infant gaze to the same object (see also Yu and Smith, 2013). Parents follow the infant's gaze to the object more often than they lead, but they follow rapidly; the mode gap between infant onset and parent onset of gaze to the same object is at the bin between 0 and .10 second. Figure 5A shows that 76% of the overlapping parent looks occurred within 1 second before or 1 second after the onset of the infant's look. This tight coordination in time rules out one uninteresting account of why infant looks might be longer when parents also look at the same object: longer looks by the infant could have provided more time for an overlapping glance by the parent and thus greater likelihood the infant look is counted as including joint attention. But this did not happen as parents looked to the object close in time to the start of the infant's own look.

Figure 6 compares the durations of infant looks that include joint parent attention to the object to those that do not. As it is evident, the frequency of infant looks with JA was greater, 9.88 per min (SE=0.33) than those without with JA, 5.88 per min (SE=0.35). More critically, the duration of infant's individual looks were longer with JA than without. That is, we observed the same pattern reported by Yu & Smith (2016): the duration of infant looks to an object was longer when parents also looked at the object (M=3.03sec, SD=0.69, Median=2.95, SE=0.11) than when they did not (M=1.26sec, SD=0.26, Median=1.18, SE=0.04), t(39)=17.64, p= 2.2e-16). We also performed a Wilcoxon Signed-rank test because whereas the difference scores used in the t-test are normally distributed, the durations of infant looks with and without JA themselves are not. The Wilcoxon Signed-rank test also confirmed that infant looks with joint attention were longer in duration than infant looks without joint attention, Z=−5.51, p<0.001. Figure 6 suggests that two distributions of infant looks –with and without JA –differ primarily in the longer tail of long looks. Accordingly, we determined the normalized count of each infant's Long looks with JA and without JA, the previously used threshold for sustained attention (Yu & Smith, 2016; Ruff & Lawson, 1990), dividing the count of Long Looks with JA, for each subject, by that subject's total number all looks with JA and likewise, the count of Long Looks without JA by total number of looks without JA. There were more Long looks with JA (Mean proportion=0.37, SD=0.10, Median=0.36. SE=0.02) than without JA (Mean proportion=0.05, SD=0.05, Median=0.05, SE=0.01), t(39)= 19.01, p=2.2e-16). The Wilcoxon Signed-rank test confirmed this finding, Z=−5.51, p<0.001. Very Long looks –at least 10-seconds-long— by the child, only occurred when parents also shared attention to that object (Mean proportion=0.03, SD=0.03, Median=0.02, SE= 0.005). The duration of infant looks with JA

did not differ as a function of who led the look to the object (parent-led: 2.91sec, SD=0.94, Median=2.67, SE=0.15; infant-led: M= 3.05sec, SD=0.70, Median=2.97, SE=0.11; t(39)=1.03, p=0.31, ns. A Wilcoxon Signed-rank test confirmed this finding, Z=−1.18, p=0.24, ns. In brief, the main finding is this: There were more Long unbroken looks to a single object by infants when their parent also looked to the same object during the infants' unbroken look.

Because infant looks to the object are defined as unbroken looks, with no glances away from the object, the infant –when looking at the object– cannot be looking directly at the parent's face. Given this, it is unlikely that the infant during their sustained looks to an object used the parent's look itself as the indicator that the parent was simultaneously attending to the same object. In addition, peripheral vision seems an unlikely source of such information as considerable evidence indicates that gaze following is very difficult for adults, older children and infants in natural contexts with freely moving heads and objects; indeed, success in following gaze —for adults, children and infants — is quite poor in any context in which there are not just two choice objects widely separated in space on opposites sides of the midline of the person directing gaze (Corkum & Moore, 1998; Doherty, Anderson & Howieson, 2009; Farroni, Johnson, Brockbank & Simion, 2000; Langton, Watt & Bruce, 2000; Loomis, Kelly, Pusch, Bailenson & Beall,2008 ; Vida & Maurer, 2012a–2012c). In principle, it is also possible that infants looked at the parent's face *just prior* to looking at the object themselves and registered a prior look to the object by the parent. This also does not seem likely since parents follow infants' looks to an object more than they lead (Figure 5A). Although we cannot definitively rule out that infants had some direct knowledge that their parent looked at the object by directly perceiving the parent's look while the infant was looking at the object, it seems more likely that other parent behaviors that overlap with infant and parent looking at the object may be the behaviors that influenced the infant's continued visual interest in the object and signaled parent's engagement with the object.

III. Multimodal Parent Behaviors

Given that an infant is looking at an object and the parent also looks to that object, what else does the parent do that might influence the duration of infant attention? Figure 7 illustrates how we identified other potentially relevant combinations of parent behaviors —talk and touch of the infant-attended object— for consideration. These additional behaviors were located as potential influences if they overlapped with *an infant's look* that was part of a joint attention episode. The additional parent behavior could be long, or short in duration and could or could not overlap in time with the parent look that defined the infant's look to the object as including joint attention. The additional parent behavior also could precede or follow the parent look to the infant attended object. That is, , the *only* —and objective — criterion for considering the additional parent behavior's effect on infant looking was that the parent behavior overlapped in time with the infant's look to the *jointly* attended object. In this way, we know the parent was attentive to the object of infant interest (not some other object) and thus that these additional behaviors likely reflected that interest.

Figure 8A shows the proportion of *all infant looks* with JA that overlapped in time, as described, with the four possible combinations of the additional parent behaviors and, as

expected, it indicates that there was much more going on than just parent and infant looking the same object. When the infant was looking at an object, and the parent also looked at the object, parents added at least one additional behavior — talk or touch— as an indicator of their interest, over 87% of all JA episodes. This fact indicates that joint attention typically occurred with more multimodal parent *engagement* with the object. Infant look durations for these four categories of infant looks with JA along with infant looks without JA are the data submitted to the main analyses. Accordingly, Figure 8B shows the mean proportions of the count of *all* infant looks (with and without JA) that were of each of the four JA categories or were without JA. Figures 4B and 4C show the histogram of the durations of parent talk and hand contact with to the infant-and parent-attended object. Both distributions are again extremely skewed with most events being quite brief, but talk is overall much briefer and consistent in its timing whereas some touches can be quite long and are more variable (note 80th percentile of utterance duration is much smaller than the 80th percentile of touch duration). Figures 5B and 5C show the timing of parent talk and touches relative to the onset of the infant look. As is the case for parent looks (Figure 5A), parent talk and parent touches are all centered on the onset of the infant look to the object and because of the tight coupling of parent look onset to infant look onset, they are also centered on the onset of the parent look. Figures 5B and 5C show the proportion of events that fall within 1 second and within 2 seconds of onset of infant look for both talk and touch respectively. These results show the onset of parent touch is less coupled in time with the onset of the infant's look as compared to the onset of parent look and talk. More than two thirds of the data are within −1 and 1 seconds for both parent look and parent talk but not for parent touch. The key question is the potential influence of these multimodal parent behaviors on the duration of infant attention to the object.

The observed often tight temporal coordination of parent looks, talk and touch of the infant-attended object emphasizes the complexity of naturalistic free-flowing parent behavior and the difficulties in singling out the role played by any individual component of parent behavior. Accordingly, the approach we took is to define four categories of parent behavior during infant attention to an object, given that the parent had also looked to that object (also shown in Figure 7): (1) the parent looked at the infant attended object but did not talk or touch; (2) the parent looked and also touched the infant attended object, (3) the parent looked at the infant-attended object and also talked, and (4) those in which parents looked at the infant-attended object, talked and touched it. We compare the durations of infant looks in these four categories with a baseline of the duration of infant looks when parent did not also look at the object, yielding 5 categories in the analysis. Because durations are not normally distributed but skewed, we submitted the logs of the durations of the infant looks to a linear mixed-effects models (code online @ https://github.com/csuarezr26/ MultimodalJA_andSA2018) using the lme4 package version 1.1.12 (Bates, Maechler, Bolker & Walker, 2015), and the lsmeans package version 2.27.2 (Lenth, 2016) in the R environment (version 3.3.2) (R Development Core Team, 2006). The model predicted the log(duration) of individual episodes of infant attention from a fixed effect, which specified the 5 possible categories. Random intercepts were specified for individual infants -which controls for infants having different durations of attention bouts-and for the specific object attended to by the infant –which controls for different objects having different durations of

attention bouts-(i.e., random intercept model, Pinheiro & Bates, 2000; for an application, see Oberauer & Kliegl, 2006). Finally, the package used restricted maximum likelihood specifying an unstructured covariance matrix. We used this linear mixed model because it allowed us to ask how the specific combinations of parent behaviors may differentially predict the duration of attention bouts while accounting for the fact that duration of infant looks are not independent of each other and could vary across infants.

Figure 9A illustrates the results, showing the predicted mean duration and 95% confidence interval of infant looks in the 5 categories specified in the model. First, the durations of infant looks given that a parent looked at an object, the four JA categories, were each significantly longer than infant looks to an object without a parent look to that same object. JA episodes that included both parent talk and touch coincided with the longest looks by the infant to the object. Table 2 shows the estimated difference, SE and p-values with respect baseline category for each of the four categories of JA. The model's intercept, in logarithmic scale, was -0.14, which corresponds to an estimated mean duration of 0.87 seconds (as the natural exponential of -0.14 is 0.87, $\exp(-0.14)=0.87$) for the duration of the baseline category (infant looks without JA). The estimated differences shown in Table 2 are in logarithmic scale and they represent the difference between the predicted mean log(durations) of each of the categories and the mean log(duration) of baseline (0.87 seconds, also shown in Figure 9A). The values in Table 2 can be converted, by computing their natural exponential, to represent also the increase in mean duration from baseline, for each category, as a factor of the duration of the baseline category. In this way, for instance, the estimated mean duration of category 2 (infant looks with JA and no additional behaviors) is 1.46 times the duration of the baseline category (as $\exp(0.38)$, the natural exponential of 0.38, is 1.46). The estimated mean duration of category 4 (infant looks with JA, touch and talk) was 3.19 times the duration of the baseline category ($\exp(1.16)=3.19$)).

As is evident, parent talk had a much greater effect on the duration of infant looks to the object than did touch. Pairwise contrasts (Figure 9B) show: (A) Infant look durations with JA that included parent talk were longer than infant look durations with JA with no additional parent behaviors; (B) Infant looks with JA that included only parent talk were longer than those with JA than included only parent touch of the object; (C) Infant look durations overlapping with JA and parent hand contact were not reliably longer than those that included no additional measured parent behaviors; (D) However, JA episodes that included both parent talk and parent hand contact with the objects were associated with longer infant looks to the object compared to JA episodes including parent talk alone, indicating that the additional behavior of touching the attended object does add to the strength of parent interest expressed by talk. The fit of the overall model decreases if the single fixed factor (the five categories based on parent behavior) is removed, leaving only the random factors as predictors: $\chi^2(4) = 715.58$, $p<.00001$. We compared the model to a linear mixed effect model that included random slopes for infants' effect of category of look on log(duration) of looks in order to test if the more complicated model, with random slopes, provided a better fit. The results suggests this was not the case as both models provided a similar fit for the data ($\chi^2(14) = 18.48$, $p=0.18$).

One open question raised by the results in Figure 9A is the role of parent looks to the objects—either with or without the additional behaviors. Because infants' rarely look directly to the parent face during the session and do not do so during the measured dependent variable of continuous gaze directed to an object, it seems unlikely that parent looks in and of themselves directly influence the duration of the infant look to the object. However, the results show that a parent look to the object is a critical component of multimodal parent behavior.

First, parent talk without an overlapping parent look was not associated with more enduring infant looks to the attended object. Infant looks that overlapped with talk but did not include joint attention (or an overlapping parent look to the same infant-attended object) were much shorter (M= 1.41 sec, SD=0.34, Median=1.34, SE=0.05) than infant looks that included talk and also joint attention (M=3.48 sec, SD=0.90, Median=3.28, SE=0.14), $t(38)$=15.46, p<0.001. A Wilcoxon Signed-rank test confirmed that infant looks *with* joint attention and with talk were longer in duration than infant looks *without* joint attention and with talk, Z= −5.44, p<0.001. Second, parent touches without an overlapping parent look as well, were not associated with longer lasting infant looks to the objects. Infant looks that overlapped with touch but did not include joint attention were significantly shorter (M= 1.30 sec, SD=0.39, Median=1.18, SE=0.06) than infant looks that overlapped with touch and also with joint attention (M=3.39 sec, SD=0.78, Median=3.29, SE=0.12), $t(39)$=17.05, p<0.001. A Wilcoxon Signed-rank test confirmed this finding, Z=−5.51, p<0.001. These results suggest parent talk and touch without a coordinated parent look to the infant-attended object have more limited influences on infant's attention compared with multimodal parent behaviors that also include parent looks.

One limitation on these conclusions is the rarity of talking or touching by a parent when the parent is not also visually attending to the object attended to by the infant. Infant looks that overlapped with talk but not with a parent look occurred on average 3.07 times per minute of eye-tracking data (SD=1.46, SE=0.23) while infant looks that overlapped talk with JA occurred on average 9.37 times per minute of eye-tracking data (SD=2.30, SE=0.37). Infant looks that overlapped with touch but not with a parent look occurred on average 0.84 times per minute of eye-tracking data (SD=0.73, SE=0.12) while infant looks that overlapped with touch with JA occurred on average 8.37 times per minute of eye-tracking data (SD=2.15, SE=0.34). This is not surprising; human behavior is a coordinated multimodal event in which we look at what we handle (Yu & Smith, 2013) and look at the referents of our talk (Henderson & Ferreira, 2004; Griffin & Bock, 2000).

However, parents sometimes looked at the attended object without also touching it or talking and these moments also were associated with longer infant looks to the object than when parents did not also look at the object (Table 2). If infants cannot directly perceive parent's eye gaze in this circumstance, then parents must be showing their interest through other unmeasured behaviors, perhaps leaning forward or hands close to but not touching the object, or perhaps even a stillness of body so as not to disrupt (and/or redirect) the infant's attention. This is a question for future research.

In summary, the main finding appears to be this: During joint parent-infant interaction with objects, parents often look at the objects to which their infants are visually attending and when they do, they often talk and touch the object during the period of the infant's attention to that object. All of these behaviors appear to increase the duration of the infant's attention to the object. The most potent combination —which also happened most often in free-flowing interaction— is a multimodal behavior that includes the parent's look, talk and touch and increases the duration of infant look to the object by about 1.5 sec (on average) as to when the parent displays no interest at all. This may seem a relatively small increase; however, as proposed by Yu and Smith (2016) and as we consider in the General Discussion, these small increases —repeated over and over within a single play session and across the days and weeks of the infant's life— may have long term consequences on infant development.

Discussion

The results show that infant sustained attention is more likely and more enduring when parents also visually attend and express interest through talking or handling the object attended to by the infant, providing converging evidence for a role of social context in infant sustained attention (Yu & Smith, 2016). The new contributions are that in addition to parent look, multimodal behaviors are embedded within these joint attention moments and that both parent hand actions and parent talk within the context of shared attention to an object sustain infant visual attention to the jointly attended object. The findings have implications for current understanding of the relation between joint attention and sustained attention, and, most critically, for experiential factors in the development of sustained attention.

Joint Attention and Sustained Attention

Joint attention and sustained attention have been studied separately but not together. Joint attention has been traditionally examined in social contexts, with parent gaze understood as a potential social signal to be read and responded to by the infant (e.g, Baron-Cohen & Cross, 1992; Brooks & Meltzoff, 2005); infant sustained attention has been considered as an individual achievement in early development (Colombo, 2001; Pérez-Edgar, McDermott, Korelitz, Degnan, Curby, Pine & Fox, 2010; Posner & Rothbart, 2000; Ruff, 1990). But parent-infant joint attention involves more than gaze behavior and contains the infant's own visual attention to the object. Parents do not just look to objects when they share attention with their infant; they express their interest in multimodal behaviors, which have direct effects on infant visual attention. Thus, joint attention may be best conceptualized not as shared visual attention to an object but as a proxy of a suite of multimodal temporally coordinated parent and infant behaviors. The present results clearly show the fine-grained temporal coordination of the onsets of parent multimodal behaviors directed to an object – gaze, talk, touch– and the onset of infant visual attention. The lag between the onset of any of the three types of parent behavior and the onset of infant looking fell at near zero in all cases (shown in Figure 5). This tight temporal entrainment of infant looking and multimodal parent behaviors marks joint attention episodes as a potentially powerful context for multiple aspects of human development, including the self-regulation of attention. One implication is that the strong predictive relations between joint attention and other developments such as

word learning (Mundy, Block, Delgado, Pomares, Van Hecke & Parlade, 2007; Tomasello & Farrar, 1986;) may emerge not exclusively through shared gaze but through the other parent behaviors that are part of naturally occurring episodes of parent-infant joint attention. This idea is further supported by a recent study (Yu, Suanda, & Smith, in press), showing that infant sustained attention but not joint attention in toy play is predictive of later language outcome.

The multimodal nature of joint attention as evidenced by the parents in this study may help understand how joint attention works in the wild, given that infants engaging in object play rarely look to the faces of their partners (Bakeman & Adamson, 1984; Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Franchak, Kretch, Soska & Adolph, 2011; Yoshida & Smith, 2008; Yu & Smith, 2013). The operational definition of joint attention used in the present study differs from the definition of "joint attention" used in many discrete trial experimental studies. (e.g, Baron-Cohen & Cross, 1992; Brooks & Meltzoff, 2005; Mundy, Block, Delgado, Pomares, Van Hecke & Parlade, 2007). In those studies, researchers often sought evidence that the child was aware of the mature partner's direction of attention, requiring infant looks to the partner's face for coordinated visual attention to the same object to count as joint attention. Here, we used a simpler and more objective measure, the degree to which parent and child directed gaze to the same object at the same time, a measure we believe is more fitting for naturalistic object play (Yu & Smith, 2013; Yu & Smith, 2017). However, just because the infant does not look at the parent face does not mean that the child is unaware of the parent's direction of attention. Hands, talk, as well as other yet unmeasured behaviors are likely well-read cues by infants about their parent's current state of interest. The close temporal coordination of parent looks, talk, and touch have been noted by other researchers as a powerful combination that makes parents referential intentions transparent (Trueswell, Lin, Armstrong, Cartmill, Goldin-Meadow & Gleitman, 2016) and indicates timely and coordinated responsivity on the parents part to infant interests (Van Egeren, Barratt & Roach, 2001). Thus, instead of focusing solely on parent gaze and child gaze in joint attention, the present study suggests a broader context to examine joint attention embedded in the multimodal parent behaviors that take place in naturalistic social interactions.

Moreover, the results also suggest the value of a more unified study of joint attention, infant sustained attention, and parental responsiveness. First, if episodes of joint attention are the context in which parents scaffold the development of infant sustained attention and if that scaffolding requires rapid responses from parents to the infant's attentional state in fractions of seconds, then the complete explanation of the development of sustained attention –and individual differences in that development– will depend on understanding how all of these components interact and how they vary across dyads. Second, numerous studies directed to these components –joint attention, infant sustained attention, parent responsiveness– show they each predict later outcomes. The present results raise the possibility that joint attention and parental responsiveness are predictive because they support the development of sustained attention and the self-regulation of attention that is essential to learning in all domains. This hypothesis can be correct only if parent behaviors have their effects on infant attention through processes related to the development of the self-regulation of attention.

A Training Ground for the Self-Regulation of Attention?

The importance of parent behavior in supporting infant attention to an object certainly lies in the in-the-moment effects that this behavior may have on the duration of the infant's attention and thus on infant learning in the moment. A recent study showed that infants engage in sustained attention more in the context of parent-infant joint play and less during contexts in which the infant plays with objects alone (Wass, Clackson, Georgieva, Brightman, Nutbrown & Leong, 2018). The accumulating evidence points to the implication that mature social partners may support in-the-moment visual learning of the attended object –documented by studies that examine infant sustained attention alone (e.g., Ruff, 1986)— through their effect on the infant's visual attention. However, the more intriguing possibility is that these day-in day-out small effects of parent behavior on sustained attention to an object may serve as the experiential training ground for longer term effects. How might such a mechanism work? There are multiple inter-related mechanisms that may be involved. For example, parent behavior could signal parent interest, and the infant's awareness of that interest may be a factor that keeps the infant attending to the object. It also possible that parent expressed interest may be rewarding to the infant, engaging reward mechanisms that have been shown in adults to support persistence in following goals and avoiding (Insel, 2003; Montague, Hyman & Cohen, 2004). Words are well known drivers on visual attention in infants as well as older children (Carvalho, Vales, Fausey & Smith, 2018; Fernald, Thorpe & Marchman, 2010; Johnson, McQueen & Huettig, 2011; Landau, Smith & Jones, 1992; Vales & Smith, 2015;), and the most powerful parent behavior observed in the present study. Finally, the dynamic coordination of parent and infant behavior in real time may operate in real time more like a dyad dancing together to socially entrain (and thus train) internal attentional control mechanisms (Marsh, Richardson & Schmidt, 2009; Takahashi, Narayanan & Ghazanfar, 2013). There is some evidence consistent with the proposal that increased exogenous (externally driven) attentional capture that is likely created by the parent is responsible for the increase in infant's sustained attention appearing in contexts of joint play (Wass et al., 2018). The distinction between top-down versus bottom-up processes as well as between endogenous versus exogenous drivers of attention during joint play with a social partner are key issues for future research.

In sum, the present findings strongly implicate that multimodal parent behaviors during joint attention may –in one way or another– influence sustained attention in real time, with the potential to influence the internal mechanisms that underlie the development of the self-regulation of attention. The effects of any single joint attention bout may be quite small with the infants' gaze to the object extended only slightly (Yu and Smith, 2016). Nonetheless, the aggregated effects on the development of self-regulated attention may be quite large as these small socially guided extensions of infant visual attention may occur multiple times a day, day-in and day-out in the social lives of infants.

References

Bakeman R, & Adamson LB (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Development, 1278–1289. 10.2307/1129997 [PubMed: 6488956]

Baldwin DA, & Markman EM (1989). Establishing word-object relations: A first step. Child Development, 381–398. 10.2307/1130984 [PubMed: 2924658]

Baron-Cohen S, & Cross P (1992). Reading the eyes: evidence for the role of perception in the development of a theory of mind. Mind & Language, 7(1-2), 172–186. 10.1111/j. 1468-0017.1992.tb00203.x

Bates D, Maechler M, Bolker B, Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. doi: 10.18637/jss.v067.i01 . 10.18637/ jss.v067.i0110.18637/jss.v067.i01. https://doi.org/10.18637/jss.v067.i01

Belsky J, Goode MK, & Most RK (1980). Maternal stimulation and infant exploratory competence: Cross-sectional, correlational, and experimental analyses. Child Development, 1168–1178. 10.2307/1129558 [PubMed: 7471921]

Bornstein MH, & Tamis-Lemonda CS (1997). Maternal responsiveness and infant mental abilities: Specific predictive relations. Infant Behavior and Development, 20(3), 283–296. 10.1016/ S0163-6383(97)90001-1

Brooks R, & Meltzoff AN (2005). The development of gaze following and its relation to language. Developmental Science, 8(6), 535–543. 10.1111/j.1467-7687.2005.00445.x [PubMed: 16246245]

Carpenter RH (1988). Movements of the Eyes, 2nd Rev Pion Limited.

Carvalho PF, Vales C, Fausey CM, & Smith LB (2018). Novel names extend for how long preschool children sample visual information. Journal of experimental child psychology, 168, 1–18. 10.1016/ j.jecp.2017.12.002 [PubMed: 29287205]

Clauset A, Shalizi CR, & Newman ME (2009). Power-law distributions in empirical data. SIAM review, 51(4), 661–703. 10.1137/070710111

Clerkin EM, Hart E, Rehg JM, Yu C, & Smith LB (2017). Real-world visual statistics and infants' first-learned object names. Phil. Trans. R. Soc. B, 372(1711), 20160055 10.1098/rstb.2016.0055 [PubMed: 27872373]

Colombo J (2001). The development of visual attention in infancy. Annual review of psychology, 52(1), 337–367. 10.1146/annurev.psych.52.1.337

Corkum V, Moore C (1998) Origins of joint visual attention in infants. Developmental Psychology 34: 28 10.1037/0012-1649.34.1.28 [PubMed: 9471002]

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

Doherty MJ, Anderson JR, Howieson L (2009) The rapid development of explicit gaze judgment ability at 3 years. Journal of experimental child psychology, 104: 296–312. 10.1016/j.jecp. 2009.06.004 [PubMed: 19640550]

Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, Pagani LS, Feinstein L, Engel M, Brooks-Gunn J, Sexton H, Duckworth K, & Japel C, (2007). School readiness and later achievement. Developmental Psychology, 43(6), 1428 10.1037/0012-1649.43.6.1428 [PubMed: 18020822]

Farroni T, Johnson MH, Brockbank M, Simion F (2000) Infants' use of gaze direction to cue attention: The importance of perceived motion. Visual Cognition, 7: 705–718. 10.1080/13506280050144399

Fernald A, Thorpe K, & Marchman VA (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. Cognitive Psychology, 60(3), 190–217. 10.1016/ j.cogpsych.2009.12.002 [PubMed: 20189552]

Franchak JM, & Adolph KE (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. Vision Research, 50(24), 2766–2774. 10.1016/j.visres. 2010.09.024 [PubMed: 20932993]

Franchak JM, Kretch KS, Soska KC, & Adolph KE (2011). Head-mounted eye tracking: A new method to describe infant looking. Child Development, 82(6), 1738–1750. 10.1111/j. 1467-8624.2011.01670.x [PubMed: 22023310]

Griffin ZM, & Bock K (2000). What the eyes say about speaking. Psychological science, 11(4), 274–279. 10.1111/1467-9280.00255 [PubMed: 11273384]

Henderson JM, & Ferreira F (2004). Scene Perception for Psycholinguists

Insel TR (2003). Is social attachment an addictive disorder?. Physiology & Behavior, 79(3), 351–357. 10.1016/S0031-9384(03)00148-3 [PubMed: 12954430]

Johnson EK, McQueen JM, & Huettig F (2011). Toddlers' language-mediated visual search: They need not have the words for it. Quarterly Journal of Experimental Psychology, 64(9), 1672–1682. 10.1080/17470218.2011.594165

Kello CT, Brown GD, Ferrer-i-Cancho R, Holden JG, Linkenkaer-Hansen K, Rhodes T, & Van Orden GC (2010). Scaling laws in cognitive sciences. Trends in Cognitive Sciences, 14(5), 223–232. 10.1016/j.tics.2010.02.005 [PubMed: 20363176]

Kochanska G, Murray KT, & Harlan ET (2000). Effortful control in early childhood: continuity and change, antecedents, and implications for social development. Developmental Psychology, 36(2), 220 10.1037/0012-1649.36.2.220 [PubMed: 10749079]

Kopp CB (1982). Antecedents of self-regulation: a developmental perspective. Developmental Psychology, 18(2), 199 10.1037/0012-1649.18.2.199

Land MF, & Hayhoe M (2001). In what ways do eye movements contribute to everyday activities?. Vision research, 41(25–26), 3559–3565. 10.1016/S0042-6989(01)00102-X [PubMed: 11718795]

Land M, Mennie N, & Rusted J (1999). The roles of vision and eye movements in the control of activities of daily living. Perception, 28(11), 1311–1328. 10.1068/p2935. [PubMed: 10755142]

Landau B, Smith LB, & Jones S (1992). Syntactic context and the shape bias in children's and adults' lexical learning. Journal of Memory and Language, 31(6), 807–825. 10.1016/0749-596X(92)90040-5

Langton SRH, Watt RJ, Bruce V (2000) Do the eyes have it? Cues to the direction of social attention. Trends in Cognitive Sciences, 4: 50–59. 10.1016/S1364-6613(99)01436-9 [PubMed: 10652522]

Lansink JM, & Richards JE (1997). Heart Rate and Behavioral measures of Attention in Six-, Nine, and Twelve-Month-Old Infants during Object Exploration. Child Development, 68(4), 610–620. 10.2307/1132113 [PubMed: 9306641]

Lenth RV (2016). Least-Squares Means: The R Package lsmeans. Journal of Statistical Software, 69(1), 1–33. 10.18637/jss.v069.i01

Loomis JM, Kelly JW, Pusch M, Bailenson JN, Beall AC (2008) Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. Perception, 37: 1443–1457. 10.1068/p5896 [PubMed: 18986070]

Marsh KL, Richardson MJ, & Schmidt RC (2009). Social connection through joint action and interpersonal coordination. Topics in Cognitive Science, 1(2), 320–339. 10.1111/j.1756-8765.2009.01022.x [PubMed: 25164936]

McClelland MM, & Cameron CE (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. Child Development Perspectives, 6(2), 136–142. 10.1111/j.1750-8606.2011.00191.x

McClelland MM, Acock AC, & Morrison FJ (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. Early Childhood Research Quarterly, 21(4), 471–490. 10.1016/j.ecresq.2006.09.003

Miller JL, Ables EM, King AP, & West MJ (2009). Different patterns of contingent stimulation differentially affect attention span in prelinguistic infants. Infant Behavior and Development, 32(3), 254–261. 10.1016/j.infbeh.2009.02.003. [PubMed: 19327841]

Montague PR, Hyman SE, & Cohen JD (2004). Computational roles for dopamine in behavioural control. Nature, 431(7010), 760 10.1038/nature03015 [PubMed: 15483596]

Mundy P, Block J, Delgado C, Pomares Y, Van Hecke AV, & Parlade MV (2007). Individual differences and the development of joint attention in infancy. Child development, 78(3), 938–954. 10.1111/j.1467-8624.2007.01042.x [PubMed: 17517014]

Mundy P, Sullivan L, & Mastergeorge AM (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. Autism research, 2(1), 2–21. 10.1002/aur.61 [PubMed: 19358304]

Oberauer K, & Kliegl R (2006). A formal model of capacity limits in working memory. Journal of Memory and Language, 55(4), 601–626. 10.1016/j.jml.2006.08.009

Parrinello RM, & Ruff HA (1988). The influence of adult intervention on infants' level of attention. Child development, 1125–1135. 10.2307/1130279 [PubMed: 3168620]

Pereira AF, Smith LB, & Yu C (2014). A bottom-up view of toddler word learning. Psychonomic Bulletin & Review, 21(1), 178–185. 10.3758/s13423-013-0466-4 [PubMed: 23813190]

Pérez-Edgar K, McDermott JNM, Korelitz K, Degnan KA, Curby TW, Pine DS, & Fox NA (2010). Patterns of sustained attention in infancy shape the developmental trajectory of social behavior from toddlerhood through adolescence. Developmental Psychology, 46(6), 1723 10.1037/a0021064 [PubMed: 20873921]

Piantadosi ST (2014). Zipf's word frequency law in natural language: A critical review and future directions. Psychonomic Bulletin & Review, 21(5), 1112–1130. 10.3758/s13423-014-0585-6 [PubMed: 24664880]

Pinheiro JC, & Bates DM (2000). Linear mixed-effects models: basic concepts and examples. Mixed-effects models in S and S-Plus, 3–56. 10.1007/978-1-4419-0318-1_1

Posner MI, & Rothbart MK (2000). Developing mechanisms of self-regulation. Development and Psychopathology, 12(3), 427–441. 10.1017/S0954579400003096 [PubMed: 11014746]

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL https://www.R-project.org/.

Reck SG, & Hund AM (2011). Sustained attention and age predict inhibitory control during early childhood. Journal of experimental child psychology, 108(3), 504–512. 10.1016/j.jecp.2010.07.010 [PubMed: 20801457]

Richards JE, & Casey BJ (1992). Development of sustained visual attention in the human infant. Attention and information processing in infants and adults: Perspectives from human and animal research, 30–60.

Ruff HA (1986). Components of attention during infants' manipulative exploration. Child Development, 105–114. 10.2307/1130642 [PubMed: 3948587]

Ruff HA (1990). Individual differences in sustained attention during infancy. Individual Differences in Infancy: Reliability, Stability, Prediction, 247–270.

Ruff HA, & Lawson KR (1990). Development of sustained, focused attention in young children during free play. Developmental Psychology, 26(1), 85 10.1037/0012-1649.26.1.85

Sigel Irving E. "The psychological distancing model: A study of the socialization of cognition." Culture & Psychology 82 (2002): 189–214. 10.1177/1354067X02008002438

Spruijt AM, Dekker MC, Ziermans TB, & Swaab H (2018). Attentional control and executive functioning in school-aged children: Linking self-regulation and parenting strategies. Journal of Experimental Child Psychology, 166, 340–359. 10.1016/j.jecp.2017.09.004 [PubMed: 29024846]

Steinmayr R, Ziegler M, & Träuble B (2010). Do intelligence and sustained attention interact in predicting academic achievement?. Learning and Individual Differences, 20(1), 14–18. 10.1016/j.lindif.2009.10.009

Suanda SH, Smith LB, & Yu C (2016). The Multisensory Nature of Verbal Discourse in Parent–Toddler Interactions. Developmental Neuropsychology, 41(5–8), 324–341. 10.1080/87565641.2016.1256403 [PubMed: 28128992]

Takahashi DY, Narayanan DZ, & Ghazanfar AA (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. Current Biology, 23(21), 2162–2168. 10.1016/j.cub.2013.09.005 [PubMed: 24139740]

Tamis-LeMonda CS, Bornstein MH, & Baumwell L (2001). Maternal responsiveness and children's achievement of language milestones. Child Development, 72(3), 748–767. 10.1111/1467-8624.00313 [PubMed: 11405580]

Tomasello M (1995). Joint attention as social cognition. Joint attention: Its origins and role in development, 103130.

Tomasello M, & Farrar MJ (1986). Joint attention and early language. Child Development, 1454–1463. 10.2307/1130423 [PubMed: 3802971]

Trueswell JC, Lin Y, Armstrong B, III, Cartmill EA, Goldin-Meadow S, & Gleitman LR (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. Cognition, 148, 117–135. 10.1016/j.cognition.2015.11.002 [PubMed: 26775159]

Vales C, & Smith LB (2015). Words, shape, visual search and visual working memory in 3-year-old children. Developmental Science, 18(1), 65–79. 10.1111/desc.12179 [PubMed: 24720802]

Van Egeren LA, Barratt MS, & Roach MA (2001). Mother–infant responsiveness: Timing, mutual regulation, and interactional context. Developmental Psychology, 37(5), 684 10.1037/0012-1649.37.5.684 [PubMed: 11552763]

Vida M, Maurer D (2012a). Fine-grained sensitivity to vertical differences in triadic gaze is slow to develop. Journal of Vision, 12: 634–634. 10.1167/12.9.634

Vida MD, Maurer D (2012b). The development of fine-grained sensitivity to eye contact after 6 years of age. Journal of experimental child psychology, 112: 243–256. 10.1016/j.jecp.2012.02.002 [PubMed: 22417921]

Vida MD, Maurer D (2012c). Gradual improvement in fine-grained sensitivity to triadic gaze after 6 years of age. Journal of experimental child psychology, 111: 299–318. 10.1016/j.jecp.2011.08.009 [PubMed: 21982081]

Vygotsky L (1978). Interaction between learning and development. Readings on the development of children, 23(3), 34–41.

Wass SV, Clackson K, Georgieva SD, Brightman L, Nutbrown R, & Leong V (2018). Infants' visual sustained attention is higher during joint play than solo play: is this due to increased endogenous attention control or exogenous stimulus capture?. Developmental science, e12667 10.1111/desc. 12667 [PubMed: 29624833]

Wei FYF, Wang YK, & Klausner M (2012). Rethinking college students' self-regulation and sustained attention: Does text messaging during class influence cognitive learning?. Communication Education, 61(3), 185–204. 10.1080/03634523.2012.672755

Welsh JA, Nix RL, Blair C, Bierman KL, & Nelson KE (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. Journal of Educational Psychology, 102(1), 43 10.1037/a0016738 [PubMed: 20411025]

Yoshida H, Smith LB (2008) What's in view for toddlers? Using a head camera to study visual experience. Infancy 13: 229–248. 10.1080/15250000802004437 [PubMed: 20585411]

Yu C, & Smith LB(2012). Embodied Attention and Word Learning by Toddlers. Cognition, 125(2), 244–262. 10.1016/j.cognition.2012.06.016 [PubMed: 22878116]

Yu C, & Smith LB (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PloS one, 8(11), e79659 10.1371/journal.pone.0079659 [PubMed: 24236151]

Yu C, & Smith LB (2016). The social origins of sustained attention in one-year-old human infants. Current Biology, 26(9), 1235–1240. 10.1016/j.cub.2016.03.026 [PubMed: 27133869]

Yu C, & Smith LB (2017). Multiple Sensory-Motor Pathways Lead to Coordinated Visual Attention. Cognitive Science, 41(S1), 5–31. 10.1111/cogs.12366 [PubMed: 27016038]

Yu C, Suanda SH, & Smith LB (in press). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. Developmental Science
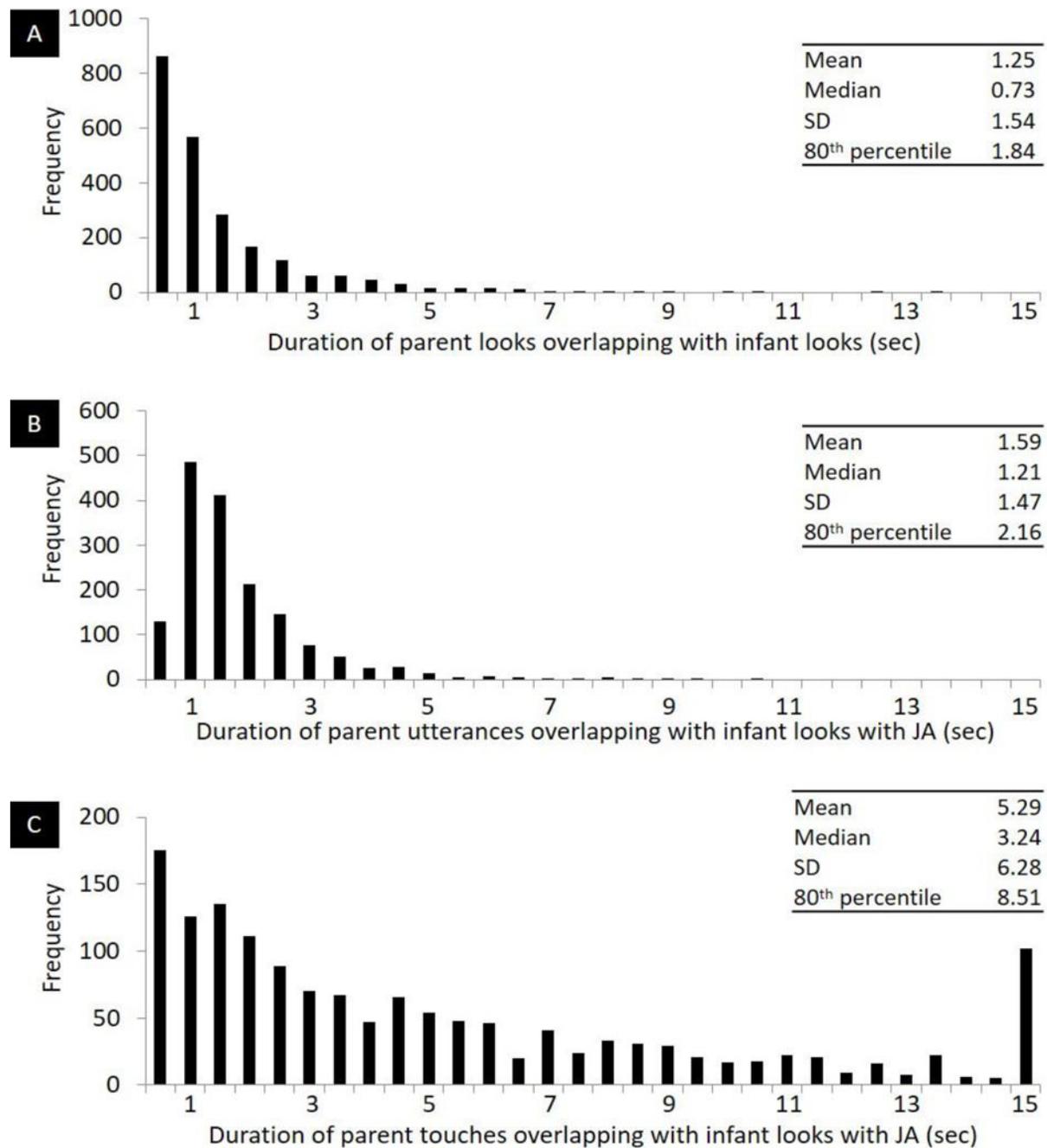
Figure 1.
Parent and infant in the tabletop with novel objects and dual head-mounted eye-tracking.
The authors received signed consent for the parent and infant's likenesses to be published in
this article.
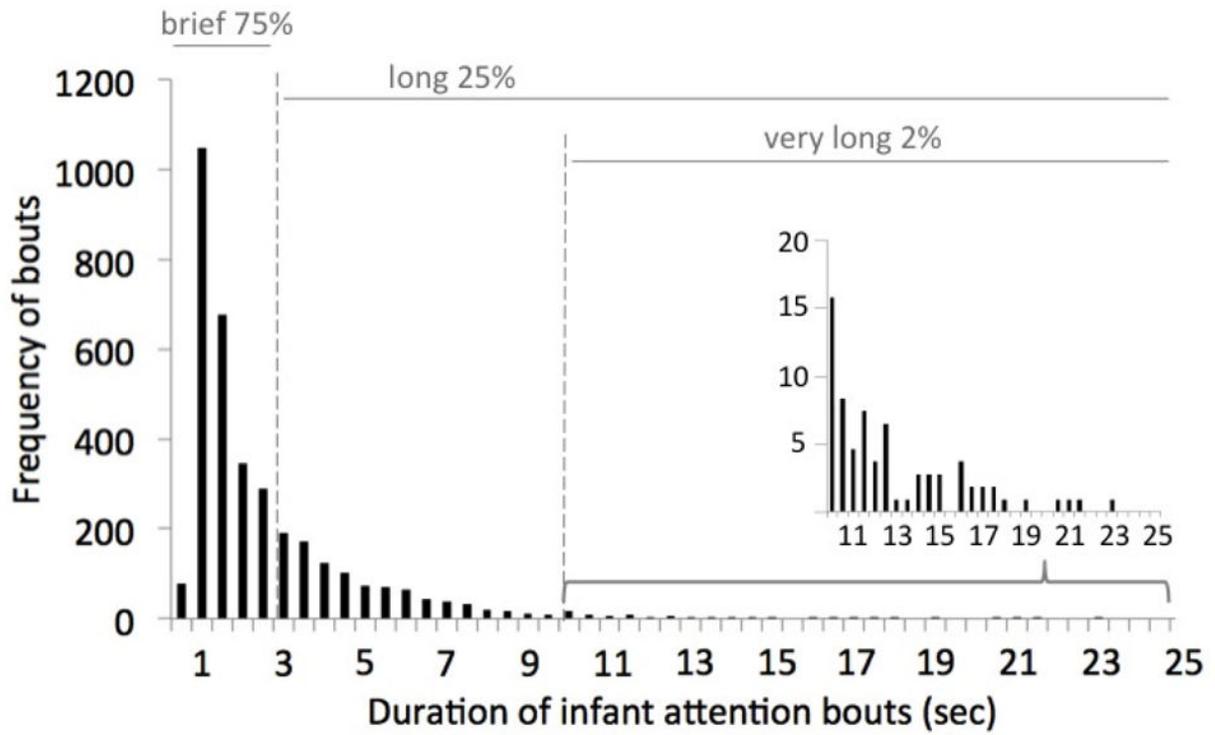
Figure 2.
Frequency distribution of the duration of infant attention bouts to objects measured in seconds. Percentage of brief, long and very long looks present in the distribution are shown.
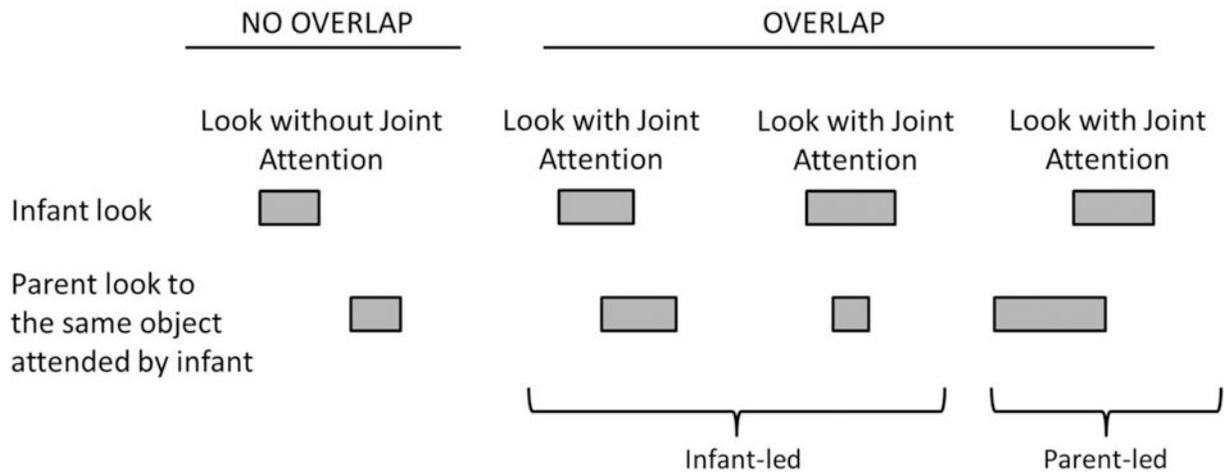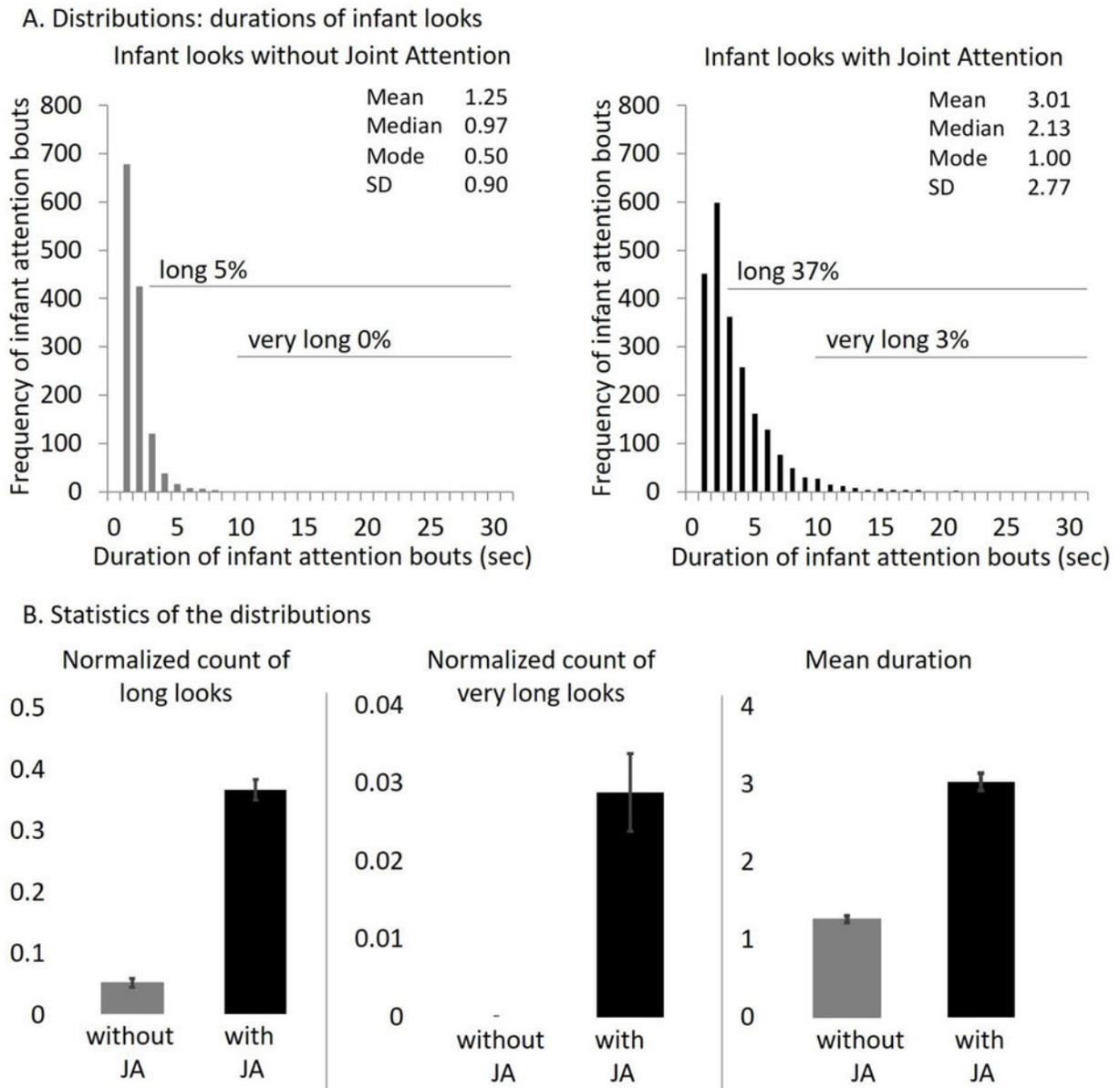
Figure 3.
Definitions for infant's look without Joint Attention and with Joint Attention (both infant-led and parent-led cases) based on overlap with a parent's look. Joint Attention was defined objectively as the temporal overlap between an infant's look at an object and the parent's look at the same infant-attended object.

A

Mean	1.25
Median	0.73
SD	1.54
80th percentile	1.84

Duration of parent looks overlapping with infant looks (sec)

B

Mean	1.59
Median	1.21
SD	1.47
80th percentile	2.16

Duration of parent utterances overlapping with infant looks with JA (sec)

C

Mean	5.29
Median	3.24
SD	6.28
80th percentile	8.51

Duration of parent touches overlapping with infant looks with JA (sec)

Figure 4.
Histograms of the duration of parent behaviors (A: parent looks, B: parent utterances, C: parent touches) overlapping with infant looks, and descriptive statistics of these distributions. Note the Y axes are on different scales reflecting the different properties of parent looks, talk and touches to objects.
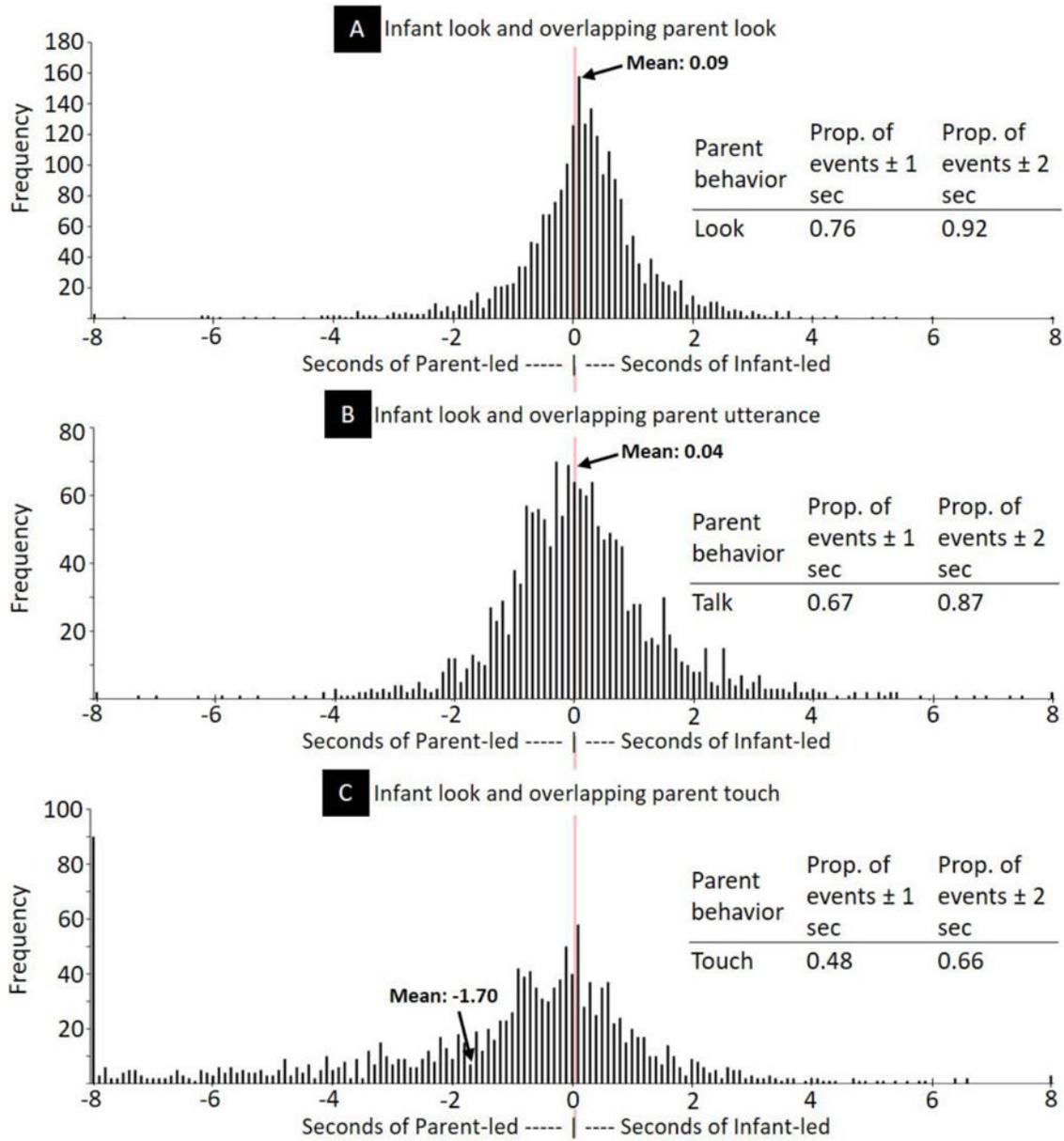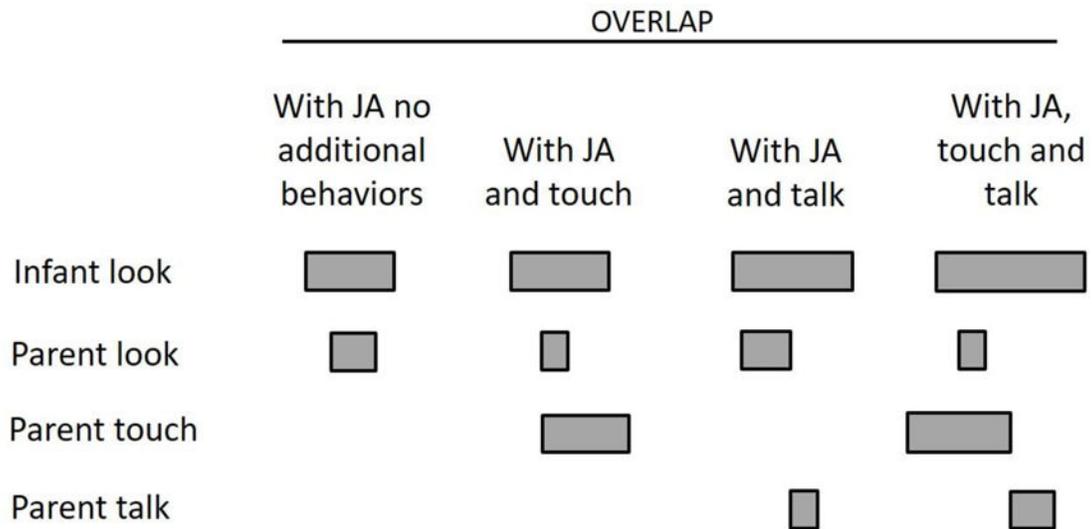
Figure 5.
Histogram showing lags in seconds between onset of first parent's look, utterance, touch (shown in A, B and C respectively) and the onset of the infant's look. The line at zero shows all panels are aligned and it represents moments in which the onsets of infant's look and the parent's behavior occurred simultaneously with lag=0. Note the Y axes are on different scales reflecting the different dynamic properties of looks, talk and touches to objects.

A. Distributions: durations of infant looks


B. Statistics of the distributions

Figure 6.
A. Histograms of duration of infant looks without and with JA. B. Three different statistics of each distribution illustrated in A are compared: Normalized count defined as the mean proportion of each infant's looks that were Long (>3sec) and Very Long (>10sec) given that the infant look did not or did include Joint Attention. We also compared the mean duration of the distributions of infant looks without and with JA computed across subjects. Error bars in this graph represent standard errors around each of the means.
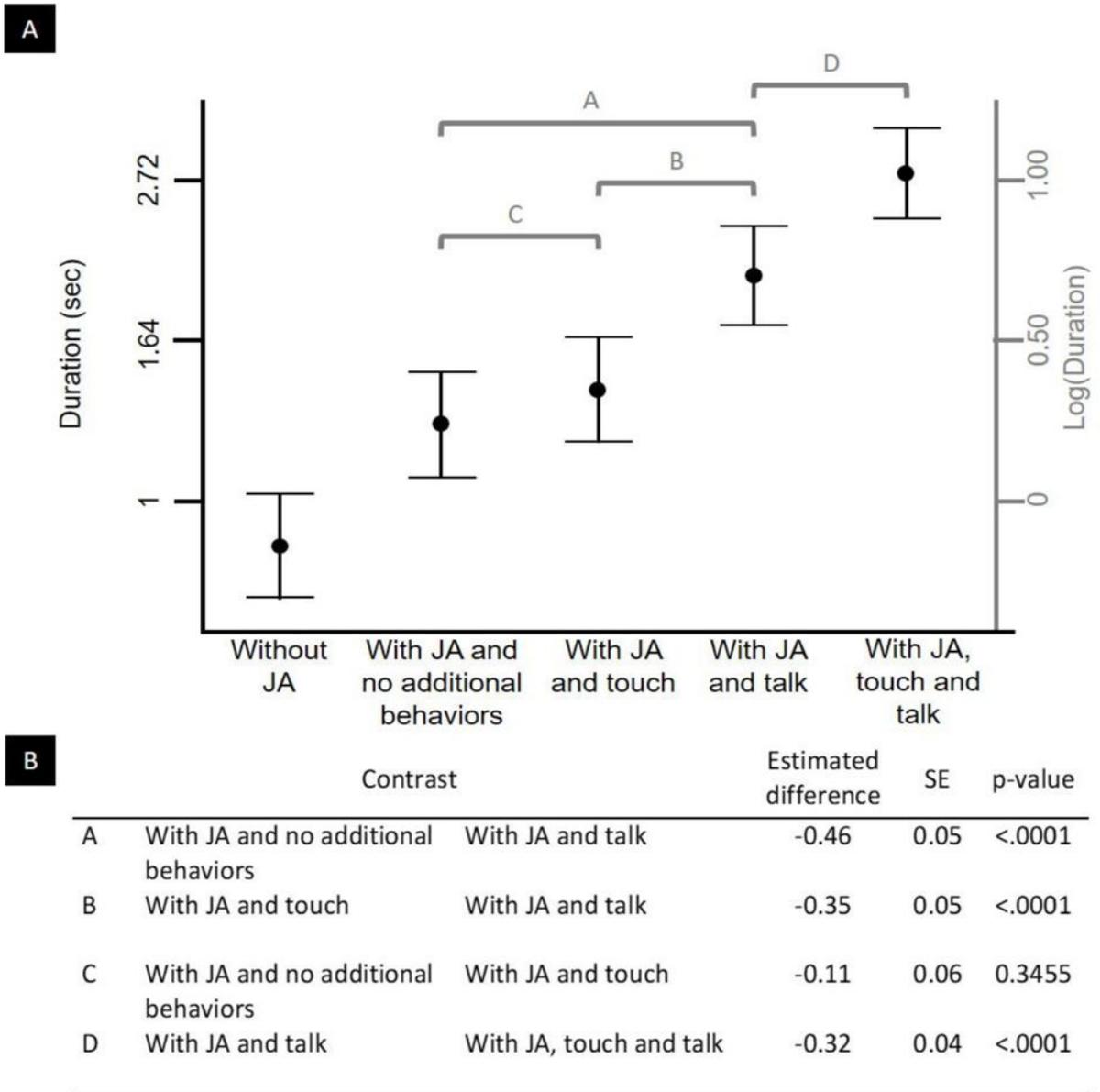
Figure 7.
Definitions for infant looks *that included Joint Attention* and different combinations of parent behaviors. Overlap with a parent behavior (touch or talk) was defined objectively as the temporal overlap between an infant look to an object and the parent behavior. The possible combinations of parent behaviors yielded four categories of infant looks that are also referred as JA categories, as they all have an overlap between the infant look and the parent look to the infant-attended object.
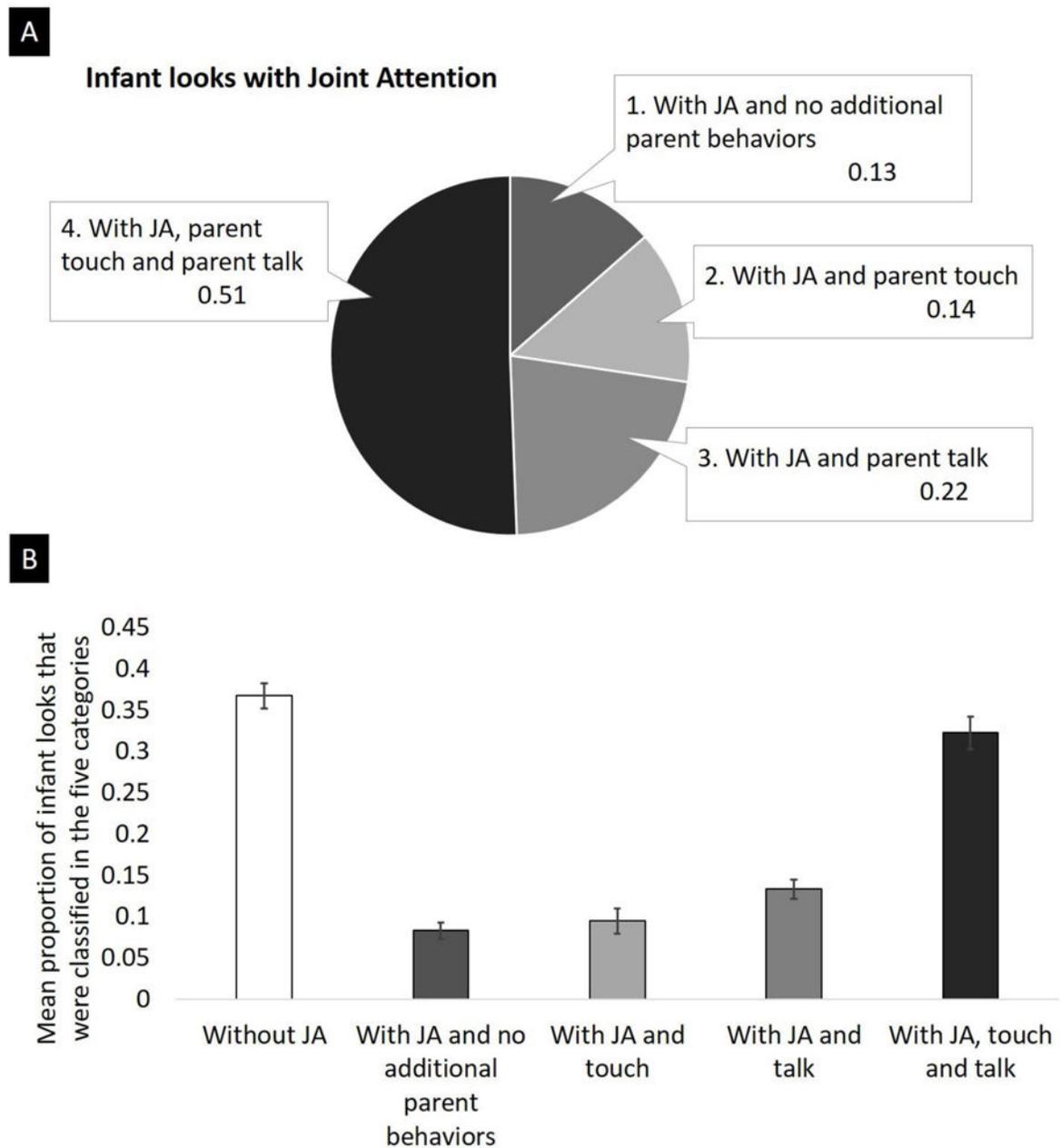
Figure 8.

A. Proportion of all *infant looks with JA* that were classified as each of the four categories of JA: JA with no additional parent behaviors, JA with parent touch, JA with parent talk, and JA with parent touch and talk. B. Mean proportion of *all infant looks* (with and without JA) that were classified as each of the five categories entered into the main analyses. The error bars in this graph represent standard errors around each of the means.

Figure 9.
A. Estimated mean log(duration) of infant looks in the five categories of infant looks with 95% confidence intervals around the estimated means. Letters A-D illustrate the contrasts tested in the model. B. Beta coefficients and standard errors of the planned contrasts A-D. The p-values are adjusted to account for all possible 5X5 pairwise comparisons according to the Tukey correction.

Table 1

Descriptive Statistics of Infant Looking Behavior to Objects and Parent Face

Infant looks	Mean prop of infant looks	SD prop of infant looks	Median prop of infant looks	Range in prop of infant looks	Mean Duration (sec)	SD Duration (sec)	Median Duration (sec)	Range in duration (sec)
At objects	0.78	0.07	0.75	0.64 – 0.97	2.40	0.51	2.26	1.60 – 3.82
Parent face	0.22	0.07	0.25	0.03 – 0.36	1.79	0.51	1.73	0.90 – 3.61

Table 2

Beta Coefficients and Standard Errors of the Linear Mixed Model

Contrast		Estimated difference	SE	p-value
2: with JA no additional behaviors	1: without JA	0.38	0.05	<.0001
3: with JA and touch	1: without JA	0.49	0.05	<.0001
4: with JA and talk	1: without JA	0.84	0.05	<.0001
5: with JA, touch and talk	1: without JA	1.16	0.04	<.0001