Title

The slow emergence of gaze- and point-following: A longitudinal study of infants from 4 to 12 months

Authors
Tang, Yueyan
Gonzalez, Marybel Robledo
Deák, Gedeon O

Publication Date
2023-11-08

DOI
10.1111/desc.13457

Received: 29 December 2022 | Revised: 14 September 2023 | Accepted: 2 October 2023

DOI: 10.1111/desc.13457

Developmental Science WILEY

RESEARCH ARTICLE

The slow emergence of gaze- and point-following: A longitudinal study of infants from 4 to 12 months

Yueyan Tang[1] | Marybel Robledo Gonzalez[1,2] | Gedeon O. Deák[1]

[1]Department of Cognitive Science, University of California, San Diego, California, USA

[2]Department of Psychiatry and Behavioral Health, The Ohio State University, Columbus, Ohio, USA

Correspondence
Gedeon Deák, Department of Cognitive Science, 9500 Gilman Dr., UC – San Diego, San Diego, CA, 92093-0515, USA.
Email: gdeak@ucsd.edu

Funding information
National Science Foundation, Grant/Award Number: SES-0527756

Abstract

Acquisition of visual attention-following skills, notably gaze- and point-following, contributes to infants' ability to share attention with caregivers, which in turn contributes to social learning and communication. However, the development of gaze- and point-following in the first 18 months remains controversial, in part because of different testing protocols and standards. To address this, we longitudinally tested $N = 43$ low-risk, North American middle-class infants' tendency to follow gaze direction, pointing gestures, and gaze-and-point combinations. Infants were tested monthly from 4 to 12 months of age. To control motivational differences, infants were taught to expect contingent reward videos in the target locations. No-cue trials were included to estimate spontaneous target fixation rates. A comparison sample ($N = 23$) was tested at 9 and 12 months to estimate practice effects. Results showed gradual increases in both gaze- and point-following starting around 7 months, and modest month-to-month individual stability from 8 to 12 months. However, attention-following did not exceed chance levels until after 6 months. Infants rarely followed cues to locations behind them, even at 12 months. Infants followed combined gaze-and-point cues more than gaze alone, and followed points at intermediate levels (not reliably different from the other cues). The comparison group's results showed that practice effects did not explain the age-related increase in attention-following. The results corroborate and extend previous findings that North American middle-class infants' attention-following in controlled laboratory settings increases slowly and incrementally between 6 and 12 months of age.

KEYWORDS
gaze following, infant social development, joint attention, longitudinal, pointing, social cognition

Research Highlights

- A longitudinal experimental study documented the emergence and developmental trajectories of North American middle-class infants' visual attention-following skills, including gaze-following, point-following, and gaze-and-point-following.

- A new paradigm controlled for factors including motivation, attentiveness, and visual-search baserates. Motor development was ruled out as a predictor or limiter of the emergence of attention-following.
- Infants did not follow attention reliably until after 6 months, and following increased slowly from 7 to 12 months.
- Infants' individual trajectories showed modest month-to-month stability from 8 to 12 months of age.

1 | INTRODUCTION

For decades researchers have been fascinated by visual attention-sharing in infancy, often focusing on gaze- and point-following (Scaife & Bruner, 1975). These behaviors seem to reflect infants' changing responsiveness to specific social-behavioral cues, and perhaps their increasing knowledge of other people's attentional states (Carpendale & Lewis, 2004). However, the origins of gaze- and point-following, hereafter called *attention-following* or *AF*, remain controversial. Ongoing controversies include the age at which infants acquire AF, the developmental trajectory of AF, and the *knowledge* that AF reflects.

Regarding the first controversy, different studies imply that AF emerges at different ages, partly due to different operational definitions. Some researchers' definition encompasses any tendency of infants to reorient in the same direction as a real or schematic face. Such a simple tendency can however be acquired by a primitive associative learner or artificial network (Ballard, 1987; Hawkins & Byrne, 2015). Other researchers' definition of AF is narrower: the ability to search for and fixate a localized target by extrapolating the direction of another person's gaze or point. This, too, is learnable by various animals, possibly via widely conserved learning mechanisms operating on structured input (see Jasso et al., 2012; Triesch et al., 2006). Mature human AF eventually includes a more specific ability to use other people's behaviors to shift attention to targets specified by those behaviors. These discrepant definitions have spawned a confusing range of claims about the age of emergence of AF, and these claims imply radically different developmental explanations. Some studies suggest that infants as young as 4 months of age (e.g., Hood et al., 1998, Astor & Gredebäck, 2019), or even newborns (Farroni et al., 2004), can follow gaze. These claims rest on the first, broader definition. However, most studies indicate that gaze-following emerges later, between 6–12 months (e.g., Butterworth & Jarrett, 1991; Deák et al., 2000; Flom et al., 2004), based on the narrower definition.

The controversy has been difficult to resolve for at least two reasons: first, although there are very few findings of gaze-following at very young ages, these must nonetheless be explained: for example, are they the result of more sensitive measurements, or do they reflect a different behavior altogether (i.e., a confound)? Second, although more studies indicate AF emerging at older ages, these do not paint a uniform picture of age-of-emergence. This might be due to between-studies differences in sampling, methodology, or analysis. In any case, the differences among results obscures the prevailing developmental trajectory. A longitudinal study with a methodologically improved method might more clearly establish the emergence and development of AF behaviors during the first year, at least for laboratory settings (see Tang et al., 2023). Most AF studies have used one of two paradigms. One, a paradigm developed by Butterworth and colleagues (e.g., Butterworth & Jarrett, 1991), tests the narrower AF definition. In it, an adult produces a directional cue (gaze or point) toward one of 4–6 targets (objects or 2D geometric shapes) distributed around a room. Targets are relatively far (~1.5 m) from the infant and adult. To succeed, the infant must follow the cue to the specific target while ignoring other targets. This paradigm has indicated emergence of AF around 6–9 months in infants in predominantly WEIRD (western, educated, industrialized, rich, and democratic; Henrich et al., 2010) samples. Simulations of the Butterworth paradigm show that a simple reinforcement learning agent can learn gaze-following in a progression mirroring human infants, if given structured teaching input (Jasso et al., 2012).

In the second paradigm, which presumes the broader AF definition (Farroni et al., 2000; Hood et al., 1998), younger infants see a real or artificial head, usually in isolation, with targets just to its left and/or right. The head (or eyes) rotates horizontally toward a target. Infants' saccades in the same direction as the head/eyes are interpreted as gaze-following. However, this paradigm can elicit primitive *motion-cued ipsilateral scanning*: a tendency to scan in the same direction as a moving object (Deák, 2015). There are reasons to believe that motion-cued scanning explains the aforementioned results: first, the studies reporting "following" by younger infants showed only one target per side, and coded *any* ipsilateral gaze-shift as gaze-following, so directional cues were confounded with rotational motion. Second, Farroni et al. (2003) showed that 4-month-olds' gaze-shifting depends on head motion, not the eyes' angle. Third, multiple studies showed that even older infants must see the head turn in the appropriate direction in order to follow (e.g., Moore et al., 1997). Motion-cued ipsilateral scanning even parsimoniously explains the original Scaife and Bruner (1975) data that sparked interest in AF.[1]

A brief summary of key differences between these two paradigms, including the implied definition of AF, is provided in Table 1 (top two cells). We note that both paradigms are limited: cross-cultural evidence reveals diverse caregiver-child AF patterns (e.g., Jurkat et al., 2023, Childers et al., 2007) that often differ from laboratory paradigms. Thus, the task used here, though a more thorough and controlled example of the most impactful laboratory paradigm for studying infant

TABLE 1 Selective comparison of two historical paradigms and the current paradigm: Key features.

Directional turning paradigm
Operationalization of AF: Tendency to reorient in the same hemifield (left/right) as a face or eyes.
Social cue: Real or video-based head, sometimes in isolation. The head (or eyes) rotate or displace horizontally toward the left or right target.
Context: Video display in an infant testing room (i.e., usually no "embodied" social partner or targets).
Targets: Images of objects (or actual objects) next to, and close to (i.e., <1 m), the side(s) of the depicted head.
Measure of AF: Infant shifts gaze in the same direction (left or right) as the head/eyes.
Butterworth-inspired paradigms
Operationalization of AF: Visual reorientation to a specific region, based on the direction of another person's gaze or point.
Social Cue: Cue-giver looks or points toward one of several distal targets.
Context: Real-time interaction with a human cue-giver.
Targets: Six objects or 2D geometric shapes distributed around the room and placed further (i.e., > 1 m) from infant and cue-giver.
Measure of AF: Infant initially shifts attention to the specified target.
Current Paradigm ('Modified Butterworth')
Operationalization of AF: Visual reorientation to a specific region, based on the direction of another person's gaze or point.
Social Cue: Cue-giver looks and/or points toward one of several distal targets.
Context: Larger testing room; real-time interaction with a human cue-giver.
Targets: Six target monitors distributed around the room, three per side, placed further (~2 m) from infant and cue-giver.
Measure of AF: Infant initially shifts attention to the specified target.

Note: Summary of paradigms and definitions used in previous studies, and the current study.

AF, does not represent all AF patterns documented in naturalistic infant-caregiver interactions.

Because previous findings purporting to show gaze-following in younger infants can be parsimoniously attributed to motion-cued scanning, the status of AF in infants 6 months and younger remains unclear. Perhaps previous studies used methods insensitive to infants' nascent AF ability: for example, 15- and 22-month-olds will follow gaze in laboratory studies, but apparently not in a cluttered environment (Deák et al., 2008). Also, in laboratory studies, AF is modulated by the targets' "interestingness" (Deák et al., 2000; Flom et al., 2004). Finally, studies using the 'Butterworth paradigm' might have imposed added motor demands that obscured younger infants' AF abilities. If AF is limited by infants' vigilance, motivation, or motor maturity, then previous negative results might reflect procedural factors that suppressed younger infants' AF: e.g., unmotivating targets, too-subtle cues, or age-inappropriate motor demands.

To resolve this question, the current study was designed to test more thoroughly and sensitively the age of emergence of infants' gaze- and point-following. We tested, longitudinally, visual AF by infants—specifically, healthy infants from North American, predominantly-WEIRD[2] families—every month from 4 to 12 months of age. The 4- to 12-month age range was chosen to cover the range most commonly investigated in infant AF studies. Infants completed a Butterworth-type paradigm, with modifications to minimize extraneous difficulty and to increase infants' motivation to follow social cues (Table 1, bottom cell). By controlling extraneous factors including motivation, motor immaturity, and attentiveness, we attempted to determine the age when gaze- and point-following—not just motion-cued scanning—typically exceeds chance levels in this population. We further sought to document the developmental trajectory of AF in infants from 4 to 12 months.

To address potential motivational limitations, a contingent reinforcement task was used. Infants were taught that responding to cues towards target locations could trigger a visual reward (specifically, videos chosen to maximize infants' interest). Infants first learned that each of six target monitors would play these videos (de Barbaro et al., 2011); this also verified infants' ability to turn to each target location, thereby establishing minimal perceptuomotor maturity. Subsequently, if infants followed a gaze or point cue to the correct target location during a test trial, a reward video played at that location. If infants did *not* follow, the video reward nevertheless played after a trial ended, so that all infants eventually saw the same number of reward videos. However, in non-following trials the reward was not contingent on the infant's actions, and started after the social cue ended, to minimize behavioral shaping during the sessions. This method was designed to motivate infants to follow social cues to obtain visual rewards, without teaching AF responses *de novo*. The design is supported by evidence that infants can rapidly learn to use social cues to orient to interesting sights (e.g., Corkum & Moore, 1998; Triesch et al, 2006, 2007; Wang et al., 2012). The design avoids differential reinforcement of certain target-responses; however, it could cause infants to increase scanning of target locations independent of social cues. To rule out this possibility, we included baseline trials in which the cue-giver did not produce a cue toward any target. These trials estimate infants' tendency to scan target locations in the absence of cues. These modifications make this procedure an unusually sensitive *and* stringent test of AF.

Other procedural details were designed to assess or compensate for immaturities that might obscure younger infants' AF capability. First, extraneous distractions were minimized by testing infants in an isolated room with no clutter. Second, on every trial, the experimenter did not produce cues until the infant looked at her. Finally, instead of holding a static cue for 7–10 sec as in previous studies (e.g., Deák et al., 2000), the cue was repeated after 4 sec, and the trial lasted ~10 sec total—long enough for even younger infants to respond (Deak et al., 2014). The cue repetition both compensates for any lapses in infants' attention, and better approximates the naturalistic attention-directing cue combinations of parents in the sample population (e.g., Deák et al., 2017). These modifications thereby address ancillary factors that might suppress AF by younger infants.

The design addresses several other outstanding questions. First, it allows us to compare developmental trajectories of gaze-following versus point-following. Infants were tested within-subjects on following gaze cues, point cues, and combined gaze and point cues. Deák et al. (2017) found that infants follow mothers' spontaneous cue combinations (see also Zukow-Goldring, 1996), suggesting that *redundant* cues normally precede AF. Also, previous laboratory studies suggest that pointing, or combined gaze and point cues, elicit following more than gaze cues alone (Deák et al., 2000, 2008; Flom et al., 2004). However, no previous study has directly compared AF of gaze cues, point cues, and both cues together, across a range of infant ages.

Second, because infants were tested monthly and might have benefited from practice effects, especially in later months, an independent comparison group of infants was recruited and tested using the same procedure, but only at 9 and 12 months of age. This comparison group allows us to estimate practice effects in the longitudinal sample. Nine months is considered a watershed for infants' social responsiveness, including attention-sharing (Tomasello, 1999; de Barbaro et al., 2016), and by 12 months infants show above-chance gaze-following in more-challenging laboratory trials (e.g., Deák et al, 2000).

We hypothesized that infants would not follow attention-cues reliably before 6 months, and then show a gradual increase from 7 to 12 months. These responses were expected to be location-dependent, with earliest, more-robust AF to targets within infants' visual field, later, less-reliable AF to peripheral targets, and latest, least-robust AF to targets outside infants' visual field (i.e., behind them). This prediction is based on previous studies (e.g., Butterworth & Jarrett, 1991; Deák et al., 2000), and simulations (Jasso et al., 2012). We also hypothesized that infants would follow pointing more than gaze cues, and follow gaze-and-point (GP) combinations more than either cue alone. Additionally, we predicted modest month-to-month stability of individual infants' AF levels (Markus et al., 2000; Mundy & Gomes, 1998), especially after 6 months. We expected these differences to be at most weakly related to individual infants' motor development. Finally, we expected no practice effects, due to the infrequency and brevity of the sessions.[3]

2 | METHOD

2.1 | Participants

2.1.1 | Main sample

A sample of convenience of 48 infant-parent dyads was recruited from middle-class suburban and urban neighborhoods in a large Southern California city. Parents were all English-fluent, and were recruited from postpartum exercise classes, posts to playgroups and parent listservs, and by word-of-mouth. Infants were excluded if they were >2 weeks premature or were reported by parents to have had significant perinatal complications or sensory or neurological problems. Five families withdrew before the second lab visit and were dropped from analyses. The 43 remaining infants (20 female) all participated with their biological mothers. Their demographic characteristics are shown in Table 2.

2.1.2 | Comparison sample

A separate comparison group of 24 infants (12 girls) was recruited from the same sources in the same community, with the same exclusion criteria, as the main sample. They were tested in the lab and visited at home at 9 and 12 months of age. Their characteristics are shown in Table 3.

2.2 | Materials

2.2.1 | Reward videos

Pilot testing of video clips from different sources showed that clips from the commercial Baby Einstein series were most interesting to infants, so these were used as reward videos. The clips show high-contrast toys or animals moving to synchronized, simplified, electronic arrangements of European mid-tempo diatonic "classical" music. Commercial and research evidence (DeLoache & Chiong, 2009; Demers et al., 2013) corroborate our pilot findings that these videos optimally capture infants' attention. Two researchers selected and rendered a set of 8-sec video clips with varied dynamic visual properties and musical selections.

2.2.2 | Testing environment

Infants were tested in a sound-attenuated room (4.0 × 3.6 m) with two chairs and six 33 cm monitors (Figure 1) positioned approximately 2 meters from the infant's head, and centered at the height of an average infant's eyes when seated on an adult's lap. Three monitors were mounted to the wall on each side of the room. Two monitors were in front of the infant (± 33° from midline), two were in the infant's periphery (± 78° from midline), and two were behind the infant (± 126°). Video cameras (JVC Everios) were fixed above the front and back monitors and pointed at the infant and/or the experimenter (EXP) (see Figure 1). A fifth camera with a fisheye lens was mounted overhead, centered on the infant and EXP. A soundbar on each monitor played the audio for any video sent to that monitor. Mothers wore opaque glasses and noise-canceling headphones playing music so they could not perceive cues or targets.

An adjacent control room displayed the cameras' timestamped output on an AV system that also pushed videos to specific monitors, and captured videos to a dedicated RAID computer. A video recorder captured backup video files.

Questionnaires and forms developed for the study can be found at https://osf.io/4nqbc/.

TABLE 2 Demographic characteristics of infants in the main sample.

Session (months)	Retained *n*	Infant age: mean days (range)	Mother age: mean yrs. (range)	Mother educ.: mean yrs. (range)	Ethnicity: *n* of: A, B, W, H, O (see *Note*)	Adults in home: *n* (see *Note*)	Parity: 1st, 2nd, 3rd+ (see *Note*)	Daycare
4	25	124.2 (114-157)	31.9 (21-39)	15.7 (12-18)	2, 0, 16, 2, 5	3, 14, 4, 2	12, 10, 3	7
5	35	156.4 (143-169)	32.4 (21-42)	16.1 (12-18)	2, 0, 25, 3, 5	3, 26, 5, 1	20, 11, 3	8
6	38	188.8 (175-209)	32.2 (26-42)	16.2 (12-21)	2, 0, 26, 2, 5	2, 30, 4, 2	23, 11, 3	10
7	39	219.7 (203-263)	32.0 (21-42)	16.2 (12-21)	2, 0, 26, 2, 5	2, 30, 4, 2	23, 11, 3	10
8	32	248.8 (239-298)	32.4 (21-42)	16.3 (12-21)	2, 0, 22, 3, 5	3, 22, 5, 1	18, 11, 3	10
9	36	277.6 (261-297)	32.5 (21-42)	16.1 (12-21)	2, 0, 27, 3, 4	3, 26, 5, 2	20, 13, 2	9
10	39	309.7 (298-318)	32.7 (27-42)	16.3 (12-21)	1, 0, 30, 2, 5	2, 31, 4, 2	24, 12, 2	9
11	37	339.3 (326-354)	31.4 (21-38)	16.0 (12-21)	2, 0, 28, 2, 5	3, 28, 4, 2	24, 11, 2	9
12	37	370.4 (357-390)	31.6 (21-39)	16.1 (12-21)	1, 0, 28, 3, 5	3, 29, 5, 0	24, 9, 3	8

Note: n = Number of infants included in analyses (i.e., 17+ valid trials; see text). 43 infants were tested each month (except 6 and 8 months: 42 tested). Key: *Age* = infant's age (days); *Mother Age* = mother's age (years); *Mother Edu.* = mother's years of formal education (12 = H.S. graduate; 16 = college graduate); *Ethnicity* = ethnicity of infants: respectively, numbers of: Asian (A), Black (B), White (W), Hispanic/Latinx (H), or "other"/Multi-ethnic[4] (O); *Adults in Home* = numbers of infants living with (respectively): 1, 2, 3, or 4+ adults; *Parity* = number of infants with, respectively: 0, 1, or 2 or more siblings; *Daycare* = number of infants spending at least 20 hrs./wk. in out-of-home daycare.

TABLE 3 Demographic characteristics of infants in the comparison sample.

Session (months)	Retained *n*	Infant age: mean days (range)	Mother age: mean yrs. (range)	Mother educ.: mean yrs. (range)	Ethnicity: *n* of: A, B, W, H, O (see *Note*)	Adults in Home: *n* (see *Note*)	Parity: 1st, 2nd, 3rd+ (see *Note*)	Daycare
9	23	280.2 (267–293)	33.0 (24-39)	17.6 (12-21)	0, 0, 16, 0, 7	0, 22, 1, 0	17, 5, 1	5
12	21	367.6 (355–381)	33.3 (24-42)	17.7 (12-21)	0, 0, 14, 0, 7	0, 20, 1, 0	15, 5, 0	5

Note: Variables are as described in Table 2.

2.3 | Procedure

Infants were seated on their mother's lap facing EXP. After an orientation phase (described below), EXP executed 20 trials. In 18 of these EXP produced specific cues toward specific monitors. Order of cues and monitors was quasi-randomized, and re-randomized every month. Randomization was constrained so that each combination of cue-type and location "latitude" (front, periphery, back) was included once in the first half and once in the second half of trials. Two remaining *baseline trials* are described below. EXP wore an earbud connected to a two-way radio to communicate with another experimenter (EXP2) in the control room, and to hear a click-track to maintain consistent cue and trial durations.

2.3.1 | Orientation phase

Infants socialized with EXP in the testing room until they seemed comfortable. EXP then began an orientation phase to show infants that each monitor played video rewards (de Barbaro et al., 2011). EXP

FIGURE 1 Quad-camera frame of a pointing trial. Note: An infant sits on his mother's lap, facing EXP, who is executing a point cue to a side monitor. The infant has followed the cue to the target monitor, so the reward video has been initiated. Note that in these images, post hoc blurring (to de-identify participants) makes the infant's head and eye angles less discernible than it was to coders.

TABLE 4 Trial types: Cue types and cue location.

Cue location	Gaze	Point	GP
Front	2 (1 left, 1 right)	2	2
Side	2	2	2
Back	2	2	2

Note: GP = Gaze and Point. Each cell represents two trials; one to a left target and one to a right target. The table does not show two additional baseline trials.

used redundant natural social cues (voice, gesture, gaze, touch, posture) to direct the infant's attention to a monitor. When the infant fixated the monitor, EXP2 triggered a video reward. Each monitor played a different video, and the videos changed in every session. EXP drew infants' attention to each monitor in quasi-random order. After the orientation trials there was a brief pause before the test phase began.

2.3.2 | Test phase

In each of 18 test trials (Table 4) EXP silently produced one of three cues: (1) Gaze: head and eyes turned smoothly from infant (midline) to the center of the target monitor; (2) Point: arm and hand swept smoothly from EXP's lap into a fully extended index-point toward the center of the target monitor, and torquing the upper body as necessary; (3) Gaze and Point (GP): EXP produced both cues simultaneously toward the target monitor. See Figures 1–2 for examples.

In each trial EXP first called the infant to attract her/his attention. When the infant fixated on EXP's face, EXP produced the cue, held it for 4 sec, then turned back to the infant (and lowered her hand, if necessary), waited ~1 sec, then produced the cue again for 4 sec before returning her gaze to midline and/or her hand to her lap to end the trial. In each trial all cues were directed to a specific monitor.

EXP2 monitored the infant and activated the reward video as soon as the infant fixated the target monitor. A fixed delay in the analog video system, plus EXP2's protocol for verifying a target fixation, meant that the video reward began 1.5 – 2.5 sec after the infant's fixation began. Alternately, if the infant did *not* fixate the target monitor, EXP2 activated the reward 2 – 2.5 sec after the cue ended. Thus every infant saw every reward video, either during the trial (if they followed), or afterwards. In very rare cases where EXP2 made a reward-activation error, the trial was dropped.

Two baseline trials were included, one in each half of the trials, in a quasi-random order that changed each month. EXP attracted the infant's attention but then looked down at her lap for 10 sec. These trials were used to estimate infants' tendency to fixate specific monitors in the absence of an attention-directing cue. One monitor was randomly designated the target monitor in each trial; this changed each month.

2.3.3 | Coding

Videos from 5 cameras (described above) were synchronized. Trained coders, blind to specific hypotheses, coded (framewise; 30 Hz) the

FIGURE 2 Overhead camera frames of a GP cue to the back, side, and front target monitors. Note: Overhead camera frames show EXP directing the infant's attention to (from left to right, respectively) a back, side, and front target monitor. Each frame is taken from a trial when the infant followed, and the reward video had been started.

onset and offset of each cue, and time and location of infants' visual fixations (defined as no saccade for > 0.1 sec) of: EXP's face, EXP's pointing hand, each of six video monitors, the mother, or other features of the room. Our analyses focus on infants' first fixations away from EXP after the cue started. Intercoder reliability (i.e. proportion agreement) for *hits* (first fixation to target) versus misses (first fixation to any other location) was 0.95 and 0.93 for, respectively, the main sample (based on 20 randomly-selected sessions), and the comparison sample (13 randomly-selected sessions).

2.3.4 | Standardized tests

At every session parents completed a Motor Development Milestones Questionnaire (MDMQ; available at https://osf.io/4nqbc/) to assess motor skill changes. Parents indicated, to the nearest week, when their infant first showed a given skill. EXP explained each item and showed photographs of the behavior to ensure mothers' understanding. Scores were summed for three milestones related to early head and trunk control, which might limit an infant's ability to turn to peripheral or back targets. Those were: *Lifts head while either laying on the belly or back*; *Raises head and chest when put on belly (supporting upper body with arms)*; and *Keeps head level when pulled to a sitting position*. For the $n = 40$ infants with valid age-estimates for these items, the average age of acquisition was 3.2 months ($SD = 1.0$; range: 0.5 to 6.0). Fourteen parents reported that their infant had not achieved one or more of these milestones by the 4-month visit.

2.3.5 | Statistical analyses

A linear mixed model analysis (LMM) using the *lmer* function (R 4.2.21 lme4 package) compared infants' AF proportions (see below) at each month as dependent variables, with infants' age, sex, and cue-type as fixed factors in all models. Infant identity was a random factor to estimate individual differences. All dependent variables were standardized. Additionally, a mixed-measures ANOVA was run to compare groups (main vs. comparison) at two ages (9 and 12 months).

2.3.6 | Dependent measures

All analyses except as noted focus on *AF hits*, defined as an infant reorienting to the cue-specified target after seeing the cue[5]. Hits could be defined using *stringent* or *liberal* criteria. *Stringent* hits were defined as infants fixating the target monitor on their first fixation away from EXP. *Liberal* hits were defined as infants fixating the target any time between the start of the cue and the end of the trial, but not necessarily on the first post-cue fixation. The analyses below apply the *stringent* definition, to minimize false alarms. However, for purposes of comparison the Liberal criterion results are reported in *Supplemental Materials* (Table S2).

2.3.7 | Data retention and missing data

Session data were retained if there were no more than 3 missing trials[6]. Preliminary examination revealed that this threshold minimized missingness without affecting the results. Notably, the number of completed trials (17 to 20) was not significantly or marginally correlated with proportion of hits (weighted, see below), either in any single month or in the averages of: the first three months (4-6 months, $r(17) = .068$, $p = .78$)[7], middle three months (7-9, $r(23) = .33$, $p = .10$), or last three months (10-12, $r(28) = -.029$, $p = .88$). Thus, we can treat missingness as at-random (Little & Rubin, 2002), and optimize power by including sessions with 1–3 missing trials.

Because the number of trials varied slightly across sessions, results are reported as proportions of hits out of the number of valid trials in the session. However, session-wise differences in *which* trials were missing might increase error variance. We corrected for this by applying a trial-type weighting to each hit. At each month, hit rates for 9 trial types (3 cue types [gaze, point, GP] x 3 location-pairs [front, side, back]) were linearly converted into 9 normalized weights (see Appendix).[8] Each total hit rate (i.e., Hits(w_t)) was adjusted by weighting the proportion of hits for each cue-and-location trial type based on the proportion of total trials represented by that trial type. This ensured that each trial type played a proportionally equivalent role in each infant's total hits, regardless of minor differences in trial type numbers across sessions. These weighted hit proportions are more comparable across infants within a month, and within infants across months. Monthly raw (unweighted) hit proportions, and weighted hit proportions, are reported in Supplemental Table S1. These show that weighting changed the mean hit rates by < 1% per month.

3 | RESULTS

3.1 | Were videos rewarding and accessible?

We first assessed whether infants were motivated to watch the videos, and able to smoothly orient to all monitor locations. Infants correctly fixated a mean of 5.49 (*SD* = 0.5) out of 6 monitors during orientation trials, confirming that videos were motivating and accessible. Monthly averages from 4 through 12 months were: 3.97 (*SD* = 1.2), 4.75 (1.5), 5.58 (0.7), 5.61 (0.6), 5.74 (0.7), 5.95 (0.2), 5.82 (0.6), 5.97 (0.2), and 6.0 (0.0). Thus, from 6 months on, infants uniformly, consistently oriented to all video locations. The means at 5 months were higher than 4 months ($Z = -2.77$, $p = .005$), and at 6 months were higher than at 4 ($Z = -4.18$, $p < .001$) or 5 months ($Z = -2.81$, $p = .005$). However, number of fixated orientation videos was not reliably correlated with AF (i.e., hit ratio; see below) at 4 months ($r = -.044$, $p = .87$) or 5 months ($r = .31$, $p = .076$). Thus, we retained infants' data at 4 and 5 months even if they did not orient in every orientation trial.

3.2 | Stringent versus Liberal Hit criteria

Analysis shows that criterion (Liberal vs. Stringent) did not reliably interact with age and target location (see Tables S3 for details). Therefore, we report only Stringent criteria results below.

3.3 | Baseline trials

In baseline trials, infants almost never fixated on a designated target monitor in their first fixation away from EXP: in fact, only one infant in one session fixated a designated target after their first saccade. Thus, the base-rate for spontaneous target looks (using the stringent criterion) was rounded down to zero for subsequent analyses.

3.4 | Overall cue following

Distributions of weighted hits by month are shown as boxplots in Figure 3. Initial analyses revealed no reliable differences in AF (i.e., weighted hits) to left- versus right-side targets. Thus, left and right trials are pooled in all analyses. Also, there were no differences (in weighted hits) between female and male infants ($Z = -0.068$, $p = .95$), and no interaction between sex of infant and cue type ($F(2, 948) = 0.166$, $p = .847$). Thus, further analyses pool female and male infants.

A linear mixed-effect model (LME) with subject as a random intercept showed that weighted total hit proportions increased with age

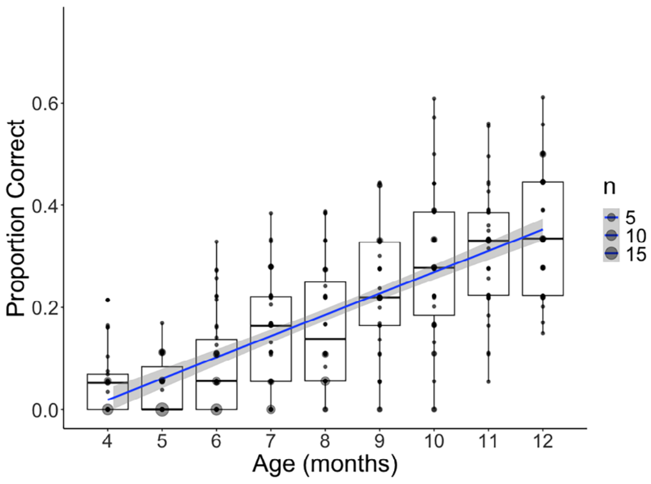

FIGURE 3 Total proportions of hits by month. Note: Boxes show proportion of weighted hits (stringent criterion—see text) each month, from sessions with >= 17 valid trials. Blue line: best-fitting linear trend ($y = 0.04\,x - 0.15$) of means (shaded region: 95% CI).

TABLE 5 Linear mixed effect results: effects of age on total proportion of hits.

Random effects	Variance	SD		
Subject	0.12	0.35		
Residual	0.41	0.64		
Fixed effects	**Estimate**	**SE**	**t**	***p***
Intercept	−0.002	0.06	−0.04	.97
Age	0.69	0.04	19	< **.001*****

Note: Weighted proportion of total hits (stringent criterion). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 6 Three-way ANOVA results based on age, cue type, and target location.

	df_{num}	df_{den}	F	*p*
Age	8	2790	65.0	<**0.001*****
Cue type	2	2790	35.4	<**0.001*****
Target location	2	2790	657	<**0.001*****
Age × cue type	16	2790	1.64	.052
Age × target location	16	2790	20.3	<**0.001*****
Cue type × target location	4	2790	16.9	<**0.001*****
Age × cue target × target location	32	2790	0.98	.503

Note: Weighted proportion of total hits (stringent criterion). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

($\beta_{Std.} = 0.69$, $p < .001$). Additionally, 23% of random effects variance was due to individual differences[9]. The full model results are shown in Table 5.[10] An additional binomial regression, treating hit rates above or below the grand median as a binary classifier (with back trials excluded due to floor effects), showed similar trends (see Table S5; Figure S4).

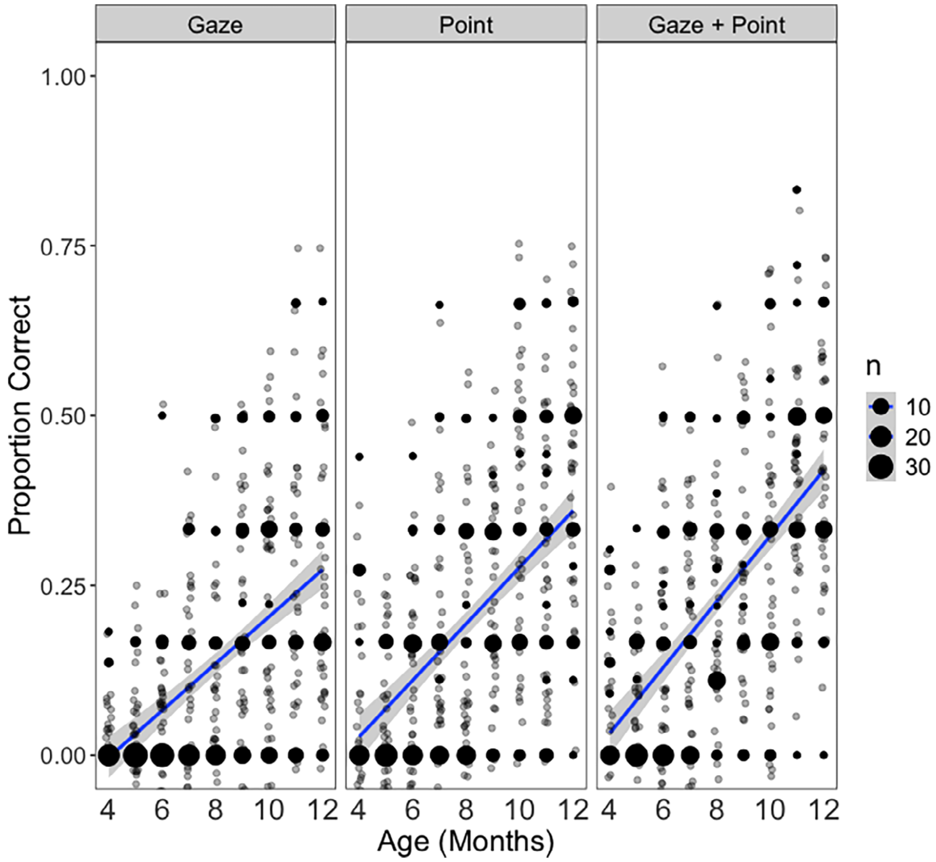

FIGURE 4 Hit proportions by month by cue type (gaze, point, GP). Note: Points indicate location-weighted hit proportions (stringent criterion) from sessions with >= 17 valid trials, for each month (X axis), for each cue type: Gaze (left panel), Point (center panel), GP (right panel). Blue lines: best-fitting linear regression trends, with 95% CI cloud.

To address interactions among all factors hypothesized to affect AF, a three-way ANOVA examined the effects of age, cue type, and target location on infants' AF from 4 to 12 months. Significant main effects were found for age, cue type, and target location, as summarized in Table 6. The age by cue type interaction was marginal; this is seen in Figure 4, which also illustrates main effects of age and cue type. Significant two-way interactions between age and target location, and between cue type and target location, are visualized in Figures 5 and 6, respectively. Each significant effect is described below.

3.5 | Age by cue type

Infants' AF for each cue type (Gaze, Point, GP) increased with age (Figure 4; Table S1). We compare individual infants' AF trajectories for each cue type by their 4-to-12 month slope coefficients. Infants followed GP cues more than point-only ($Z = -2.7$, $p = .006$), or gaze-only cues ($Z = -8.2$, $p < .001$). To compare trajectories between cue types we examined cue type by age interactions. The trajectory for GP cues was steeper than for gaze-only cues, $t(83) = 2.38$, $p = .02$. Trajectories did not differ reliably between GP and point-only cues, $t(83) = 0.83$, $p = .41$, or between gaze-only and point-only cues, $t(83) = -1.42$, $p = .16$.[11]

3.6 | Age by target location

AF to each location (front, side, back) increased with age (Figure 5, Table S1). Infants followed cues to front targets more than to side ($Z = -12.8$, $p < .001$) or back targets ($Z = -13.9$, $p < .001$). To compare development across locations we examined month by location interactions. The age trajectory was steeper for front than back locations, $t(50) = 14.0$, $p < .001$, for front than side locations, $t(81) = 6.81$, $p < .001$, and for side than back locations, $t(54) = 6.68$, $p < .001$.

3.7 | Cue type by target location

AF to different target locations varied by cue type (Figure 6). For front targets infants followed a higher proportion of GP than either point-only ($t(42) = 2.88$, $p = .006$) or gaze-only cues ($t(42) = 8.98$, $p < .001$). For side targets infants followed GP cues more than gaze-only cues ($t(42) = 2.58$, $p = .013$), but not more than point-only cues ($t(42) = 0.592$, $p = .557$). Similarly, on back target trials infants followed GP more than gaze-only cues ($t(42) = 2.58$, $p = .014$), but not more than point-only cues ($t(42) = 1.70$, $p = .097$).[12]

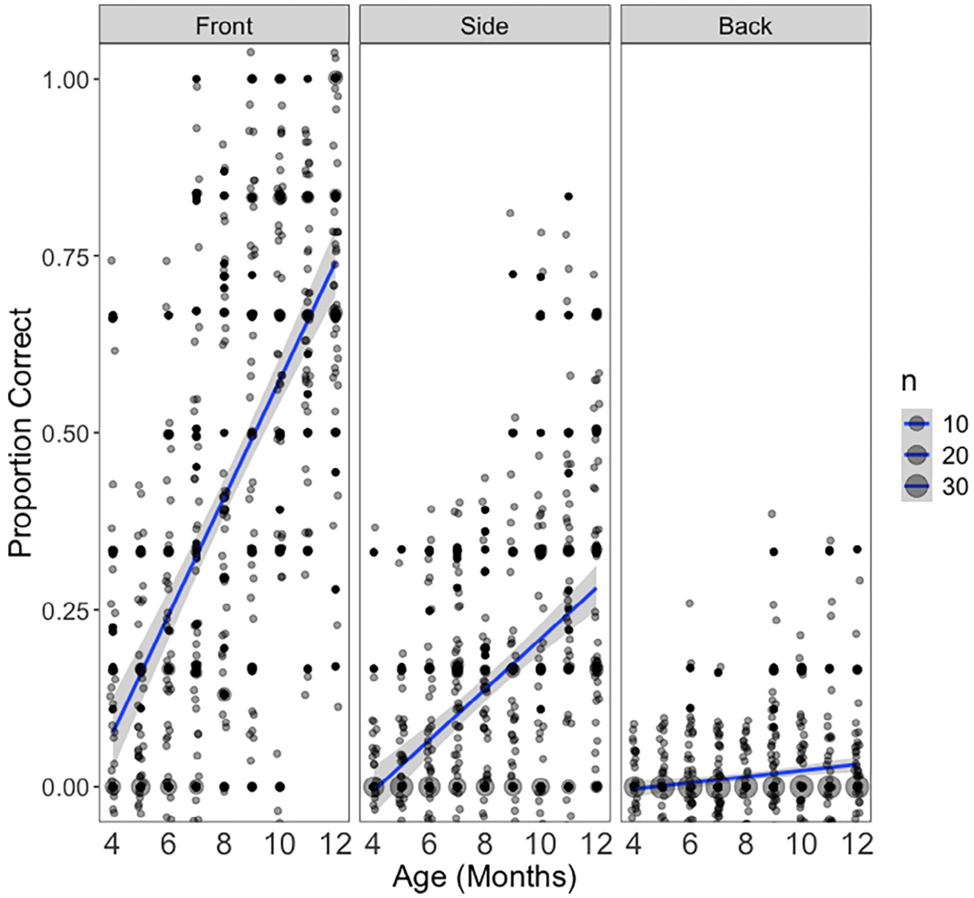

FIGURE 5 Hit proportions by month by target location (front, side, back). Note: Points indicate cue-type-weighted hit proportions (stringent criterion) from sessions with >= 17 total valid trials, for each month (X axis), at each target location: Front (left panel), Side (middle), and Back (right). Blue lines: best-fitting linear regression trends, with 95% CI cloud.

3.8 | Analysis of practice effects

To test whether practice effects contributed to age-related improvement, we compared the proportions of weighted hits between the main and comparison groups, at 9 months (i.e., the main sample's fifth session vs. the comparison group's first), and at 12 months (eighth session vs. second). At 9 months one comparison infant completed only one trial and was excluded (final $n = 23$), and at 12 months two infants dropped out, and one completed only 12 trials (final $n = 21$). A mixed-measure ANOVA indicates a significant group effect (main vs. comparison), $F(1, 49) = 17.84$, $p < .001$, but no group by age interaction, $F(1, 49) = 0.34$, $p = .56$. Notably, as Figure 7 shows, the group effect is in the *opposite* direction as predicted for practice effects.

3.9 | Individual differences

Longitudinal stability. Stability of AF was moderate and inconsistent from 4 to 12 months, with inter-month correlations (Pearson *r*s) ranging from -.23 to .65. Figure 8 shows correlations between months above the diagonal, scatterplots below the diagonal, and linear trends for each month on the diagonal. Only months 7–12 are shown, as the distributions from 4 to 6 months show large floor effects.[13] The positive skewness persists in 7–8 months, further reflecting the late acquisition of AF, and individual infants' AF becomes more stable after 8 months, with low-to-moderate positive correlations (*r*s: .12 to .51). In fact, correlations between consecutive months from 8 to 12 months are all positive and statistically reliable (8–9 months, $r = .47$, $p < .05$; 9–10 months, $r = .51$, $p < .01$; 10–11 months, $r = .37$, $p < .05$; 11–12 months, $r = .47$, $p < .01$).

3.10 | Relations to motor indices

Early motor milestones (i.e., mean age of achieving three head and torso control skills) did not predict AF in later months (r = -.00 to .01, $p = .56$ [11 months] and $p = .91$ [12 months]), confirming that age of attaining these milestones did not predict AF responses. To rule out late motor maturation as a cause of late AF acquisition, we re-analyzed 4- to 6-month AF data with only infants who passed all milestones by 4 months (n = 17, 24, and 29 at 4, 5, and 6 months, respectively). Those infants produced mean weighted hit proportions of 0.039 (SD = 0.063), 0.041 (0.053), and 0.096 (0.099) at 4, 5, and

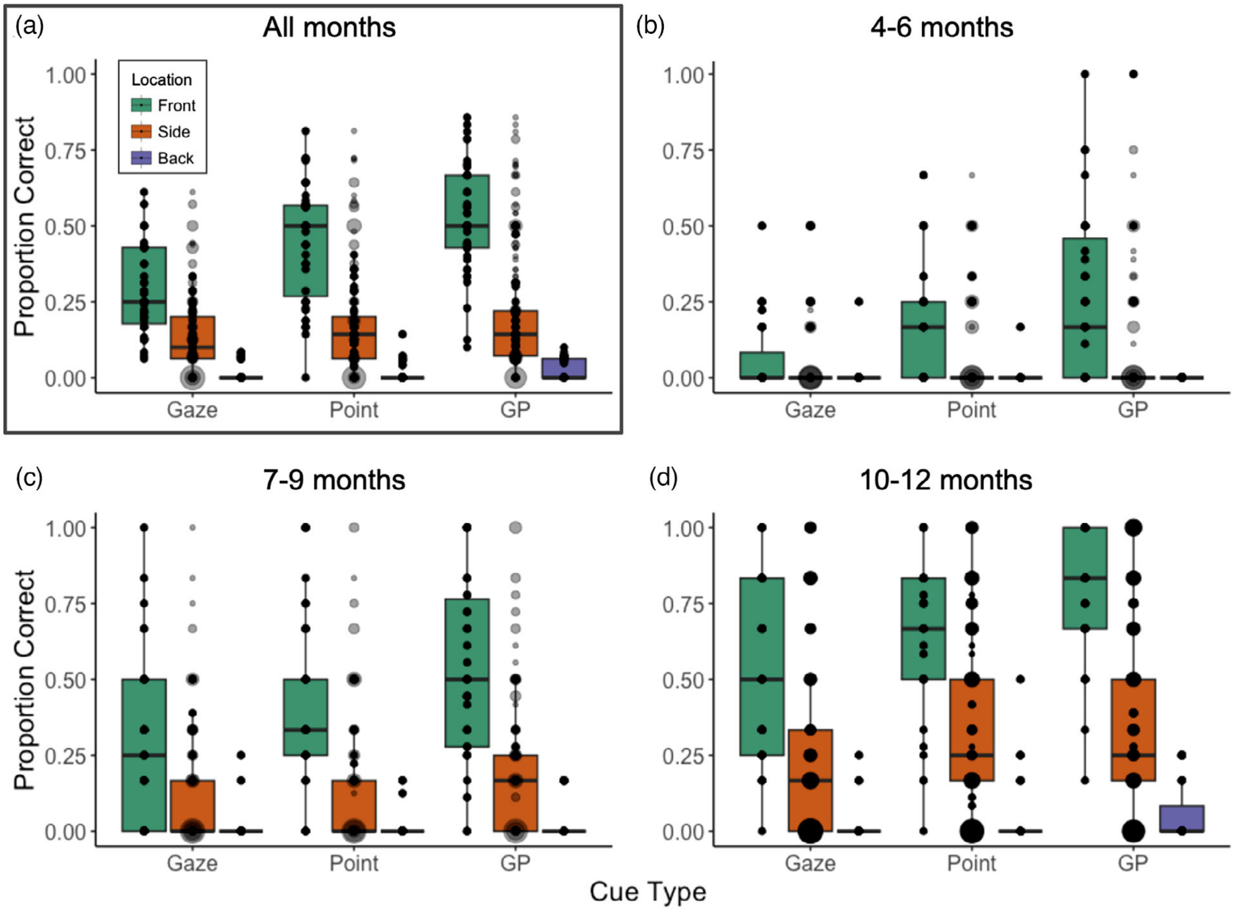

FIGURE 6 Hit proportions by cue type by target location. Note: Points and boxplots indicate weighted proportion hits (stringent criterion) from sessions with > = 17 valid trials in 4–12 months (panel A), 4–6 months (panel B), 7–9 months (panel C) and 10–12 months (panel D), for each cue type at each location (front: green; side: orange; back: purple). The circle sizes represent the number of scores at that proportion, and circles are semi-transparent and jittered, so larger and darker circles represent more scores (i.e., cases).

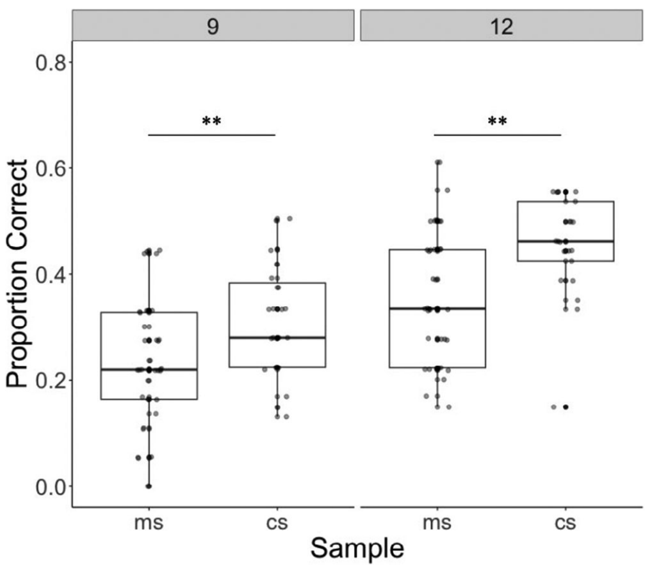

FIGURE 7 Hit proportions by main and comparison groups at 9 and 12 months. Note: Total stringent hit proportions. MS = main sample; CS = comparison sample. *$p < 0.05$, **$p < 0.01$.

6 months, respectively. These results and the orientation trial results together offer no evidence that lack of AF from 4–6 months was due to motor immaturity.

4 | DISCUSSION

The literature on the development of attention following (AF) has been controversial. Although most studies suggest that AF emerges around 6 months or later, some researchers have argued that infants follow gaze as early as 4 months or even at birth. These discrepant claims are related to different methods: results supporting early acquisition used methods that confounded motion-tracking, and usually did not establish that infants' responses were specific to social cues like eye direction (e.g., Hood et al., 1998), making the results ambiguous (Deák, 2015). Laboratory studies that controlled for these confounds have not found AF before 6 months.

The debate about age of acquisition is consequential because some authors have argued that attention-sharing is evidence of cognitive intersubjectivity (e.g., Tomasello et al., 2005), and even that attention-sharing shows the "innateness" of social cognition (e.g., Trevarthen,

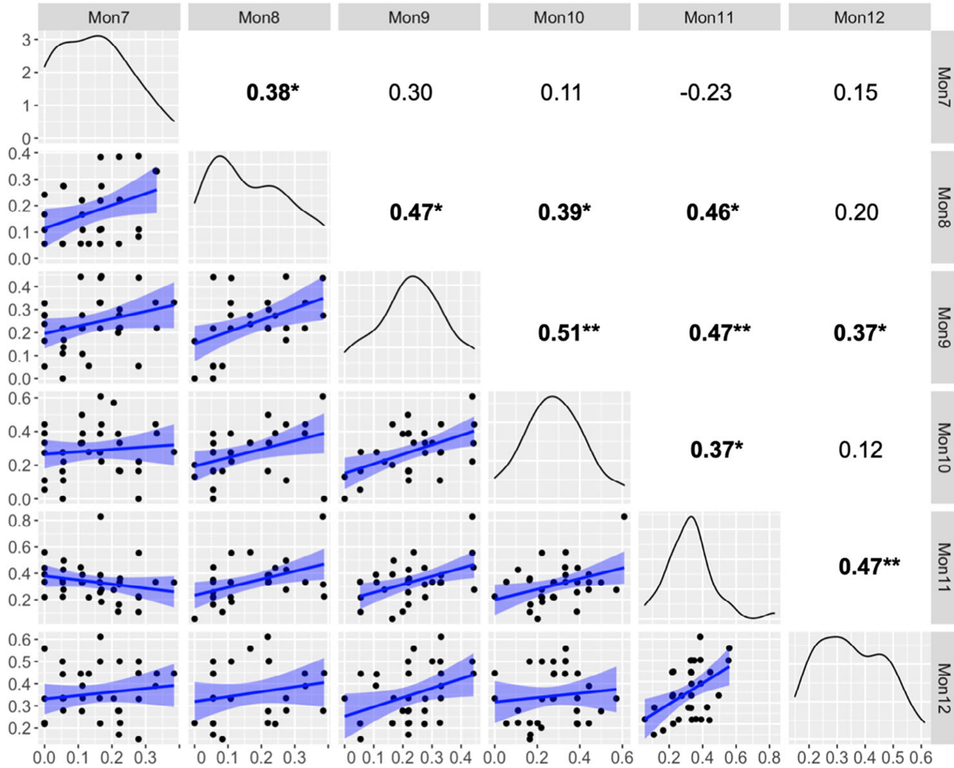

FIGURE 8 Correlations of hit proportions from 7 to 12 months. Note: Above the diagonal: Pearson's correlations between weighted stringent hit rates between months. Below the diagonal: Scatterplots of individual means between months, with lines showing the best-fitting linear trends (shaded region: 95% CI). Data include infants with valid data in both months. Curves on the diagonal: smoothed histogram by month. Axis labels at left and bottom indicate scale. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

2011)—notwithstanding the conceptual error of attributing behavioral patterns to "innate" causes (e.g., Spencer et al., 2009). Thus, broad claims about social development have been informed by claims about the development of AF. To resolve this debate we carried out a longitudinal experimental study of AF from 4 to 12 months, focusing on gaze- and point-following behaviors. This is one of a few longitudinal studies of attention-sharing development, and the first to implement controls for three factors that could interfere with infants' AF: specifically, motivation, attentiveness, and motor skills; and for one possible confound: infants' baseline tendency to look at targets.

To ensure that infants were motivated to follow gaze and point cues, thereby making lack of motivation a less plausible explanation for negative findings, we used a modified Butterworth paradigm in which infants first learned that monitors in six target locations could play reward videos after social cues. To rule out inattentiveness, trials did not start unless the experimenters (EXP and EXP2) judged that the infant's attention was focused on EXP. Also, cues were produced relatively slowly, to simulate naturalistic infant-directed "motionese" (Brand et al., 2002), and were repeated to further ensure attentiveness. To confirm that our task did not impose motor demands beyond younger infants' abilities, we tracked infants' early motor milestones. Additionally, orientation trials established that infants could orient to each monitor. Finally, infants were given ~10 sec to respond—slightly longer than infants' slowest AF responses in a naturalistic dataset (Deák et al., 2014)—to ensure that they had enough time to respond.

These measures ensured that infants attended to every cue, and had the time and ability to respond to the cues, effectively ruling out explanations for negative results including disinterest, inattentiveness, slow encoding, slow responding, or motor immaturity. To verify this, behavioral recordings (30 f/s) were made from 5 locations (in each corner, and overhead) and coded frame-by-frame by trained observers who annotated every visual fixation in every trial. This micro-behavioral coding of infants' visual fixations and shifts makes this an unusually detailed study of infants' AF to multiple spatially-distributed locations within a laboratory setting.

The control measures seem to have been effective: almost no trials had to be dropped due to inattentiveness. When infants did not follow a cue, they virtually always turned to the target monitor after the reward video began, confirming that they could reorient to, and were motivated to watch, the target video. Infants also almost always looked at target monitors during orientation trials, and remained oriented to the monitors for as long as the video played, further confirming the videos' reward value. Although these responses were less consistent at 4 and 5 months, that inconsistency did not predict infants' later AF, suggesting that modest early differences in orienting to target monitors did not predict AF. Also, motor immaturity did not predict infants' lack of AF in earlier months: age of reaching head and trunk control milestones did not predict infants' later AF; and regardless, even at 4 and 5 months infants successfully turned to most of the monitors during the orientation phase. This is the first study to rule out motivation, attentiveness, and motor development as factors that might have obscured younger infants' AF ability in previous studies. Thus, our results support the hypothesis that infants do not reliably follow specific gaze or point cues before 6 months because they have not yet learned that those cues indicate interesting distal stimuli.

Previous studies have not controlled for infants' base rates of spontaneous (i.e., uncued) looking at targets. Infants might fixate targets due to a tendency to visually explore their surroundings, and the monitors were among the most salient objects in a mostly bare room. To test this we added two baseline trials per session. However, spontaneous first-fixations to a specific monitor were so rare during baseline trials that no adjustment of scores was necessary. No previous study has controlled for this, however, and it remains possible that in studies with less-stringent definitions of "following," as well as fewer targets (e.g., Scaife & Bruner, 1975), spontaneous visual scanning was a non-trivial confounding factor.

In this paradigm infants could not solely use motion-cued scanning to find a target, unlike studies of younger infants (see Deák, 2015). Although direction-of-motion might have biased infants to turn to the correct side of the room, their first fixation target had to be the correct monitor out of three on that side of the room. This required more precision than simple motion-cued searching would support. Perhaps some correct fixations to front monitors were due to motion–cued scanning—this could partly explain the robust location effect (i.e., more following to front targets) found here and elsewhere (e.g., Butterworth & Jarrett, 1991; Deák et al., 2000; Flom et al., 2004). Yet from 4 to 6 months infants almost never followed cues even to front targets, suggesting that in settings with more numerous and spatially-distributed targets, motion-tracking is insufficient for precise AF responses.

Our modified paradigm tested divergent claims about the age of emergence of AF. This is the first longitudinal study of monthly micro-behavioral changes in AF from 4 to 12 months, and among the first to track developmental trajectories within and between individual infants. Furthermore, it is the first longitudinal study to document both cue-type (gaze, point, GP) and target-location effects on AF from 4 to 12 months, thus expanding previous reports (Butterworth & Jarrett, 1991; Deák et al., 2000, 2008).

The results show virtually no AF at 4 to 6 months. This confirms prior studies that partly controlled for motion-cued scanning by providing multiple ipsilateral targets (e.g., Butterworth & Jarrett, 1991). The results are also the first to show a slow linear growth curve from 7 to 12 months in following gaze, point, and gaze-and-point cues (see Deák et al., 2017), albeit with a steeper slope for the latter. The results further confirm limited gaze- and point-following even at 9 months, and increasing AF from 6 to 9 to 12 months (e.g., Butterworth & Cochran, 1980; Butterworth & Jarrett, 1991; Flom et al., 2004).

The results are among the first to assess longitudinal stability of individual AF during the first year. They show moderate inter-month stability starting at 8 months. Although a prior study found low-moderate stability of AF from 6 to 24 months (Morales et al., 2000), the many differences between the samples and methods of that study and this one make it difficult to reconcile the respective results.

Importantly, our sample size was modest, and limited to North American, English-using, predominantly-WEIRD infants—a population that is already over-represented in the literature (Nielsen et al., 2017). We cannot infer whether these results generalize to infants from other cultural and familial backgrounds, and other evidence indicates that infant-caregiver triadic interactions differ between WEIRD and other cultural groups (e.g., Bard et al., 2021; Mastin & Vogt, 2016).

4.1 | Gaze- vs. point-following

Although most infant AF studies have focused on gaze-following, several have implied that point-following emerges slightly later (Butterworth & Itakura, 2000), but becomes more effective by the second year (e.g., Deák et al., 2000, 2008). Point-following might become more effective than gaze-following for several reasons: e.g., points are more salient than gaze shifts (Tomasello et al., 2007), people use points to intentionally re-direct another's attention (Tomasello et al., 2007), and points have higher cue-validity for indicating interesting targets.

It has been difficult to adjudicate between these possible reasons partly because no study had directly compared the developmental trajectories of point- vs. gaze-following. Our study suggests that gaze- and point-following have similar trajectories from 4 to 12 months, with a non-significant, modest advantage for the latter. Of course, other tasks or contexts might reveal different patterns between gaze- and point-following across the first year.

Though infants did not follow point cues significantly more than gaze cues, the combination of gaze and pointing yielded more AF than gaze alone. Previous studies found that redundant cues predict infant AF. For example, Deák et al. (2008) showed that adding points to gaze cues increased following from ~10% to ~50% in 1-year-olds. Other studies reported that adding pointing or verbal cues to gaze cues increases infants' following (Butterworth & Grover, 1988; Leekam et al., 1998; Deák et al., 2008, 2017), consistent with the current findings.

4.2 | Target location

Our results also corroborate and extend findings that infants initially follow cues to targets in front of them, and later follow to peripheral and finally back targets (e.g., Butterworth & Jarrett, 1991; Deák et al., 2000). Infants typically do not follow cues to targets behind them until after 9 months of age, but may do so at above chance level by 12 months (Deák et al., 2000). One possible explanation is that it is easier to follow a cue to a target if both are simultaneously in the infant's visual field. By comparison, it requires more advanced spatial working memory to maintain a trace of the cue (e.g., head and/or arm angle) when turning away to search for a target (e.g., back targets). Another possibility is that motor and language skills facilitate attention-following to targets far from the infant's midline (i.e., back targets), perhaps because improved postural control and/or receptive and productive vocabulary promote continuity of social interaction. However, cross-cultural evidence indicates that infant motor development and face-to-face interactions are not necessarily positively correlated (Karasik & Robinson, 2022). Thus we cannot assume that these patterns generalize across culturally diverse populations. Finally, the energy cost of looking at the back target is likely higher than looking at front or side targets, possibly contributing to the target location effect (Wells & Davies, 1996)—although our orientation trial results, and infants' ubiquitous orientation to the reward videos even on non-following test trials, fail to support an energy expenditure account. In sum, these results confirm that infants rarely follow gaze *or* point cues to targets behind them even by 12 months, in laboratory settings, and this is not attributable to motor immaturity, failure to detect the cues, or lack of motivation to turn around.

4.3 | Practice effects

Previous studies showed that infants or children can increase AF responses through training (e.g., Whalen & Schreibman, 2003). Conversely, Deák et al. (2000) found in a split-half analysis that infants' AF declined in later trials if targets were simple and repetitive, suggesting that infants habituate to low-validity gaze and point cues. To test for possible practice effects we compared the main sample to a comparison sample tested at 9 and 12 months. The results show the opposite of practice effects: the comparison group performed better in both sessions. This might be due to habituation: perhaps the main sample grew bored in later months, or learned that they would see video rewards regardless of following. However, specific reward videos differed in every location and every session (which were a month apart), and a strong habituation effect seems unlikely in this scenario. It is also possible that the group difference reflects sampling error: the comparison sample might have been more predisposed to make AF responses. Although reasons for the difference remain unclear, it shows that practice effects cannot explain age-related increases in AF. Thus, the overall results suggest that AF is slowly learned across weeks and months and consolidates during the second half of the first year, but the monthly test sessions did not contribute to this growth; to the contrary, repeated testing might have attenuated the observed increase in AF from 9 to 12 months.

4.4 | Implications

Our results suggest that AF emerges from a gradual learning process. One possible objection to this conclusion could be that the slow group-level growth curve might be the result of averaging across individual infants' *nonlinear* trajectories with inflection points that fall at different ages. However, closer examination of individual trajectories does not support this possibility (see Figure S3): few individual trajectories show an obvious nonlinear inflection (i.e., a rising sinusoidal curve) across successive months.

The current results are consistent with arguments that AF can emerge from general learning processes operating on structured social events, constrained by general cognitive and affective phenotypes (e.g., Deák et al., 2013; Nagai et al., 2003; Triesch et al, 2006, 2007). The results support the late acquisition of gaze- and point-following in adequately controlled paradigms. The results also support predictions of bottom-up reinforcement learning and visual attention models (e.g., Jasso et al., 2012). For example, such models predict that infants should respond more to combined gaze and point cues than to either cue alone, and should respond more to targets within infants' visual field than outside it.

The results also address hypotheses about the relation of attention-following to other social and symbol-using skills. AF is eventually integrated with cognitive intersubjectivity, and Tomasello et al. (2007) claimed that "Pointing is a special gesture [that] requires…serious 'mindreading'." (p. 705). However, practical responses to others' points can begin in humans or apes before point *production* (see, e.g., Blake et al., 1994; Leavens, 2021). Moreover, general learning processes can explain point-following responses without assuming intersubjective representations. AF might nonetheless be a precursor of early intersubjective mental-state representations, and might contribute to such representations later in infancy (e.g., in the second year; Goubet et al., 2006).

Low levels of attention sharing are sometimes considered an early marker of childhood autism spectrum disorder (i.e., ASD; e.g., Osterling & Dawson, 1994). Previous studies reported that AF is reduced or delayed in children with ASD (e.g., Mundy et al., 1994). Triesch et al. (2006) suggest that the gradual associative learning process implied by the current data might be slowed in infants with non-optimal learning parameters (e.g. infants with ASD-like phenotypes). This is consistent with evidence that training can boost AF responses in preschoolers with ASD (Leekam et al., 1998; Whalen & Schreibman, 2003). Notably, the slow growth curve found from 7 to 12 months suggests an optimal age range during which AF training interventions might be most effective for infants at risk for ASD (see Grzadzinski et al., 2021).

4.5 | Limitations and future directions

One limitation of our paradigm is the lack of verbal cues during trials. Verbal cues effectively recruit or direct infants' shared attention in conjunction with gaze or pointing cues (Butterworth & Grover, 1988; Leekam et al., 1998). Deák et al. (2008) found that verbal cues improve AF over gaze cues alone in 1-year-olds, and Tang et al. (2023) found the mothers in unscripted interactions typically incorporate verbal cues in bids to elicit their infants' AF. Thus, it would be informative to extend this study by incorporating verbal cues. However, the use of verbal cues to AF seems to vary across cultures, and Mastin and Vogt (2016) reported that the association between joint attention and later vocabulary development was reversed in a small sample of non-WEIRD toddlers in rural settings. Thus, associations between language patterns, language development, and infant AF might differ across cultural groups with distinct infant-caregiver communication patterns. Future studies could investigate how parents use verbal cues in bids to share attention, and how this relates to infants' developing language skills.

Two other procedural limitations might have affected the results: first, infants might be less interested in video targets than in real world objects (Diener et al., 2008); second, isolated gaze and point cues are rare in naturalistic settings (Deák et al., 2017), and might reduce infants' AF. Another limitation might be that our criterion for counting AF hits was *too* stringent, and underestimated infants' AF responses. However, using the more liberal hit criterion (Tables S2, S3; Figure S1) reveals a reliably higher but still modest rate of AF—and, importantly, that increase is larger in later months (i.e., 9–12) than in earlier months. Notably, hit proportions using the liberal criteria averaged only 2.0%, 5.8%, and 6.4% higher than the stringent criterion at 4, 5, and 6 months, respectively. Thus, low AF levels through 6 months were not due to a too-stringent criterion.

Finally, AF was tested only in infants from North American, English-using, predominantly-WEIRD families. Previous studies showed that infants AF is moderated by demographic factors including socioeconomic status (Nielsen et al., 2017; Reilly et al., 2021). Also, attention-sharing and triadic interaction patterns differ systematically across cultures (Hernik & Broesch, 2019). To determine whether our results generalize to infants from other populations, further research should focus on families from more diverse socioeconomic and cultural backgrounds.

5 | CONCLUSION

The age of emergence of AF has remained controversial, partly due to differences in sampling, methods, and analyses across previous studies. This study used a modified paradigm that controlled for infants' motivation, attentiveness, motor maturity, and visual search-and-fixation baserates, to chart the development of AF behaviors in North American, mostly-WEIRD infants. Infants did not follow attention cues reliably before 7 months, and AF thereafter increased slowly and incrementally through 12 months. Infants also showed moderate longitudinal AF stability, but only from 8 months on. These results address open questions about the development of individual infants' AF, and its modulation by variables including cue type and target location.

ACKNOWLEDGMENTS

The project's human research participants treatment protocol was approved by the university's institutional review board. We thank Jochen Triesch for crucial intellectual contributions during many formative discussions, the students of the Cognitive Development Lab for assistance in data collection and coding, and the families who participated in this research. This research was supported by grants from the National Science Foundation (SES-0527756 to the senior author) and from the Univ. of California—San Diego Academic Senate.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data are available at https://osf.io/4nqbc/.

ORCID

Yueyan Tang https://orcid.org/0000-0002-6521-8604
Gedeon O. Deák https://orcid.org/0000-0001-8656-8796

ENDNOTES

[1] Scaife and Bruner's (1975) results are in fact uninterpretable: infants responded to only two trials, and no baseline visual search data were reported, so it is possible that infants responded to directional motion, and that their responses did not differ from default scanning biases.

[2] Hereafter we refer to the North American infants from predominantly WEIRD families of this study simply as "infants" for brevity, without presuming that the results generalize normatively to infants from other populations (see, e.g., Karasik & Robinson. 2022).

[3] Some of the data in this study (i.e., infants AF performance, unweighted, based on stringent criteria, between 6 and 9 months) are also reported in Tang et al. (2023).

[4] Almost all infants identified by parents as multi-ethnic were partly of Asian descent.

[5] EXP only produced cues if the infant was looking at her; and EXP2 verified this on every trial.

[6] Trials could be missing due to technical malfunctions, experimenter errors, or infant fussiness.

[7] Months were averaged to more closely approximate normal distributions.

[8] This also corrected for a systematic experimenter error in 8-month sessions, whereby infants were inadvertently given three front/GP trials, but only one front/point trial.

[9] Follow-up tests confirm that non-linear components (i.e., quadratic and cubic) are not significant.

[10] For proportion of hits using the liberal criterion: age factor $ß_{Std.} = 0.74$, $p < 0.001$, and 30% of random effects variance was due to individual differences. See Table S4 for liberal criterion LME results.

[11] To further explore the age-by-cue type interaction, another two-way ANOVA excluded all back trials due to floor effect. The results confirm the main effects of age ($F(8) = 49.2$, $p < 0.001$), cue type ($F(1) = 26.2$, $p < 0.001$), and the interaction between age and cue type ($F(16, 650) = 1.13$, $p = 0.33$).

[12] Another two-way ANOVA (as above) tested the cue type by target location interaction, excluding back trials due to the floor effect. This confirms the main effects of location ($F(2) = 509$, $p < 0.001$) and cue type ($F(2) = 27.4$, $p < 0.001$), and the interaction between location and cue type ($F(4, 200) = 13.1$, $p = < 0.001$).

[13] See Figure S2 for correlations among AF scores for all months.

REFERENCES

Astor, K., & Gredebäck, G. (2019). Gaze following in 4.5- and 6-month-old infants: The impact of proximity on standard gaze following performance tests. *Infancy*, *24*(1), 79–89. https://doi.org/10.1111/infa.12261

Ballard, D. H. (1987). Modular learning in neural networks. *Proc. AAAI*, 279–284.

Bard, K. A., Keller, H., Ross, K. M., Hewlett, B., Butler, L., Boysen, S. T., & Matsuzawa, T. (2021). Joint attention in human and chimpanzee infants in varied socio-ecological contexts. *Monographs of the Society for Research in Child Development*, *86*(4). https://doi.org/10.1111/mono.12435

Blake, J., O'Rourke, P., & Borzellino, G. (1994). Form and function in the development of pointing and reaching gestures. *Infant Behavior and Development*, *17*(2), 195–203. https://doi.org/10.1016/0163-6383(94)90055-8

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*, 72–83. https://doi.org/10.1111/1467-7687.00211

Butterworth, G. E., & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal* *of Behavioral Development*, *3*, 253–272. https://doi.org/10.1177/016502548000300303

Butterworth, G. R., & Grover, L. (1988). The origins of referential communication in human infancy. In (L. Weiskrantz Ed.), *Thought without Language*, pp. 5–21. Oxford University Press.

Butterworth, G. E., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*, 55–72. https://doi.org/10.1111/j.2044-835X.1991.tb00862.x

Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology*, *18*(Pt 1), 25–50. https://doi.org/10.1348/026151000165553

Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*, 79–151. https://doi.org/10.1017/S0140525X04000032

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior & Development*, *21*(2), 315–330.

Childers, J., Vaughan, J., & Burquest, D. (2007). Joint attention and word learning in Ngas-speaking toddlers in Nigeria. *Journal of Child Language*, *34*(2), 199–225. https://doi.org/10.1017/S0305000906007835

Corkum, V., & Moore, C. (1998). The origins of joint attention in infancy. *Developmental Psychology*, *34*, 28–38. https://doi.org/10.1037/0012-1649.34.1.28

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

Deák, G. O., Flom, R. A., & Pick, A. D. (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*(4), 511–523. https://doi.org/10.1037/0012-1649.36.4.511

Deák, G. O., Krasno, A., Jasso, H., & Triesch, J. (2017). What leads to shared attention? Action and gaze sequences in naturalistic infant-caregiver interactions. *Infancy*, https://doi.org/10.1111/infa.12204

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*(2), 270–281. https://doi.org/10.1111/desc.12122

Deák, G. O., Triesch, J., Krasno, A., de Barbaro, K., & Robledo, M. (2013). Learning to share: The emergence of joint attention in human infancy. In (B. Kar Ed.), *Cognition and brain development: Converging evidence from various methodologies* (pp. 173–210). APA.

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

de Barbaro, K., Chiba, A., & Deák, G. O. (2011). Micro-analysis of infant looking in a naturalistic social setting: Insights from biologically based models of attention. *Developmental Science*, *14*(5), 1150–1160. https://doi.org/10.1111/j.1467-7687.2011.01066.x

de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother-infant-object interactions. *Child Development*, *87*(2), 494–512. https://doi.org/10.1111/cdev.12464

DeLoache, J. S., & Chiong, C. (2009). Babies and baby media. *American Behavioral Scientist*, *52*(8), 1115–1135. https://doi.org/10.1177/0002764209331537

Demers, L. B., Hanson, K. G., Kirkorian, H. L., Pempek, T. A., & Anderson, D. R. (2013). Infant gaze following during parent–infant coviewing of baby videos. *Child Development*, *84*(2), 591–603. https://doi.org/10.1111/j.1467-8624.2012.01868.x

Diener, M. L., Pierroutsakos, S. L., Troseth, G. L., & Roberts, A. (2008). Video versus reality: Infants' attention and affective responses to video and live presentations. *Media Psychology*, *11*(3), 418–441. https://doi.org/10.1080/15213260802103003

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. https://doi.org/10.1080/13506280050144399

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology*, *85*, 199–212. https://doi.org/10.1016/S0022-0965(03)00022-5

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

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. https://doi.org/10.1016/j.infbeh.2003.09.007

Goubet, N., Rochat, P., Maire-Leblond, C., & Poss, S. (2006). Learning from others in 9-18-month-old infants. *Infant and Child Development*, *15*, 161–177. https://doi.org/10.1002/icd.446

Grzadzinski, R., Amso, D., Landa, R., Watson, L., Guralnick, M., Zwaigenbaum, L., Deák, G., Estes, A., Brian, J., Bath, K., Elison, J., Abbeduto, L., Wolff, J., & Piven, J. (2021). Pre-symptomatic intervention for autism spectrum disorder (ASD): defining a research agenda. *Journal of Neurodevelopmental Disorders*, *13*(1), 1–23. https://doi.org/10.1186/s11689-021-09393-y

Hawkins, R. D., & Byrne, J. H. (2015). Associative learning in invertebrates. *Cold Spring Harbor Perspectives in Biology*, *7*(5), a021709. http://doi.org/10.1101/cshperspect.a021709

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Hernik, M., & Broesch, T. (2019). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, *22*(4), e12779. https://doi.org/10.1111/desc.12779

Hood, B. M., Willen, D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, *9*, 53–56. https://doi.org/10.1111/1467-9280.00024

Jasso, H., Triesch, J., Lewis, J., & Deák, G. O. (2012). A unified account of gaze-following. *IEEE Transactions on Autonomous Mental Development*, *4*(4), 257–272. https://doi.org/10.1109/TAMD.2012.2208640

Jurkat, S., Köster, M., Hernández Chacón, L., Itakura, S., & Kärtner, J. (2023). Visual attention across cultures: Similarities and differences in child development and maternal attention styles. *Developmental Science*, e13368. https://doi.org/10.1111/desc.13368

Karasik, L. B., & Robinson, S. R. (2022). Milestones or millstones: How standard assessments mask cultural variation and misinform policies aimed at early childhood development. *Policy Insights from the Behavioral and Brain Sciences*, *9*(1), 57–64. https://doi.org/10.1177/23727322211068546

Leavens, D. A. (2021). The Referential Problem Space revisited: An ecological hypothesis of the evolutionary and developmental origins of pointing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(4), e1554.

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and cues: Gaze-following in children with autism. *Journal of Child Psychology and Psychiatry*, *39*(7), 951–962. https://doi.org/10.1111/1469-7610.00398

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd edn.). Wiley.

Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child–caregiver joint attention and language. *Social Development*, *9*, 302–315. https://doi.org/10.1111/1467-9507.00127

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, *43*(2), 235–264. https://doi.org/10.1017/S0305000915000148

Moore, C., Angelopoulos, M., & Bennett, P. (1997). The role of movement in the development of joint visual attention. *Infant Behavior and Development*, *20*, 83–92. https://doi.org/10.1016/S0163-6383(97)90063-1

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298. https://doi.org/10.1016/S0193-3973(99)00040-4

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behaviour and Development*, *21*(2), 373–377.

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology*, *6*(3), 389–401. https://doi.org/10.1017/S0954579400006003

Nagai, Y., Hosoda, K., Morita, A., & Asada, M. (2003). A constructive model for the development of joint attention. *Connection Science*, *15*(4), 211–229. https://doi.org/10.1080/09540090310001655101

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Reilly, E. B., Stallworthy, I. C., Mliner, S. B., Troy, M. F., Elison, J. T., & Gunnar, M. R. (2021). Infants' abilities to respond to cues for joint attention vary by family socioeconomic status. *Infancy*, *26*, 204–222. https://doi.org/10.1111/infa.12380

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

Spencer, J. P., Blumberg, M. S., McMurray, B., Robinson, S. R., Samuelson, L. K., & Tomblin, J. B. (2009). Short arms and talking eggs: Why we should no longer abide the nativist–empiricist debate. *Child Development Perspectives*, *3*(2), 79–87. https://doi.org/10.1111/j.1750-8606.2009.00081.x

Tang, Y., Triesch, J., & Deák, G. O. (2023). Variability in infant social responsiveness: Age and situational differences in attention-following. *Developmental Cognitive Neuroscience*, *63*, 101283. https://doi.org/10.1016/j.dcn.2023.101283

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*, 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

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

Wang, Q., Bolhuis, J., Rothkopf, C. A., Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in control: rapid anticipation of action outcomes in a gaze-contingent paradigm. *PloS One*, *7*(2), e30884. https://doi.org/10.1371/journal.pone.0030884

Wells, J. C., & Davies, P. S. (1996). Relationship between behavior and energy expenditure in 12-week-old infants. *American Journal of Human Biology*, *8*(4), 465–472. https://doi.org/10.1002/(SICI)1520-6300(1996)8:4%3c465::AID-AJHB6%3e3.0.CO;2-X

Whalen, C., & Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry*, *44*, 456–468. https://doi.org/10.1111/1469-7610.00135

Zukow-Goldring, P. (1996). Sensitive caregivers foster the comprehension of speech: when gestures speak louder than words. *Early Development and Parenting*, *5*, 195–211. https://doi.org/10.1002/(SICI)1099-0917(199612)5:4%3c195::AID-EDP133%3e3.0.CO;2-H

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tang, Y., Gonzalez, M. R., & Deák, G. O. (2023). The Slow Emergence of Gaze- and Point-Following: A Longitudinal Study of Infants From 4 to 12 Months. *Developmental Science*, e13457. https://doi.org/10.1111/desc.13457

APPENDIX

Calculation of monthly condition-specific hit rates for individual infants

For each infant in each month, we

1. first calculate the raw hit rates for 9 cue type by target location combination as follows:
   a. For each of 9 combinations (FrG [front, gaze]; SdG [side, gaze]; BkG [back, gaze]; FrP [front, point]; SdP [side, point]; BkP [back, point]; FrGP [front, GP]; SdGP [side, GP]; BkGP [back, GP]),
   b. Calculate the raw hit rate as the # of hits / total trials of that type (e.g., raw FrG hit rate = # FrG hits / total # FrG trials)
   c. This yields 9 raw hit rates, one for each type of trial defined by cue type and target location
2. Calculate the weighting coefficients for 9 cue type by target location combinations as follows:
   a. $w_{l,c}$ = coefficient for a given target location and cue type, in a given month
   b. $w_{l,c}$ = # of Target location (l) x Cue type (c) trials/# of all monthly trials for all infants
3. Calculate each individual infant's total weighted (i.e., trial-number-standardized) hit rate in a certain month (i.e., Hits(w_t) as:
   a. $\text{Hits}(w_t) = w_{Fr,G}{*}FrG + w_{Sd,G}{*}SdG + w_{Bk,G}{*}BkG + w_{Fr,P}{*}FrP + w_{Sd,P}{*}SdP + w_{Bk,P}{*}BkP + w_{Fr,GP}{*}FrGP + w_{Sd,GP}{*}SdGP + w_{Bk,GP}{*}BkGP$