Title
What Leads To Shared Attention? Maternal Cues and Infant Responses During Object Play

Journal
Infancy, 23(1)

ISSN
1525-0008

Authors
Deák, Gedeon O
Krasno, Anna M
Jasso, Hector
et al.

Publication Date
2018

DOI
10.1111/infa.12204

Peer reviewed

INFANCY

THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES

WILEY Blackwell

Infancy, 1–25, 2017

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12204

What Leads To Shared Attention? Maternal Cues and Infant Responses During Object Play

Gedeon O. Deák and Anna M. Krasno
Department of Cognitive Science
University of California at San Diego

Hector Jasso
Department of Computer Science and Engineering
University of California at San Diego

Jochen Triesch
Frankfurt Institute for Advanced Studies
Goethe University at Frankfurt

Attention sharing provides an important context for infant learning, but it is not fully understood how infants respond to parents' isolated or combined actions to shift from nonsharing to attention-sharing states. To investigate this, we recorded unscripted toy-play interactions of infants (3 to 11 months old, $N = 35$) and mothers at home, and coded attention-related behaviors. These included infants' and mothers' visual fixations, and mothers' attention-directing actions including gaze shifts, pointing gestures, object manipulations, verbalizations, and object sounds. In addition, dyadic attention was continuously classified into one of seven states of shared or nonshared attention. Results showed that mothers usually produced a combination of attention-directing cues within the 7 sec. before infants shifted their attention to match the mother's focus. Mothers' cue combinations usually included object manipulation and either a gaze shift or a verbalization. Infants seldom looked at mothers' faces and followed a very small proportion of isolated gaze shifts or pointing gestures. However, infants frequently shifted attention to watch mothers manipulate objects. The results indicate that during toy play, combinations of maternal attention-specifying actions selectively elicit infants' attention following.

WHAT ACTIONS PRECEDE SHARED ATTENTION? MATERNAL CUES AND INFANT RESPONSES DURING OBJECT PLAY

By the end of their first year, most infants can monitor and follow an adult's direction of attention (Bakeman & Adamson, 1984; Butterworth & Jarrett, 1991). One-year-old

Correspondence should be sent to Gedeon O. Deák, Department of Cognitive Science, UC-San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0515. E-mail: gdeak@ucsd.edu

infants can follow an adult's gaze or pointing gestures or re-orient in response to verbal exhortations (Adamson, Bakeman, Deckner, & Romski, 2009; Baldwin, 1993; Deák, Walden, Yale Kaiser, & Lewis, 2008). These early *attention-sharing* skills might support social cognition, and facilitate communication and learning. For example, infants' attention sharing predicts later language and social skills (e.g., Dawson et al., 2004; Markus, Mundy, Morales, Delgado, & Yale, 2000).

Despite these far-reaching implications, it remains unclear how attention-sharing emerges in spontaneous infant–caregiver interactions. To address this, microbehavioral studies have documented the richly structured patterns of action coordination in infant–parent interactions (e.g., Beebe et al., 2010; Harris, Jones, & Grant, 1983; Hsu & Fogel, 2003; Tamis-LeMonda & Bornstein, 1991; Tamis-LeMonda & Bornstein 1994; Van Egeren, Barratt, & Roach, 2001; Yoshida & Smith, 2008; Yu & Smith, 2013; Zukow-Goldring & Arbib, 2007). Such studies confirm that coordinated action during attention sharing can support infant learning (e.g., Bigelow & MacLean, 2004; Rossmanith, Costall, Reichelt, López, & Reddy, 2014; Zukow-Goldring, 1996). However, many further questions remain. Here, we address one such question: In dyadic interactions, what actions compel infants to follow parents' focus of attention?

Studies of infants' attention sharing have commonly focused on *gaze following* (e.g., Moore, 2008; Shepherd, 2010; Tomasello, 1999), which is sometimes presumed to be a specialized skill (e.g., Deaner & Platt, 2003; Frischen, Bayliss, & Tipper, 2007). Indeed, there is converging evidence that some infants by 6 months can follow an adult's head turn (e.g., D'Entremont, Hains, & Muir, 1997; Tremblay & Rovira, 2007) and that gaze-following skills improve from 6 to 18 months (Butterworth & Jarrett, 1991; Deák, Flom, & Pick, 2000; Flom, Deák, Phill, & Pick, 2004). However, recent evidence suggests that gaze following might not play a pivotal role in attention sharing (Akhtar & Gernsbacher, 2008; Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; Yu & Smith, 2013). For example, although 12-month-olds follow adults' gaze in the context of relatively bare or "stripped-down" laboratory settings (e.g., Butterworth & Jarrett, 1991; Deák et al., 2000), they rarely do so in more naturalistic "cluttered" settings (Deák et al., 2008). However, even in cluttered settings infants remain attentive to adults' manual actions: pointing gestures (Deák et al., 2008) and object handling (Deák et al., 2014; Yu & Smith, 2013; Zukow-Goldring, 1996).

This last point implies that a wide range of caregiver actions—not just gaze shifts—can recruit and redirect infants' attention. The current study investigates five types of caregiver actions during home-based unscripted[1] interactions (i.e., dyadic toy play) and the occurrence of these actions before attention sharing. These actions have not all been documented in detail within a single data set. To test how they relate to attention sharing, we continuously coded infants' and parents' foci of visual attention, and their engagement in shared and nonshared attention. From this dataset, we quantified how parents' particular actions preceded shifts from nonshared to shared attention. Our goal was to determine how parents' actions, alone or in combination, preceded infants' shifts to follow their parent's attention. Based on prior studies, we chose to document mothers' gaze shifts, pointing gestures, object manipulation, verbalizations, and object sounds.

[1]By *unscripted* we mean interactions in which specific actions are not explicitly prescribed, suggested, or primed.

Pointing gestures can be defined as extending an arm and hand or finger(s) toward a target, to draw another person's attention to that target (Kita, 2003). In laboratory studies, infants by 9 months follow points more readily than gaze (Deák et al., 2000; Flom et al., 2004). For example, 1-year-olds in a visually distracting environment followed fewer than 10% of parents' gaze shifts, versus almost half of parents' pointing gestures (Deák et al., 2008). Pointing gestures might be effective for several reasons: the salient sweeping motion of the arm and hand, the purposefulness of the gesture to direct another's attention (Bangerter, 2004), and/or the relative infrequency and high cue validity of pointing. Although it is unknown how these characteristics contribute to infants' point following, we predicted that pointing gestures would compel infants to follow their parent's attention, more than the parents' gaze shifts.

Manual activity—object manipulation—also elicits infants' attention and facilitates attention sharing (Zukow-Goldring, 1996). Most simply, movement attracts the attention of animals (Abrams & Christ, 2003; Kawahara, Yanase, & Kitazaki, 2012) including human neonates (Girton, 1979; Haith, 1966). In fact, even the directional motion of adults' gaze shifts can compel infants' attention shifts, and is critical for early 'gaze-following' (Deák, 2015; Farroni, Johnson, Brockbank, & Simion, 2000; Moore, Angelopoulos, & Bennett, 1997). The movement of adults' object handling is also salient to infants (Brand, Baldwin, & Ashburn, 2002; Deák et al., 2014; Rader & Zukow-Goldring, 2010; Yu & Smith, 2013) and tends to attract their attention. It seems that infants' attention is not drawn by adults' hands *per se* (Deák et al., 2014; Frank, Vul, & Johnson, 2009), but by hands manipulating objects. In a prior analysis of the current dataset, Deák et al. (2014) found that infants ranging from 3 to 11 months of age preferred watching their mother handle objects to any other sight during toy-play interactions. This complements reports that parents modify object handling to maintain infants' interest (Brand et al., 2002; Zukow-Goldring & Arbib, 2007). Based on these findings, we expected adults' object handling to reliably elicit infants' attention following.

In addition to visual cues, parents' interactions with infants incorporate acoustic cues, notably speech. Speech can direct infants' attention (Cooper & Aslin, 1990; de León, 2000), facilitate attention sharing (Adamson et al., 2004, Flom & Pick, 2003; Tremblay & Rovira, 2007; Zukow-Goldring, 1996), and enhance learning (Tamis-LeMonda, Kuchirko, & Song, 2014). Speech effects are context- and content-specific: For example, verbal exhortations by an adult do not increase the likelihood that 12- or 18-month-olds will follow gaze or points in the laboratory (Deák et al., 2000), but can *prolong* infants' attention to target stimuli (Flom & Pick, 2003). Also, in a visually cluttered setting, mothers' verbal exhortations increase the chances that 1-year-olds will follow their gaze (Deák et al., 2008).

How do verbalizations affect younger infants' attention following? Infants tend to look at a talking adult's mouth (e.g., Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017; Hunnius & Gueze, 2005), suggesting that vocal signaling attracts infants' attention. However, in naturalistic social contexts infants show more complex responses such as looking from the adult's mouth to an object she is holding, if speech and object motion are synchronized (Gogate, Bolzani, & Betancourt, 2006; Rader & Zukow-Goldring, 2010, 2015). There are regularities of speech/action timing during parents' interactions with infants (e.g., Chang, de Barbaro, & Deák, 2016; Tamis-LeMonda, Kuchirko, & Tafuro, 2013). However, these regularities are complex: For example, within several seconds after mothers verbalize, 4-month-olds are more likely to look *either* at the object she is handling *or* at her face (Van Egeren et al., 2001), suggesting that verbalizations have diverse effects on infants' attention. However, only a few studies have documented how parents produce verbalizations in combination with other cues (e.g., pointing; gaze). These suggest that such combinations promote infants' attentive engagement (Harris et al., 1983; Zukow-Goldring, 1996). The current study builds upon those findings by detailing how verbal *and* nonverbal cues jointly influence infants' dyadic-attention states. We predicted that parents of infants younger than one year might use verbal cues less than nonverbal cues (e.g., object manipulation; pointing). We also predicted based on prior reports (Flom & Pick, 2003; Rader & Zukow-Goldring, 2010) that combinations of verbal and nonverbal (manual object handling) cues will commonly precede attention following.

In addition to speech, adults produce sounds while manipulating objects. Banging, rattling, or squeezing toys produces localized, predictable, and potentially interesting sounds. These cues differ from speech cues in several ways, despite the shared modality. For example, verbalizations are highly structured, but that structure is arbitrary to infants, with respect to meaning. By contrast, object sounds are determined by the physical properties of objects and actions and are therefore nonarbitrary and hypothetically predictable. Therefore, even 3-month-old infants have expectations about how a visible event should sound (Bahrick, 1988). Also, verbalizations emit from a speaker's mouth and therefore might draw infants' attention *away* from more tangentially located objects (Frank et al., 2009; Rader & Zukow-Goldring, 2010), whereas sounds produced during manipulation are localized at the objects. Even very young infants can localize sounds (Chun, Pawsat, & Forster, 1960; Morrongiello, Fenwick, Hillier, & Chance, 1994), so just as object movement might attract infants' attention, object sounds might also attract and re-orient infants' attention. Such differences between verbalizations and object sounds might contribute to quite different attention-eliciting properties. To test this possibility, we coded mothers' actions that produced object-sounds. A few studies indicate that such sounds that can contribute to infant-parent attention sharing (O'Neill, Bard, Linnell, & Fluck, 2005; Waxman & Spencer, 1997; Zukow-Goldring, 1997). The current study extends those findings by documenting how often object sounds accompany parents' other action cues, and whether sounds contribute to infants' attention following in the context of other cues.

Current study

This study considers how mothers produce five cues—gaze shifts, pointing gestures, object manipulation, verbalization, and object sounds—during unscripted toy play, and how often infants then follow into shared attention. Although parents may produce other attention-directing cues (e.g., facial expressions, postural changes, touch, e.g., Flom & Pick, 2005; Kaye & Fogel, 1980; Mondloch, Horner, & Mian, 2013; Stein, 2012; Zukow-Goldring, 1997), the cues investigated here include both cue types that were prominent in prior research, and less-studied by potentially robust cues for attention sharing. No previous study has investigated these five types of cues in detail. By documenting how the cues are produced, in high temporal detail and in relation to changes in the dyads' patterns of shared and nonshared attention, the dataset will provide a unique perspective on how parents' actions facilitate infants' attention following.

We recorded infant–mother interactions at participants' homes and coded participants' actions frame-by-frame (30 Hz). Microbehavioral analyses can reveal input patterns that structure behaviors and promote learning (de Barbaro, Johnson, Forster, & Deák, 2013; Fricker, Zhang, & Yu, 2011; Hutchins, 1996; Kaye & Fogel, 1980). The current project uses microbehavioral analysis to document the prevalence, patterning, and effectiveness of the five actions described above, as cues that precede attention sharing. By documenting the prevalence, patterning, and effectiveness of the five actions, we can establish both their *independent* occurrence and their *co*-occurrence (see also Zukow-Goldring, 1996).

To estimate the effect of the action cues on attention sharing, we parsed infant–mother interactions into continuous, mutually exclusive, exhaustive *dyadic-attention states*. These included three *attention-sharing* states and four *nonsharing* states. This allowed us to document when each maternal action occurred, relative to the onset of each attention-sharing state. A by-product of this effort is a database of sequential state transitions. Other studies have reported on infant–parent attention sharing and nonsharing in rich contexts (Harel, Gordon, Geva, & Feldman, 2011; Hsu & Fogel, 2003; Messer & Vietze, 1988; Tamis-LeMonda et al., 2013). The current study complements those by reporting temporally precise data from 35 healthy English-learning infants between 3 and 11 months of age: a relatively broad age range that should represent common patterns of mother–infant action, reaction, and dyadic states during the first year.

METHOD

Participants

A sample of convenience of infant–mother dyads (N = 58) in middle-class neighborhoods in San Diego county was recruited from postpartum exercise classes and via flyers or notices to local preschools and play groups and word of mouth. All mothers gave informed consent. Infants ranged from 3 to 11 months old and had no known medical problems or developmental delays. The first n = 8 dyads were pilot participants who completed somewhat different procedures; n = 4 other infants were 12 months old and are excluded because their advanced locomotor and communication abilities complicate the current analyses. Of the N = 46 remaining dyads who completed the final procedure, 11 (three females; eight males) were excluded due to recording or equipment problems (n = 4), procedural error (n = 1), or infant fussiness (n = 6).

The 35 remaining dyads included 23 infant girls. Infants averaged 6 months, 28 days (range: 3 months 10 days to 11 months 8 days). For some analyses below, this group was divided *post hoc* into three age-defined subgroups based on age-related changes in social attention reported in the literature. These included a younger group (11 infants; mean age = 4 months 5 days, range = 3 months 10 days to 4 months 24 days; six girls); a middle group (15 infants; mean = 6 months 29 days, range = 5 months 21 days to 8 months 8 days; 12 girls), and an older group (nine infants; mean = 9 months 28 days, range = 8 months 14 days to 11 months 8 days; five girls).

All caregivers were biological mothers, averaging 32.5 years of age (range = 23 to 41) and 17 years of education (range = 14 to 22). Most infants (61%) were firstborn; all but two of the rest were second-born. No infant had suffered critical perinatal medical problems. Parents identified 6% of infants as Asian, 24% as biracial, and 67% as

Caucasian (two parents did not report their infant's ethnicity). Parents reported that 27% of their infants regularly heard at least one language besides English; all infants heard at least 50% English except for one (60% Spanish exposure). Most families (74%) could be recontacted when their infant was 2 to 4 years old. None of these infants had been diagnosed in the interim with any cognitive, social, or communicative disorder.

Materials

Infant–mother dyads were video-recorded in their homes using two Canon GL-series camcorders. Two simultaneous videos of infant and caregiver were captured at 30 f/s (NTSC quality), synchronized offline, and coded using QuickTimePro and Excel on Mac G4 computers with dual 51 cm monitors.

Procedure

Two trained researchers (RES) visited participants' homes at a time of day when the caregiver had indicated that the infant was typically awake and had recently been fed. After obtaining informed consent, RES prepared a location where caregivers said that they typically interact with the infant (most often a den or playroom). Caregivers were asked to seat or recline their infant as they typically would for face-to-face floor play, using their own infant seat or bouncer.[2] Parents were asked to select several of the infant's preferred toys for the toy-play interaction. After the infant and caregiver were seated, RES began recording from locations that optimized video angles and distances. RES changed position only if the mother shifted her position so that the field of view was occluded. One RES focused on the infant (INF) and the other on the mother (MO). Camera angles were set so that the focal participant's face and upper body dominated the frame, but the other participant remained partly visible (to more easily code gaze direction or hand location). Because gaze direction is harder to code in infants than in adults, the INF camera was zoomed in closer than the MO camera. Figure 1 shows an example of synchronized frames. RESs remained as quiet and unobtrusive as possible while recording, and did not speak to MO unless she addressed them (this was rare, and all such segments were excluded from analysis; see below).

After ~7–9 min of free play, the object play period began. MOs were simply instructed to try to get INF interested in the toys, with no further specification. MOs were told that their INF might not respond, and this was normal. Object play was recorded for 5 min or until infants became fussy (mean duration = 4.5 min, SD = 1.0). Only the object play period is reported here. Note that MOs freely chose all actions: No specific actions were suggested. MOs also selected the toys, seating apparatus, and play location.

Coding

Video files were synchronized off-line. Synchronized videos were preprocessed by a RES (blind to specific hypotheses) who identified segments to be excluded due to

[2]Although infants varied widely in age and motor maturity, all were able to maintain postural stability well enough to keep the mother's head in view, once seated. Mothers of the youngest infants typically placed the infant in a semi-reclined infant seat and positioned themselves so that the infant could readily see them.

Figure 1 Example of composite video frames from a dyadic object–play interaction (synchronized screen capture from two videos).

occlusion of MO's or INF's face, MO moving herself or INF, or distracting events (e.g., occlusion by a pet). Excluded segments were verified by the first author. Less than 10% of video was excluded. Remaining video was coded frame-by-frame (30f/s) by trained coders blind to specific hypotheses. In separate coding passes, coders annotated the event categories described below (see also Table 1).

Mother's gaze direction and target

All MO fixations (i.e., intervals with no saccade for ≥0.15 sec) were coded from the frame after a saccade (onset) to the frame before the next fixation (offset). Blinks and saccades were not coded. Horizontal direction of each fixation was coded into one of

TABLE 1
Summary of Coding Categories and Values.

Participant	*Parameters*
Mother	
Gaze fixations	Horizontal direction (far left, near left, center, near right, far right)
	Vertical direction (up, center, down)
	Target (object [specified], infant [face/hand], own [hand], other)
Manual actions	Horizontal location (far left, near left, center, near right, far right)
	Vertical location (up, center, down)
	Target (object [specified])
	Extraneous movement (yes/no)
Object sounds	Quality (e.g., rattle, squeak, tap)
Pointing	Horizontal direction (far left, near left, center, near right, far right)
	Vertical direction (up, center, down)
	Target (object [x], interlocutor [face/hand], own [hand], other)
Verbalizations	Verbatim transcription
Infant	
Gaze fixations	Horizontal direction (left, center, right)
	Vertical direction (up, center, down)
	Target (object [x], mother [face/hand], own [hand], other)
Dyad	
Attention state	Four shared states (see text)
	Three nonshared states (see text)

six ±36° regions (midline/center, near right, far right, near left, far left, or back [90–270° from midline]). Vertical direction was coded as center (relative to INF's face), up, or down (±60° each). Coders were trained to recognize angles from a practice set of videos and screenshots. The target of each fixation was coded: a specific toy or object (each with a unique numeric code), own body (i.e., hand), INF's face, INF's body, or "other."

Mother's manual actions

Mother's left and right hands were coded separately. Location was coded using the same 6 × 3 grid as gaze fixations, from the frame when the hand stopped in a location for ≥0.15 sec to the frame before entering a new location. Hand status was coded as empty (no object), touching/holding object, or pointing to an object. Object identity was coded as above. Touching time was coded from the first to last frame in which at least one finger contacted the object. Extraneous movement (e.g., tapping, waving, shaking) was annotated, ignoring pauses of ≥0.5 sec (so repetitive actions like tapping received a single event code). Pointing events were coded from the frame when the forearm angle was static for ≥0.15 sec to the frame before the next hand action (forearm angle is a more stable indicator than finger angle). Pointing target (i.e., object) and location were annotated.

Object sounds

Mother actions that caused sounds loud enough to attract an interlocutor's attention (e.g., rattling, squeaking, tapping) were coded, specifying the on/offset times, hand, object, and location.

Mother's verbalizations

Verbalizations were transcribed, from the initial phone to the last phone, with gaps of ≥1 sec defining a new utterance. Off-task verbalizations (e.g., comments to RES) were excluded; these were rare and mostly occurred during excluded video segments.

Cue perceptibility

The analyses included only maternal actions that occurred in INF's visual field (e.g., not if INF was looking down or back) or were clearly audible. This criterion eliminated only <5% of potential cue actions.

Infant gaze direction and target

The direction of each INF fixation was coded similar to MO gaze, except horizontal direction was divided into only four regions (center, left, right, and back) because infant gaze is harder to resolve. Gaze targets were a specific object, INF's body, MO's face, MO's hand, MO's body, RES, or "other." A {hand+object} code was used when INF watched MO handling an object.

Dyadic-attention states

To identify periods of shared and nonshared attention, dyadic-attention status was coded into discrete mutually exclusive and exhaustive states, based on INF's gaze target and MO's attention or cue action. This method differs from previous efforts that classified dyadic attention into qualitative or highly abstract categories, often with low temporal resolution (e.g., Adamson et al., 2004; Camaioni, Aureli, Bellagamba, & Fogel, 2003). Dyadic states were coded in a separate pass. Trained coders classified each dyadic state lasting at least 0.15 sec and locked its onset to specific actions that initiated the new state (e.g., onset of an INF gaze shift to the object that MO was holding). There were four shared-attention states and three non-shared-attention states.

Non-shared-attention states

1. *INF-Looks-at-MO:* INF fixates MO's face or body while MO is looking away from INF's face.
2. *MO-Looks-at-INF:* MO fixates INF while INF is looking away from MO's face or hand.
3. *Both-Look-away:* Both MO and INF fixate different objects; neither looks at the other's face (or, for INF, MO's hand).
4. *Mutual Gaze*: INF and MO both look at each other's face.

Shared-attention states encode whether INF or MO followed into shared attention (i.e., was second to look at the target):

1. *INF-Follows-MO:* MO is fixating, manipulating, or pointing to a target; INF shifts gaze to that target. This state is the focus of most of the analyses below.
2. *MO-Follows-INF:* INF is fixating a target; MO shifts gaze to that target.
3. *Imposed Attention:* MO holds an object in front of the INF's face, close enough to fill most of INF's visual field. Zukow (1990, 1996) coded this as a type of "SHOW" action; we restricted the definition to looming an object to fill INF's visual field.

Researchers blind to the hypotheses independently re-coded 23% of quasi-randomly selected videos (stratified to accurately represent the age range). Agreement was defined as selecting the same code within 0.1 sec. Agreement coefficients (kappa, Cohen, 1960) averaged $\kappa = .75$ for MO gaze, $\kappa = .75$ for INF gaze, and $\kappa = .76$ for MO manual actions, indicating high agreement (Landis & Koch, 1977).

RESULTS

Two sets of analyses are presented. The first set describes the incidence and transitions between dyadic states, focusing on transitions to INF-Follows-MO states. This provides the context for the second set of analyses, concerning the incidence of the five MO action types (gaze shifts, points, etc.) especially preceding INF-Follows-MO states, relative to other dyadic-attention states.

Dyadic-attention states and transitions

Dyadic states: Overall distributions

Distribution of time spent in each of the seven dyadic-attention states is shown in Figure 2, for the younger, middle, and older age groups (see also Deák et al., 2014). Dyads averaged 38.1% of time in shared-attention states (*SD* = 10.8%). Most attention-sharing time occurred in INF-Follows-MO states (mean = 76.6% of shared time; *SD* = 10.5%); very little (mean = 3.8% [5.0%]) was spent in imposed-attention states (i.e., looming objects).

Nonshared states also were asymmetrically distributed. The most common (67.9% of nonshared time; *SD* = 11.1%) was MO-Looks-at-INF (INF-Looks-elsewhere). Conversely, INF-Looks-at-MO (MO-Looks-elsewhere) averaged 1.4% of total time (1.5%). Mutual Gaze subsumed 10.1% total time (7.9%).

The temporal distribution of dyadic states did not change substantially with age (Figure 2). A oneway ANOVA comparing age groups showed a marginal trend of decreasing INF-Follows-MO time, $F(2,32)=3.1$, $p = .058$. This fits prior reports that infants guide attention sharing more as they get older (Bakeman & Adamson, 1984; de Barbaro, Johnson, & Deák, 2013). However, no other state reliably differed between

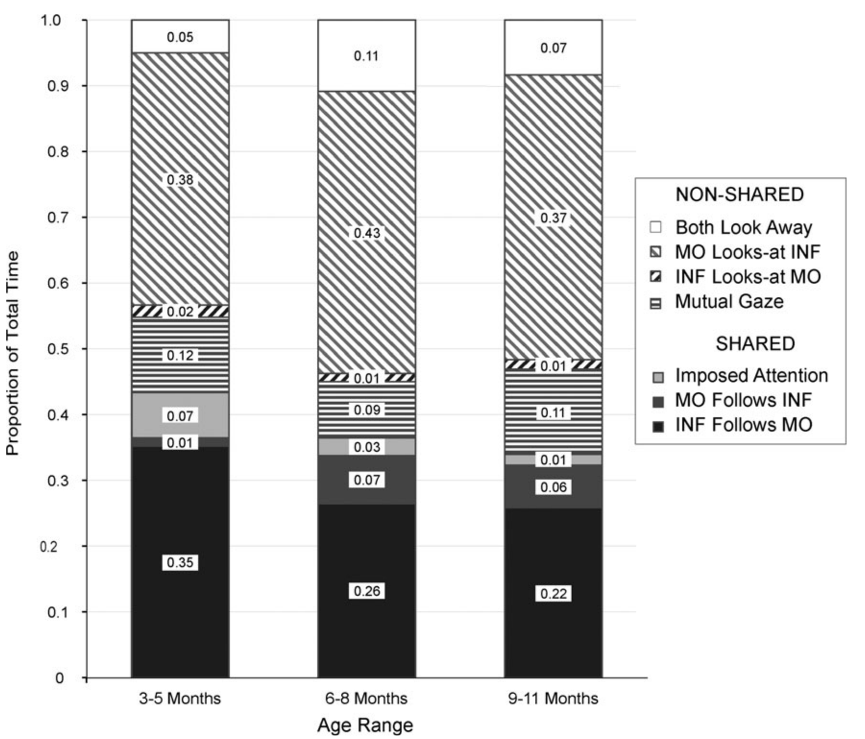

Figure 2 Proportion of time participants spent in seven dyadic-attention states (see text for descriptions), for infants in three *ad hoc* age ranges (3–5 months, 6–8 months, 9–11 months). Numbers show average proportions of total coded video when dyads were in each state. Proportions averaged over the entire sample are displayed as percentages in Figure 3, below. [Note: MO = mother; INF = infant].

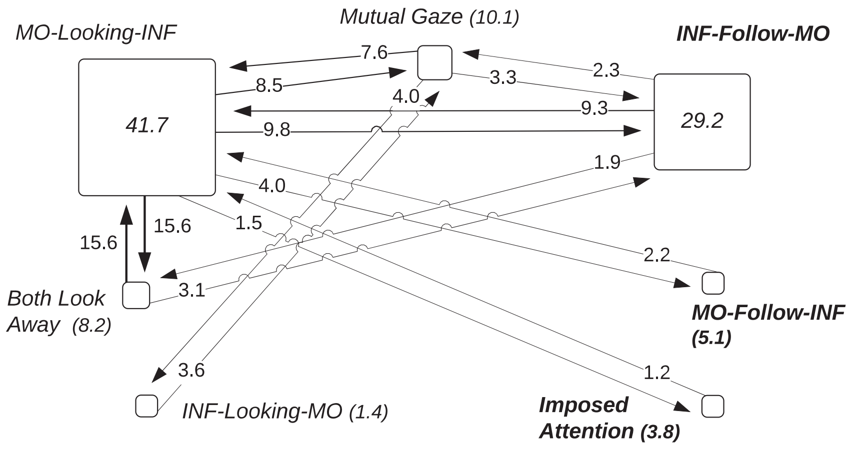

Figure 3 Transitional state-space diagram for seven dyadic-attention states, with shared-attention states labeled in bold, and on the right side. Box sizes are proportional to mean percentage of total time in that state (percentage either in the box or in parentheses in box label). Numbers on arrows are the percentage of that transition out of all transitions. Rare transitions (<1% of total) are omitted. See text.

age groups, in either frequency or proportion of time. (Note: age differences were not compared in the three least prevalent dyadic states, because such analyses would have violated statistical assumptions.)

Figure 3 shows the state-transition frequency space involving all (nearly 5,000) dyadic states, illustrating the distribution of attention transitions, including states that preceded INF-Follows-MO. Only one-back transitions are shown (for simplicity, because there are at least 165 possible two-back transitions). The size of each box is proportional to the mean prevalence of that state. Numbers on each arrow indicate the percentage of transitions due to that transition. Rare transitions (<1% of total) are omitted for readability.

Transitions to attention following

The state most commonly preceding infants' attention following, MO-Looks-at-INF, was also most prevalent state overall, suggesting that INF attention following did not emerge from specialized dyadic states. Table 2 shows conditional probabilities (column 2) of INF-Follows-MO following each possible prior nonsharing state (column 1), relative to the total frequency base rate of that prior state. For purposes of comparison, column 3 shows the conditional probabilities that a MO-Follows-INF state (i.e., the other main shared-attention state) followed each prior nonsharing state, also relative to the latter's base rate.

Notably, INF-Follows-Mo succeeded MO-Looks-at-INF states 160% as often as expected (p = .008, two-tailed one-sample z-test of [normalized observed %]/[expected %]). Relatedly, MO-Follows-INF states succeeded MO-Looks-at-INF states 240% as often as expected ($p < .0001$). Thus, both attention-sharing states were disproportionately likely to occur right after MO looked at INF while INF was looking at something other than MO. This suggests that MOs monitored INFs during periods of nonshared attention. Consistent with this interpretation, the only other state that

TABLE 2
Normalized Frequencies of Attention-Following States After Four Non-Following States, Relative to Base-Rates of Each State

1. Prior Nonshared State	2. Relative Frequencies of INF-Follows-MO Proportional to Base Rate of Prior State	3. Relative Frequencies of MO-Follows-INF, Proportional to Base Rate of Prior State
Mutual Gaze	1.30	.10
MO-Looks-at-INF	1.59	2.40
INF-Looks-at-MO	.70	NA
Both-Look-Away	.92	.43

Relative frequencies of INF-Follows-MO states (column 2) directly following each type of nonsharing state (column 1), as proportions of the frequency basis of the latter. For example, if the prior state accounted for 10% of all states, but occurred immediately before 20% of INF-Follows-MO states, the relative frequency would be 2.0. Column 3 shows analogous relative frequencies of MO-Follows-INF states following each non-sharing state, also relative to the frequency of the latter. Note that it was virtually impossible for two attention-sharing states to occur in immediate succession.

disproportionately preceded INF-Follows-MO states was Mutual Gaze (130% as often as expected; $p = .018$). This is presumably because in some Mutual Gaze states, MO was handling a toy to get INF's attention while monitoring INF's response (Deák et al., 2014; also see below). These patterns of dyadic-attention transitional probabilities indicate that MOs monitored INF's attention to promote attention sharing.

These patterns contextualize the results below that describe MOs' actions prior to INF attention following. For a more complete perspective, proportions of all transitions to and from attention-sharing states (both INF-Follows-MO and MO-Follows-INF) are shown in Figure S1. For comparison, analogous transition proportions to and from each non-attention-sharing state are shown in Figure S2. These patterns merit future study, but for the present, they illustrate the importance of assessing dyadic-attention states to contextualize the effects of an adult's actions on infants' attention.

Caregiver cues and combinations

To determine which MO cues preceded INF-Follow-MO states, we first examined the frequency of maternal gaze shifts, pointing, object manipulation, verbalizations, and object sounds. Table 2 shows the rate of each cue, and its correlation with INF's age. These statistics do not consider whether or not the INF followed, but rather indicate the overall availability of various cues.

Three cue types—gaze shifts, object manipulation, and verbalizations—were more frequent, and thus provided more opportunities to influence INFs' attention. Pointing gestures were relatively rare (eleven MOs pointed fewer than six times), as were object sounds (seven MOs produced fewer than six). Cue rates did not reliably correlate with infant age (Table 2, right column). This was confirmed by one-way ANOVAs comparing cue rates across the *post hoc* age groups: Only one marginal effect was found: a trend of more pointing by MOs of middle (6–8 months) than younger (3–5 months) infants (means: 1.3 versus 0.5 per min, *SD*s = 0.8 and 0.3), $F(2, 32) = 2.74$ ($p = .080$).

Cue rates varied across individual MOs, as shown in Table 3. This might reflect different habits or preferences for different cues, or more general differences in parenting

TABLE 3
Frequencies of Maternal Cues and Correlations With Infant's Age

MO Action	*Times/Min (SD)*	*Range (Times/Min)*	*Correlation (r) With Infant Age*
Gaze shift	5.1 (2.1)	1.7 to 9.7	.021
Pointing	1.9 (1.2)	0.0 to 3.9	.083
Object manipulation	3.7 (2.0)	0.7 to 10.4	−.045
Verbalization (n = 34)	4.8 (1.6)	1.5 to 8.8	.064
Sound	2.4 (1.6)	0.0 to 6.7	−.169

Rates of MO attention-indicating cues (mean, *SD*, range), and Pearson's *r* between rates and INF age (right). Verbal data exclude one MO who spoke Hindi to her 11-month-old infant.

style or in traits like motivation or persistence. To explore whether individual differences were cue specific or were more general (e.g., persistence), Table 4 (lower right quadrant) presents correlations among MOs' rates for different cues. Individual differences in a general trait such as persistence would be reflected in uniformly high positive correlations in production rates of different cues. However, only two pairs of cue rates were significantly correlated: gaze and pointing, $r(34) = 0.60$, $p < .001$, and gaze and object manipulation, $r(34) = 0.41$, $t(33) = 2.56$, $p = .015$. No other cues were correlated, suggesting that general traits such as persistence or motivation did not determine MOs' production of cuing actions. Although the data are consistent with the possibility that individual MOs are moderately consistent in how often they produce some common cues, an alternative possibility is that the correlation between gaze and object-manipulation rates is driven by the latter, because adults tend to watch their hands while manipulating objects (Land, Mennie, & Rusted, 1999).

Maternal cues preceding shared attention

Cue distributions

The next analyses consider how often the five cues, singly or combined, preceded INF attention following. For this, we identified all cues produced in the 7 sec. before each INF-Follows-MO state.[3] Production rates varied considerably across cue types. Object manipulation preceded 79.0% of INF-Follows-MO states. Gaze shifts, though more frequent (Table 3), preceded only 50.7% of INF-Follows-MO states. Although it might seem odd that MOs did not always look at the target before attention sharing, recall that MOs often looked at INF's face while manipulating a toy to induce following (Deák et al., 2014). Verbalizations also preceded more than 40% of INF-Follows states. Pointing gestures and object sounds were less common, preceding 19.8% and 6.0% of INF-Follows states, respectively.

Although different cues occurred and preceded attention sharing at very different rates, these data are complicated by the fact that more than one cue could be produced in the seconds before attention following. Thus, cue types might not be

[3]The 7-sec. criterion was chosen because almost all cues occurred within 7 sec. of shared-attention state. Neither an 8-sec window nor a shorter (5 sec) window reliably changes the pattern of results. Moreover, because infants in laboratory studies sometimes take up to ~7 sec to respond to a gaze or point cue (Flom et al., 2004), a 7-sec window is likely to maximize "hits" without introducing false alarms.

TABLE 4
Correlations Among Three Dyadic Social-Attention States, and Among Rates of Three Most Common Maternal Attention-Cues

CUE	*% MO-Follow*	*% Mutual Gaze*	*MO Gaze Rate*	*MO Point Rate*	*MO Manip. Rate*	*MO Verbal Rate*
% INF-follows time	−.34^{+}	.09	−.13	−.05	−.18	−.19
% MO-follows time		−.29	−.02	−.05	.29	.27
% Mutual gaze time			−.01	−.10	−.54*	−.19
MO gaze rate				.60*	.41^{+}	−.10
MO point rate					.17	.18
MO manipulation rate						−.08

Correlations among three dyadic states (% time in INF-Follows-MO, MO-Follows-INF, and Mutual Gaze), and rates of MO cues (gaze shifts, points, object manipulation, verbalization). $*p < .01$; $^{+}p = .025$; a critical $\alpha < .01$ threshold was adopted to control type I error across multiple tests.

independent (e.g., Zukow-Goldring, 1996). We therefore examined the occurrence of cue combinations in the 7 sec before INF-Follows-MO states. These rates are shown in Figure 4, for each cue, at three levels of cue "density": the cue alone before INF-Follows-MO (darkest bar for each cue); the cue plus one other cue type (middle bar); and the cue plus two or three other cues (lightest bar). Note that across cue types the latter combined-cue bars will include nonexclusive or overlapping events that contribute to at least one other combined-cue bar for another cue type.

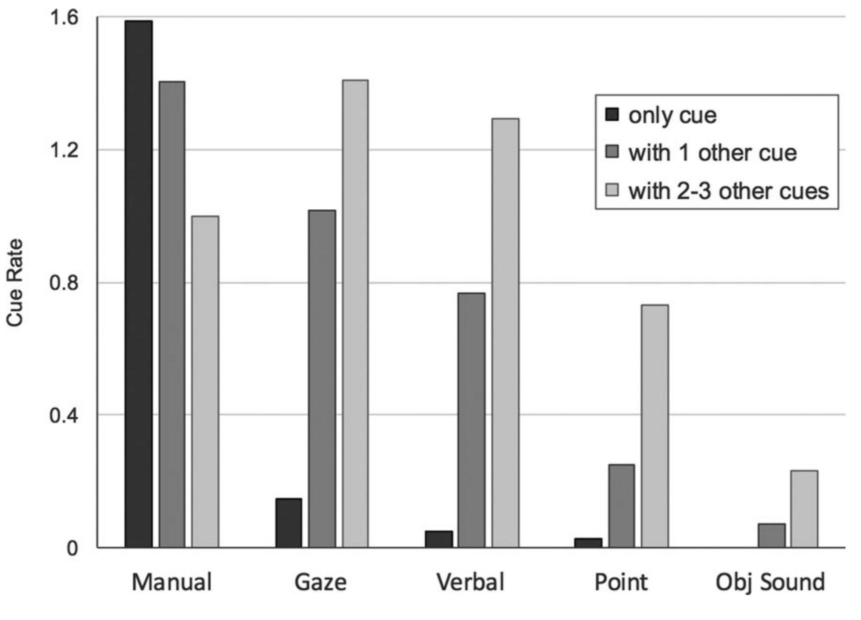

Figure 4 Maternal cue distributions before INF-Follows-MO states, for the five cue types. For each cue type, the three bars show (from darker to lighter) the rate of the cue alone, the cue occurring with one other cue, and the cue occurring with two or three other cues. Thus, the lighter bars across cue-types include non-exclusive events (i.e., that contribute to more than one bar).

The results reveal that INF attention following often was preceded by multiple cues. Single cues preceded only 35.9% of INF-Follows-MO events. This proportion did not significantly covary with infant age (r = .186; *ns*). When a single cue preceded INF following, it was usually object manipulation (mean: 87.6%). By contrast, gaze and verbal cues accounted for (respectively) only 8.1% and 2.8% of single-cue events, suggesting that those cues, though common, preceded attention following only in combination with other cues. No specific individual-cue rate correlated significantly with infant age.

Table 5 reveals that two-cue combinations preceded an average of 34.7% of INF-Follows-MO events (see Figure 4). Object manipulation was the most common participating cue (76.3% of cue pairs), but more than half of cue pairs (57.5%) were preceded by a gaze cue. Thus, the most common cue pairs overall were Gaze + Manipulation (41.6% of pairs) and Verbalization + Manipulation (33.6%). Rate of total two-cue combinations was not reliably related to INF age, r = 0.155.

Finally, an average of 29.4% of INF-Follows-MO states were preceded by three or four different cue types (Figure 4). Most of these combinations included a gaze cue (mean: 94.8%), verbalization (87.1%), and object manipulation (62.7%). The rate of three to four cue combinations was not correlated with INF age, r = 0.164.

Relation to attention sharing

To evaluate whether individual differences in MOs' cue rates affected attention sharing, we examined correlations between MO's cue rates and proportion of time that the dyad spent in INF-Follow-MO states. Correlations among the three most common cues are shown in Table 4 (top right). These weak and nonsignificant associations suggest that individual MOs' cue production frequency did not predicted the frequency of attention following. Another possibility, however, is that certain cue *combinations* predicted INF attention following. To test this, INF-Follows-MO time was correlated with rates of cue combination that included object manipulation. The correlation, controlling for INF age, was not significant: r_{part} = 0.27, p = .122.

Other associations among dyadic states and cues

Table 4 also shows (top left) correlations between MO cue production and the proportion of time in two other states: MO-Follows-INF shared attention, and Mutual

TABLE 5
Rates and Proportions of Different Cue Pairs

CUE	*Gaze*	*Point*	*Manip.*	*Verbal*
Gaze		0.17/min	0.73	0.12
Point	9.5%		0.02	0.06
Object manipulation	41.6%	1.1%		0.59
Verbalization	6.6%	3.7%	33.6%	

Cue composition of cue pairs produced before INF-Follows-MO states. Top/right values: pair rates per min; bottom/left: percentages of each pair out of all cue pairs. (Note: values do not sum to 100% because pairs with an object sound cue [7.0% of total] are not shown.)

Gaze, which has been claimed to facilitate attention sharing (e.g., Farroni, Csibra, Simion, & Johnson, 2002; Tomasello, 1999). Neither state was reliably associated with MOs' rate of cue combination.

Finally, Table 4 reveals an unpredicted but interpretable negative correlation between object manipulation and Mutual Gaze time: MOs who spent more time handling objects to get INF's attention also spent less time spent in mutual gaze with their INF.

DISCUSSION

This study detailed mothers' actions that could attract infants' attention, and their 3- to 11-month-old infants' subsequent attention following. Five maternal actions were coded from unscripted toy-play videos: gaze shifts, pointing gestures, object manipulation, verbalizations, and object sounds. In order to assess how these action cues facilitated attention sharing, dyadic-attention states were continuously coded, and changes in these states were related to the type and timing of the mothers' actions.

The record of dyadic-attention states showed that infants followed mothers' attention more than the reverse. Infants typically redirected their attention to an object cued by the mother when the mother was looking at the infant, and manipulating a toy to draw the infant's attention even as she monitored the infant's response. From these states, infants frequently shifted attention to the object the mother was handling, thereby entering into an attention-sharing state. Note that by our definition, attention-sharing subsumed states in which the mother was looking at the infant but moving a toy to attract the infant's attention, and the infant was looking at that toy. By recognizing object manipulation in the toy-play context as a bid for infants' attention, we found that this cue was most likely to precede infants' attention following, although it was not the most frequent cue action overall. Moreover, it was the cue most likely to co-occur with another cue—usually gaze or verbalization—in the seconds before attention-following. By comparison, although mothers produced many potentially informative gaze shifts, infants almost never followed gaze shifts in the absence of additional cues.

Extensions of previous findings

These results confirm other recent evidence that toddlers pay little attention to caregivers' gaze during object play. The current data neatly confirm the results reported by Yu and Smith (2013), including details such as the distribution of infants' gaze and the frequency of attention sharing. This is especially noteworthy because that study and this one tested nonoverlapping age groups. Given the differences between studies in infant age as well as methods and settings (e.g., home versus laboratory), the similarity of results is striking. Other studies have also reported that infants rarely fixate on adult faces or follow gaze during social interactions (Franchak, Kretch, Soska, & Adolph, 2011; Yoshida & Smith, 2008). For example, Yoshida and Smith (2008) reported that 18-month-olds' attention was attracted by parents' object handling but not their faces. The current results (and those in Deák et al., 2014) confirm these results and extend them to a younger age range, including infants as young as 3 to 5 months. It is striking that no infant in our study preferred looking at their mother's face.

The finding that infants seldom follow adults' gaze during unscripted play would seem to undermine or qualify claims by Csibra (2010), among others, that infants' "preferential attention to faces" (p. 146) is reflected in their enjoyment of eye contact. The weight of evidence from naturalistic studies fails to support the assumption that infants particularly prefer to look at parent's faces. It is not that infants were disengaged from their mothers, as shown by their persistent attention to her object-handling actions. But infants seldom engaged in mutual gaze, even though mothers' persistent attention to the infant's face (Deák et al., 2014) made mutual gaze possible at almost any time. Although mothers did sometimes glance away from the infant to look at an object, the fact that infants almost never followed these cues *or* took advantage of mutual gaze opportunities suggests that they were not especially motivated to use their mother's face or eyes to guide their attention.

The current results contextualize our finding (Deák et al., 2014) that during object play infants aged 3 to 11 months preferred to look at handled or static objects more than faces, whereas mothers monitored their infant and only occasionally glanced at objects. Infants nevertheless get enough valid sequential information to slowly learn gaze-following responses, because when infants happened to look at their mother's face, if she was *not* looking at them she was usually looking at an object she was holding, and infants tended to look at that object with their next fixation. This pattern could serve as a sparse but reliable teaching signal (Deák et al., 2014). The current results looked at a wider range of maternal actions. The analyses confirm that (1) maternal object handling, more than other actions, account for infants' following; and (2) object handling often occurs in conjunction with other actions (e.g., speech) that support attention following.

Other implications: Maternal cues and infant following

When a single cue preceded infants' attention following, it was almost always (88% of instances) object manipulation. However, mothers' object handling during the play episodes was distinct from adult manual activity in other contexts (e.g., Hayhoe & Ballard, 2005). Here the objects were toys and mothers' actions had no purpose except to engage the infant. The quality of their actions tended to reflect this purpose (e.g., shaking or waving the toys). This is consistent with evidence that parents modify their actions to orient or direct infants (e.g., Brand et al., 2002; Matatyaho & Gogate, 2008; Nagai & Rohlfing 2008).

The results also speak to the role of pointing gestures in infant–parent communication (e.g., Tomasello, Carpenter, & Liszkowski, 2007). There has been limited evidence concerning how, and when, caregivers point for infants during unscripted interactions, and how reliably infants follow such points. In the current context mothers seldom pointed, possibly due in part to the short distances separating mothers, infants and toys, and in part to the age range (i.e., some infants were too young to follow pointing; see Flom et al., 2004). Future studies could investigate how and when caregivers point, and whether infants of different ages respond to those points, given various ecological conditions and task demands.

Object sounds during play might attract infants' attention (O'Neill et al., 2005; Zukow-Goldring, 1997): Many infant toys are designed to produce salient sounds (e.g., squeaking, rattling), but any object can be manipulated to produce a sound (e.g., paper can crinkle, pens can tap, keys can jingle). However, our results suggest that objects sounds contributed only weakly and incrementally to infants' attention shifts. Perhaps in other contexts (e.g., with harder-to-find objects or extreme visual clutter) or in special populations (e.g., infants with visual deficits), object sounds will be found to play a larger role.

In the current results, mothers' gaze shifts preceded attention sharing only if the shifts occurred in conjunction with other cues, especially object manipulation and verbalization (see also Tamis-LeMonda et al., 2013). This suggests that the teaching signal for gaze following that was identified by Deák et al. (2014) might be enhanced by redundant cues—notably, manual actions and verbalizations. More broadly, the results show that *redundancy* of caregiver actions facilitated attention sharing. Most INF-Follows-MO episodes (64%) were preceded by two or more cues (most commonly object manipulation, verbalization, and/or gaze), suggesting that redundant cues often precede infants' attention following. The fact that these combinations included gaze shifts is not surprising because adults shift gaze frequently (see Table 3), and tend to look at what they are handling or speaking about (e.g., Griffin & Bock, 2000; Hayhoe & Ballard, 2005). However, there was no association between mothers' gaze shifts and infants' attention following ($r = -.02$), further suggesting that gaze cues played at most a secondary role in attention following. Regardless, the results corroborate reports that cue combinations are the norm in mother–infant attention-sharing interactions (Zukow-Goldring, 1996).

Although we did not assess what infants learned during these interactions, the results have two possible implications for research on infant learning. First, adults' object handling is a salient event that might help infants resolve uncertainty about ambiguous verbal cues. Converging evidence suggests that caregivers modify their speech in response to infants' exploratory actions (e.g., Chang et al., 2016; Tamis-LeMonda et al., 2013; Zukow-Goldring, 1996; Zukow-Goldring & Arbib, 2007). Caregivers also modify object handling to fit their utterances, and this can facilitate infants' learning (Gogate et al., 2006; Rader & Zukow-Goldring, 2010, 2015). Thus, caregivers shape their actions and speech to help infants learn speech–referent associations. Second, infants are sensitive to coordination of gaze and manual gestures (Amano, Kezuka, & Yamamoto, 2004; Wu & Kirkham, 2010), and so this cue combination might also facilitate learning. These results thereby suggest that redundant social cues might affect not only infants' attention but also learning (see also Jasso, Triesch, Lewis, & Deák, 2012).

Dyadic states leading to attention sharing

Attention following was more likely to follow some nonsharing states than others. Most attention sharing followed a state wherein mothers watched the infant looking at some object. However, the results do not support a hypothesis (Farroni et al., 2002; Tomasello, 1999) that mutual gaze facilitates attention sharing. Although mutual gaze preceded INF-Follows-MO states fairly often relative to its base rate, in general attention sharing did not usually follow mutual gaze (Figure 3). Also, there was no correlation between dyads' mutual gaze time and their attention-sharing time (Table 4). Finally, it might be argued that our fine-grained coding protocol obscured a sequential relation from mutual gaze to attention sharing, because an intermediate state might be necessary. However, a post hoc analysis of two-back sequences disconfirms this possibility: that is, the two-step sequence that would lead to attention sharing is $[t-2]$ =

mutual gaze; [$t-1$] = mother-looks-away, [t] = infant follows gaze. However, that sequence accounted for only 0.7% of all two-step state transitions, which is not more than expected by chance. Thus, we find no evidence that mutual gaze promoted attention sharing.

The results also address long-standing questions about how caregivers lead or follow into attention sharing. It has been argued that *following* infants into shared attention might have a different social-communicative impact than *leading* infants' attention (Akhtar, Dunham, & Dunham, 1991; Mundy et al., 2007; Tomasello & Farrar, 1986). However, evidence for this claim has been limited. In the current results, mothers' tendency to follow infants varied (range: 0 to 20% of time, $SD = 5.3$), and was marginally negatively correlated with infants' tendency to follow mother's attention (Table 4), consistent with the hypothesis that individual caregivers vary in their tendency to lead or follow infants' attention. However, we found no reliable relations between this variable and other caregiver attention-cuing behaviors. Thus, our data do not support any theories about the importance of this individual difference.

Our results confirmed an infrequent form of caregiver-driven shared attention (Yoshida & Smith, 2008; Zukow-Goldring, 1996). In *imposed-attention* events (mean=3.4% of time), mothers loomed a toy to fill the infant's visual field. Zukow-Goldring (1996) described a larger category of "showing" actions, but did not analyze looming actions separately. Thus, the current data corroborate and extend the evidence that caregivers occasionally direct infants' attention by fiat, essentially filling their visual field with an object. Notably, the size of an object in toddlers' visual field predicts their memory for the object (Yu & Smith, 2012), so this action, though rare, might be relatively effective. Also, mothers imposed attention more with younger infants than older infants (correlation with age: $r = -0.37$, $p = .03$), consistent with findings that parents increasingly let older infants guide dyadic activity (Bakeman & Adamson, 1984; de Barbaro, Johnson, & Deák, 2013; Zukow-Goldring, 1996). However, mothers' tendency to impose attention was not associated with other maternal behaviors.

The results revealed several age effects. The proportion of time spent in shared attention after infants followed mothers' cues decreased with age, as did imposed-attention time. These trajectories are consistent with previously reported age differences (e.g., de Barbaro, Johnson, & Deák, 2013). In addition, there was a trend for mothers to point marginally more for infants aged 6 to 8 months than for younger infants. This is consistent with the fact that younger infants do not follow pointing gestures (Flom et al., 2004). Considering the profound changes in motor abilities between 3 and 11 months, it is likely that further analysis of infants' responses would reveal additional age-related changes. It is possible that we did not find more age differences for three reasons: First, most of the maternal cues studied in this dyadic context are easy for even 3- to 5-month-old infants to detect. Second, infants' responses, as coded here, did not require very mature or nuanced motor skill. A third possible reason is that mothers were asked to attract their infants' attention, and they tailored their actions appropriately—in other words, mothers worked to minimize age differences. In any case, the results indicate fairly stable tendencies by mothers in this population to produce certain cues, for infants spanning a fairly wide age range that is accompanied by considerable developmental change.

Limitations and future directions

The current analyses are limited in several regards, suggesting directions for future research. One limitation is that we did not code the details of cue execution—for example, specific actions during object manipulation, or hand shape during pointing. Such variables might mediate the results, but because there was considerable variability in cue execution, dividing the cues into subclasses would have reduced the statistical power of our analyses. We are, however, currently analyzing the content of maternal utterances from a much larger longitudinal dataset (Chang et al., 2016). That project and others (e.g., Wu & Gros-Louis, 2014) directly consider how specific verbal content relates to other events in infant–parent attention sharing.

Another limitation is that our mothers were middle-class, relatively old and well-educated North American English speakers. The results might not generalize to dyads from other cultures: Although many attention-sharing patterns are similar across cultural groups (e.g., Bornstein, Toda, Azuma, Tamis-LeMonda, & Ogino, 1990; Tamis-LeMonda et al., 2013), some specific behaviors differ across cultures (e.g., Bornstein et al., 1990; de León, 2000; Tamis-LeMonda et al., 2013) or across families with different risk factors. Similarly, we cannot assume that the patterns would be similar with other types of caregivers such as fathers or unrelated adults (e.g., Colonnesi, Zijlstra, van der Zande, & Bögels, 2012). Finally, the results were collected during in-home, dyadic toy-play interactions, and might not generalize to other interactions or settings (see Gros-Louis, West, & King, 2016; Nomikou & Rohlfing, 2011; O'Neill et al., 2005; Rossmanith et al., 2014; for examples).

A third limitation is that in order to collect dense observations, we asked mothers to get their infants interested in toys. This yielded comparable data across dyads, but might have influenced mothers' behaviors, for example by encouraging them to produce more cues than normal. It is therefore prudent to treat the reported cue rates as tentative, and possibly inflated. However, comparing our results to those from other reports of unscripted infant–parent interactions, our observed cue rates do not seem greatly inflated. For example, Yu and Smith (2013), in a laboratory setting, found that attention sharing subsumed 33% of interaction time between 1-year-olds and mothers, comparable to our mean of 38%.

A final caveat is that although we coded a wider range of caregiver cues than previous studies, we did not consider all caregiver actions that might initiate attention sharing. Other possibly effective cue actions include facial expressions (Flom & Pick, 2005) and infant touch (Stack & Arnold, 1998; Stein, 2012; Zukow-Goldring, 1996). Similarly, we coded a limited subset of infant behaviors. Coding a wider range of mother than infant actions might have made mothers appear to have a larger influence on dyadic attention sharing. Currently, we are conducting more detailed analyses of infant as well as caregiver actions from dyadic at-home toy-play sessions recorded for a longitudinal study (see de Barbaro, Johnson, & Deák, 2013). Those data will more fully address how both infants' and parents' actions affect attention sharing.

These results inform a growing literature that is providing a rich account of infant–parent interactions. Franchak et al. (2011) and Yu and Smith (2013) have shown the benefits of head-mounted eye trackers for estimating infant and parent visual attention in unscripted social interactions—a method that might be used to validate some of the current results. Future studies could utilize motion- and gaze-tracking technology (e.g., Essig et al., 2012) to investigate the fine-grained dynamics of infants' and caregivers' actions, including object use (e.g., Bambach, Franchak, Crandall, & Yu, 2014). Most generally, the results highlight the importance of investigating infants' and children's social interactions during unscripted infant–parent interactions, using dense microbehavioral methods and event-sequential analyses to understand changes in their social skills and sensitivities, *and* related changes in the social environment.

ACKNOWLEDGMENT

The study was supported by grants from the MIND Institute and the National Alliance for Autism Research to J. Triesch, J. Movellan, and G. Deák, by a grant to G. Deák and J. Triesch from the National Science Foundation (SES-0527756), and by a grant from the National Science Foundation (DLS-1227222) to G. Deák and S. Makeig. J. Triesch is supported by the Quandt foundation. Thanks to Catherine Best, Austin Hilliard, Cindy Nam, Louisa Salisbury, Leigh Sepeda, Shefali Sharma, and Yuri You for assistance, and to Ian Fasel and Javier Movellan for helpful discussions. Anna Krasno, PhD, is at Child Abuse Listening & Mediation (CALM), Santa Barbara, CA, USA.

REFERENCES

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*, 427–432. https://doi.org/10.1111/1467-9280.01458

Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The development of symbol-infused joint engagement. *Child Development*, *75*, 1171–1187.

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism and Developmental Disorders*, *39*(1), 84–96.

Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language 18*, 41–49.

Akhtar, N., & Gernsbacher, M. A. (2008). On privileging the role of gaze in infant social cognition. *Child Development Perspectives*, *2*, 59–65. https://doi.org/10.1111/j.1750-8606.2008.00044.x

Amano, S., Kezuka, E., & Yamamoto, A. (2004). Infant shifting attention from an adult's face to an adult's hand: A precursor of joint attention. *Infant Behavior and Development*, *27*(1), 64–80.

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, *59*, 197–209.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*, 1278–1289.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*, 395–418.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*, 415–419.

Bambach, S., Franchak, J. M., Crandall, D. J., & Yu, C. (2014). Detecting hands in children's egocentric views to understand embodied attention during social interaction. *Proceedings, 36th annual meeting of the Cognitive Science Society*, *36*, 134–139.

de Barbaro, K., Johnson, C., & Deák, G. O. (2013). Twelve-month 'social revolution' emerges from mother-infant sensory-motor coordination: A longitudinal investigation. *Human Development*, *56*, 223–248.

de Barbaro, K., Johnson, C., Forster, D., & Deak, G. O. (2013). Methodological considerations for investigating the microdynamics of social interaction development. *IEEE Transactions on Autonomous Mental Development*, *5*, 258–270.

Beebe, B., Jaffe, J., Markese, S., Buck, K., Chen, H., Cohen, P, ... Feldstein, H. (2010). The origins of 12-month attachment: A microanalysis of 4-month mother-infant interaction. *Attachment & Human Development*, *12*, 3–141.

Bigelow, A. E., MacLean, K., & Proctor (2004). The role of joint attention in the development of infants' play with objects. *Developmental Science*, *7*, 518–526.

Bornstein, M. H., Toda, S., Azuma, H., Tamis-LeMonda, C., & Ogino, M. (1990). Mother and infant activity and interaction in Japan and in the United States: II. A comparative microanalysis of naturalistic exchanges focused on the organisation of infant attention. *International Journal of Behavioral Development*, *13*, 289–308.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1), 72–83.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*(1), 55–72.

Camaioni, L., Aureli, T., Bellagamba, F., & Fogel, A. (2003). A longitudinal examination of the transition to symbolic communication in the second year of life. *Infant and Child Development*, *12*, 1–26.

Chang, L., de Barbaro, K., & Deák, G. O. (2016). How do infants' actions affect when parents name objects? *Developmental Neuropsychology*, *41*, 342–361. https://doi.org/10.1080/87565641.2016.1274313

Chun, R. W., Pawsat, R., & Forster, F. M. (1960). Sound localization in infancy. *The Journal of Nervous and Mental Disease*, *130*, 472–476.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220.

Colonnesi, C., Zijlstra, B. J., van der Zande, A., & Bögels, S. M. (2012). Coordination of gaze, facial expressions and vocalizations of early infant communication with mother and father. *Infant Behavior and Development*, *35*, 523–532.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*, 141–168.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology*, *40*, 271–282.

Deák, G. O. (2014). Development of adaptive tool-use in early childhood: Sensory-motor, social, and conceptual factors. In J. Benson (Ed.), *Advances in child development and behavior*, *46* (pp. 149–181). San Diego: Academic Press.

Deák, G. O. (2015). When and where do infants follow gaze? *Proceedings of International Conferences on Development and Learning and on Epigenetic Robotics*. Providence, RI, USA.

Deák, G. O., Flom, R. A., & Pick, A. (2000). Effects of gesture and target on 12 -and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, *36*, 511.

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

Deák, G. O., Walden, T. A., Yale Kaiser, M., & Lewis, A. (2008). Driven from distraction: How infants respond to parents' attempts to elicit and re-direct their attention. *Infant Behavior and Development*, *31*(1), 34–50.

Deaner, R. O., & Platt, M. L. (2003). Reflexive social attention in monkeys and humans. *Current Biology*, *13*, 1609–1613.

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*, *20*, 569–572.

Essig, K., Dornbusch, D., Prinzhorn, D., Ritter, H., Maycock, J., & Schack, T. (2012). Automatic analysis of 3D gaze coordinates on scene objects using data from eye-tracking and motion-capture systems. *ACM Proceedings: Symposium on Eye Tracking Research and Applications* (pp. 37–44).

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 9602–9605.

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

Flom, R., Deák, G. O., Phill, C., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, *27*, 181–194.

Flom, R., & Pick, A. D. (2003). Verbal encouragement and joint attention in 18-month-old infants. *Infant Behavior and Development*, *26*(2), 121–134. https://doi.org/10.1016/S0163-6383(03)00012-2

Flom, R., & Pick, A. D. (2005). Experimenter affective expression and gaze following in 7-month-olds. *Infancy*, *7*, 207–218.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*, 1738–1750.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170.

Fricker, D., Zhang, H., & Yu, C. (2011). Sequential pattern mining of multimodal data streams in dyadic interactions. *International Conference on Development and Learning*, 1–6.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*, 694–724.

Girton, M. R. (1979). Infants' attention to intrastimulus motion. *Journal of Experimental Child Psychology*, *28*, 416–423. https://doi.org/10.1016/0022-0965(79)90072-9

Gogate, L. J., Bolzani, L. H., & Betancourt, E. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, *9*, 259–288.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279.

Gros-Louis, J., West, M. J., & King, A. P. (2016). The influence of interactive context on prelinguistic vocalizations and maternal responses. *Language Learning and Development*, *12*, 280–294.

Haith, M. M. (1966). The response of the human newborn to visual movement. *Journal of Experimental Child Psychology*, *3*, 235–243. https://doi.org/10.1016/0022-0965(66)90067-1

Harel, H., Gordon, I., Geva, R., & Feldman, R. (2011). Gaze behaviors of preterm and full-term infants in nonsocial and social contexts of increasing dynamics: Visual recognition, attention regulation, and gaze synchrony. *Infancy*, *16*, 69–90.

Harris, M., Jones, D., & Grant, J. (1983). The nonverbal context of mothers' speech to infants. *First Language*, *4*(10), 21–30.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*, 188–194.

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*(3) https://doi.org/10.1111/desc.12381

Hsu, H. C., & Fogel, A. (2003). Stability and transitions in mother-infant face-to-face communication during the first 6 months: a microhistorical approach. *Developmental Psychology*, *39*, 1061–1082.

Hunnius, S., & Gueze, R.H. (2005). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, *6*, 231–235.

Hutchins, E. (1996). *Cognition in the wild*. Cambridge: Bradford.

Jasso, H., Triesch, J., Deák, G., & Lewis, J. (2012). A unified account of gaze following. *IEEE Transactions on Autonomous Mental Development*, *4*, 257–272.

Kawahara, J., Yanase, K., & Kitazaki, M. (2012). Attentional capture by the onset and offset of motion signals outside the spatial focus of attention. *Journal of Vision*, *12*, 10.

Kaye, K., & Fogel, A. (1980). The temporal structure of face-to-face communication between mothers and infants. *Developmental Psychology*, *16*, 454–464.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. New York: Psychology Press.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311–1328.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

de León, L. (2000). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology*, *8*, 131–161.

Markus, J., Mundy, P., Morales, M., Delgado, C., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development*, *9*, 302–315.

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word–object relations. *Infancy*, *13*, 172–184.

Messer, D. J., & Vietze, P. M. (1988). Does mutual influence occur during mother-infant social gaze? *Infant Behavior and Development*, *11*(1), 97–110.

Mondloch, C. J., Horner, M., & Mian, J. (2013). Wide eyes and drooping arms: Adult-like congruency effects emerge early in the development of sensitivity to emotional faces and body postures. *Journal of Experimental Child Psychology*, *114*, 203–216.

Moore, C. (2008). The development of gaze following. *Child Development Perspectives*, *2*, 66–70.

Moore, C., Angelopoulos, M., & Bennett, P. (1997). The role of movement in the development of joint visual attention. *Infant Behavior and Development*, *20*(1), 83–92.
Morrongiello, B. A., Fenwick, K. D., Hillier, L., & Chance, G. (1994). Sound localization in newborn human infants. *Developmental Psychobiology*, *27*, 519–538.
Mundy, P., Block, J., Delgado, C., et al. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, *78*, 938–954.
Nagai, Y., & Rohlfing, K.J. (2008). Parental action modification highlighting the goal versus the means. *Proceedings, 7th International Conference on Development and Learning*, 1–6. IEEE.
Nomikou, I., & Rohlfing, K. J. (2011). Language does something: Body action and language in maternal input to three-month-olds. *IEEE Transactions on Autonomous Mental Development*, *3*, 113–128.
O'Neill, M., Bard, K. A., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science*, *8*, 352–359.
Rader, N., & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture*, *10*, 202–221.
Rader, N., & Zukow-Goldring, P. (2015). The role of speech-gesture synchrony in clipping words from the speech stream: Evidence from infant pupil responses. *Ecological Psychology*, *27*, 290–299.
Rossmanith, N., Costall, A., Reichelt, A. F., López, B., & Reddy, V. (2014). Jointly structuring triadic spaces of meaning and action: Book sharing from 3 months on. *Frontiers in Psychology*, *5*, 1390.
Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, *4*, 5.
Stack, D. M., & Arnold, S. L. (1998). Changes in mothers' touch and hand gestures influence infant behavior during face-to-face interchanges. *Infant Behavior and Development*, *21*(3), 451–468.
Stein, R. (2012). The impact of synchronized tactile stimulation on joint attention in 11-month-old infants. Theses and Dissertations Paper 1774. digitalcommons.ryerson.ca/dissertations.
Tamis-LeMonda, C. S., & Bornstein, M. H. (1991). Individual variation, correspondence, stability, and change in mother and toddler play. *Infant Behavior and Development*, *14*, 143–162.
Tamis-LeMonda, C. S., & Bornstein, M. H. (1994). Specificity in mother-toddler language-play relations across the second year. *Developmental Psychology*, *30*, 283.
Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126.
Tamis-LeMonda, C. S., Kuchirko, Y., & Tafuro, L. (2013). From action to interaction: Infant object exploration and mothers' contingent responsiveness. *IEEE Transactions on Autonomous Mental Development*, *5*, 202–209.
Tomasello, M. (1999). *The cultural origins of cognition*. Cambridge, MA: Harvard.
Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*, 705–722.
Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–1463.
Tremblay, H., & Rovira, K. (2007). Joint visual attention and social triangular engagement at 3 and 6 months. *Infant Behavior and Development*, *30*, 366–379.
Van Egeren, L., Barratt, M., & Roach, M. (2001). Mother- infant responsiveness: Timing, mutual regulation, and interactional context. *Developmental Psychology*, *37*, 684–697.
Waxman, R. P., & Spencer, P. E. (1997). What mothers do to support infant visual attention: Sensitivities to age and hearing status. *Journal of Deaf Studies and Deaf Education*, *2*, 104–114.
Wu, Z., & Gros-Louis, J. (2014). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *Journal of Child Language*, *41*(4), 1–24.
Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*, 118–136.
Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience *Infancy*, *13*, 229–248.
Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244–262.
Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, *8*, e79659.
Zukow, P. G. (1990). Socio-perceptual bases for the emergence of language: An alternative to innatist approaches. *Developmental Psychobiology*, *23*, 705–726.
Zukow-Goldring, P. (1996). Sensitive caregivers foster the comprehension of speech: when gestures speak louder than words. *Early Development and Parenting*, *5*, 195–211.

Zukow-Goldring, P. (1997). A social ecological realist approach to the emergence of the lexicon: Educating attention to amodal invariants in gesture and speech. In C. Dent-Read, & P. Zukow-Goldring (Eds.), *Evolving explanations of development* (pp. 199–250). Washington, DC: APA.

Zukow-Goldring, P., & Arbib, M. A. (2007). Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention. *Neurocomputing*, *70*(13), 2181–2193.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. Proportional distribution of previous states (left-hand boxes) and subsequent states (right-hand boxes), for each of two attention-sharing states (center boxes: figure in parentheses is the proportion of all 4,411 states accounted for by that state).

Figure S2. Proportional distribution of previous (left) and subsequent (right) states, for each non-attention-following state (center boxes, percentages indicate occurrences out of all states).