Developmental Science
Developmental Science 17:2 (2014), pp 270–281 DOI: 10.1111/desc.12122

PAPER

Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects

Gedeon O. Deák,[1] Anna M. Krasno,[2] Jochen Triesch,[3] Joshua Lewis[1] and Leigh Sepeta[4]

1. *Department of Cognitive Science, University of California at San Diego, USA*
2. *Department of Counseling, Clinical, and School Psychology, University of California, Santa Barbara, USA*
3. *Frankfurt Institute for Advanced Studies, Goethe University Frankfurt, Germany*
4. *Division of Pediatric Neuropsychology, Children's National Medical Center, Washington, DC,USA*

Abstract

Infants gradually learn to share attention, but it is unknown how they acquire skills such as gaze-following. Deák and Triesch (2006) suggest that gaze-following could be acquired if infants learn that adults' gaze direction is likely to be aligned with interesting sights. This hypothesis stipulates that adults tend to look at things that infants find interesting, and that infants could learn by noticing this tendency. We tested the plausibility of this hypothesis through video-based micro-behavioral analysis of naturalistic parent–infant play. The results revealed that 3- to 11-month-old infants strongly preferred watching caregivers handle objects. In addition, when caregivers looked away from their infant they tended to look at their own object-handling. Finally, when infants looked toward the caregiver while she was looking at her own hands, the infant's next eye movement was often toward the caregiver's object-handling. In this way infants receive adequate naturalistic input to learn associations between their parent's gaze direction and the locations of interesting sights.

Research highlights

- Infants strongly prefer watching their parents manipulate objects over watching either their parents' face, or static objects, during naturalistic face-to-face play.
- Parents playing with infants focus on infants' faces and use object movement to attract infants' attention.
- Infants occasionally see their parent looking toward what she is holding; infants then tend to look in that direction. The consequent reward (i.e. seeing object-handling) could serve as input for infants to learn to follow others' gaze.

Introduction

Toddlers monitor and adapt their actions to the attention states of caregivers and other adults. During their first 2 years, human infants learn to share attention (i.e. attend to the same stimulus as a partner; Deák & Triesch, 2006). Attention-sharing is a fundamental element of face-to-face interactions (Argyle & Cook, 1976). For infants, attention-sharing skills facilitate language and action learning (Zukow-Goldring & Arbib, 2007) and support the capacity to infer others' mental states (Wellman, Lopez-Duran, LaBounty & Hamilton, 2008). Given the importance of attention-sharing in development, it is striking how little we understand about how infants acquire these skills.

Attention-sharing often involves *gaze-following*, or shifting gaze to intersect with another person's direction of gaze. However, it is not trivial to learn gaze-following. People frequently shift their gaze, and eye movements are affected by many contextual factors including the person's concurrent activity, their traits and states (e.g. emotions), and the properties of the setting (e.g. Foulsham, Cheng, Tracy, Henrich & Kingstone, 2010). This variety of contextual variables complicates infants' task of learning to discriminate and respond to other people's gaze direction and gaze shifts.

Address for correspondence: Gedeon O. Deák, Department of Cognitive Science, University of California at San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0515, USA; email: deak@cogsci.ucsd.edu

Moreover, infants only sometimes attend to adult behaviors. Thus, infants receive limited input about adults' gaze actions, and therefore have limited opportunities to learn the relations between adults' gaze and other properties of their complex everyday environments.

To explain how infants follow gaze, some psychologists have proposed that gaze-following is a specialized innate ability (Batki, Baron-Cohen, Wheelwright, Connellan & Ahluwalia, 2000). However, gaze-following might emerge from gradual, experience-dependent learning. Consistent with this, there is evidence that infants show progressive improvements in gaze-following skill. At 3 or 4 months, infants merely respond to an adult's left or right gaze shifts by scanning in the same direction, but only if a salient object is close to the adult's head (D'Entremont, 2000; D'Entremont, Hains & Muir, 1997; Gredebäck, Fikke & Melinder, 2010). By 6 months, infants occasionally follow gaze-shifts to targets further from the adult but still in their own visual field (Butterworth & Jarrett, 1991). By 9 months, infants sometimes follow adults' gaze to targets in their peripheral visual field (Flom, Deák, Phil & Pick, 2004), and by 12 months infants sometimes follow gaze to targets behind them (Deák, Flom & Pick, 2000). If this progression of gaze-following ability does indeed result from general learning processes, there is no need to stipulate additional specialized (e.g. innate) mechanisms. A recent theory has outlined how gaze-following might result from general learning, perceptual, and affective capacities (Deák and Triesch, 2006; Triesch, Teuscher, Deák & Carlson, 2006). These capacities include smooth attention-shifting (Butcher, Kalverboer & Geuze, 2000), a preference for faces and objects (Macchi Cassia, Turati & Simion, 2004; Goren, Sarty & Wu, 1975), and generalized learning processes. Those processes include habituation and reinforcement learning (Sutton & Barto, 1998; Botvinick, Niv & Barto, 2009). *Reinforcement learning* is the modern theoretical framework for describing classical and instrumental conditioning as well as higher-level planning abilities in animals and humans. In basic computational models of reinforcement learning, an agent registers the 'value' of its past actions in the context of particular states of the environment. Learning entails adapting action policies to the agent's 'history' of outcomes in various contexts, allowing for some exploration of new actions.

The plausibility of this theoretical framework for explaining attention-sharing abilities has been established through computer simulations (Jasso, Triesch, Lewis & Deák, 2012; Lewis, Deák, Jasso & Triesch, 2010; Triesch *et al.*, 2006; Triesch, Jasso & Deák, 2007). However, its plausibility has not been verified through infant behavioral data. One reason for this is that it is unclear what patterns of caregiver actions could facilitate infants' learning of gaze-following. To address this issue, we used *micro-behavioral ethnographic* methods (de Barbaro, Forster, Johnson & Deák, 2013a). These methods document types, rates, durations and sequences of behaviors (e.g. Hsu & Fogel, 2003). These variables in caregiver behaviors can powerfully modulate infants' social responses (e.g. Rochat, Querido & Striano, 1999).

The current study examined in-home play interactions between caregivers (CG) and infants. Caregiver and infant behaviors, including gaze direction and manual actions, were coded from synchronized videos. We focused on behaviors that were hypothesized to provide sufficient information for infants to learn gaze-following. To prove the sufficiency of parent-provided actions, several patterns must be demonstrated: (1) Infants must have the *opportunity* to observe a variety of parents' gaze 'states'; (2) Infants' *looking preferences* must allow them to register different amounts of accumulated reward by modifying their actions; (3) Parents' gaze-patterns must have *cue validity*; that is, their gaze direction must be related to, and allow prediction of, the location of infants' preferred stimuli. These three criteria require further explanation.

(1) *Opportunity*: A necessary condition for any learning account of gaze-following is that the infant has adequate opportunity to see different directions of, and changes in the direction of, adults' gaze (i.e. head and eye). That is, infants must have enough input to form some abstract association of different gaze directions. This knowledge must be abstract because parents look at different objects, from different positions or distances, in different settings or situations. Although this might seem obvious, there is some evidence that in complex settings infants seldom respond to adults' gaze shifts (Deák, Walden, Yale & Lewis, 2008). More generally, there are no data documenting how often infants in natural settings register their parent's gaze direction or gaze shifts.

(2) *Looking preferences*: Reinforcement learning models presume that agents' learning is driven by the amount of reward they obtain as a result of choosing different actions in various situations (Sutton and Barto, 1998). The reward value of infants' visual actions (e.g. gaze fixations or saccades) corresponds to how much they prefer (i.e. value) whatever stimuli can be seen wherver they are looking (Siqueland & Delucua, 1969). Over time, infants should learn to adapt their looking and attention-shifting decisions to obtain more visual reward. If some cue or contextual information is associated with the locations of preferred sights, infants might adapt their looking decisions to those cues. Such cues might include adults' actions. However, an

important question is whether infants strongly prefer looking at their parent's face. It is widely assumed that infants strongly prefer faces, especially their parents' (e.g. Hayes & Watson, 1981; Johnson, Dziurawiec, Ellis & Morton, 1991). If this preference is very strong, it would make a reinforcement learning account less plausible. Such an account would be unable to explain why infants systematically look away from their parent toward other locations. However, there is some evidence that infants' preferences for face and non-face stimuli vary considerably with age, stimulus properties, and task context (e.g. Robledo, Deák & Kolling, 2010). It is not known how much infants prefer to look at their parent versus other stimuli during play episodes in naturalistic settings. Thus, we cannot make assumptions about infants' everyday looking preferences.

(3) *Cue validity*: Even if infants are rewarded by visual stimuli other than parents' faces, it must be established that parents' gaze direction is associated with rewarding stimuli. Even if there is such an association, it would be effective only if infants actually see parents' gaze direction at the appropriate time – that is, just before looking at the rewarding stimuli, in the region specified by their parent's gaze direction. If infants seldom see this contingency, it could falsify a reinforcement learning account.

In sum, this study is designed to establish whether infants' responses to the range of action 'states' produced by their caregiver in the context of play interactions could provide infants with the *opportunity* to learn informed policies for looking around. It was also designed to determine what sights infants prefer to look at during these play interactions, in order to estimate relative *reward values* of infants' looking actions. Finally, it is designed to determine whether caregivers' gaze direction is statistically related to the location of infants' preferred (i.e. rewarding) sights – that is, whether caregivers' gaze direction has *cue validity.* If this final condition is met, it will indicate that infants' actions after seeing caregivers' gaze direction can provide enough information to learn a gaze-following policy.

Method

Participants

A sample of convenience comprising 35 dyads was recruited in San Diego County, California. Infants ranged from 3 to 11 months old. Caregivers (CGs) were English-speaking biological mothers. Another 22 infants were excluded for fussiness (n = 12), interruptions of the play session (n = 5), or poor video quality (n = 5). The final sample included 11 3-to 5-month-old infants (six female, five male; mean age = 128 days, range = 101–146), 15 6–8-month-olds (12 female, three male; mean = 213 days, range = 174–253), and nine 9–11-month-olds (five female, four male; mean = 304 days, range = 260–345). These age ranges reflect progressive changes in attention-sharing skills (see Deák and Triesch, 2006).

Demographic surveys were completed by all except for two CGs. CGs' mean age was 32.5 yrs (range = 23–41) and they had completed an average of 17 years of education (range = 14–22). All were married except for one. Infants were 60% first-born and 40% later-born; their ethnic backgrounds were Asian (6%), Caucasian (67%), Hispanic-Caucasian (3%), and multi-ethnic (24%). No infant had either a history of major medical problems or any diagnosed perceptual, cognitive, or communication disability.

Material

Dyads were video-recorded with Canon GL camcorders. NTSC videos were coded at 30 Hz using QuickTimePro and FinalCutPro. Several of the infant's own toys were used for object-play.

Procedure

Two trained researchers visited the dyad's home at a time when the infant was alert. After obtaining informed consent and identifying a suitable taping location (e.g. family room), the researchers video-recorded the dyad with one camera focused on the infant and one on the CG, angled so that the other actor and nearby toys remained visible (Figure 1).

Dyads were recorded in two phases. An initial 7–9 min period of free play on the floor allowed dyads to get accustomed to the cameras. Next, object play was recorded for 7–10 min. During this phase CGs sat facing their infant, with infants in a high chair or infant seat. CGs had 2–3 familiar infant toys that they could use to engage the infant. A substitution toy was placed further away. Other features of the environment (e.g. fixtures of the high chair) could also be examined or manipulated. CGs were asked to play as they usually did with the infant, and get them interested in the toys. The procedure was approved by the Institutional Review Board of the University of California, San Diego.

Coding

The main analyses focus on the object-play phase. Videos were synchronized and a researcher watched all sessions to mark and exclude periods of unusable data based on *a priori* criteria. These included: occlusion of

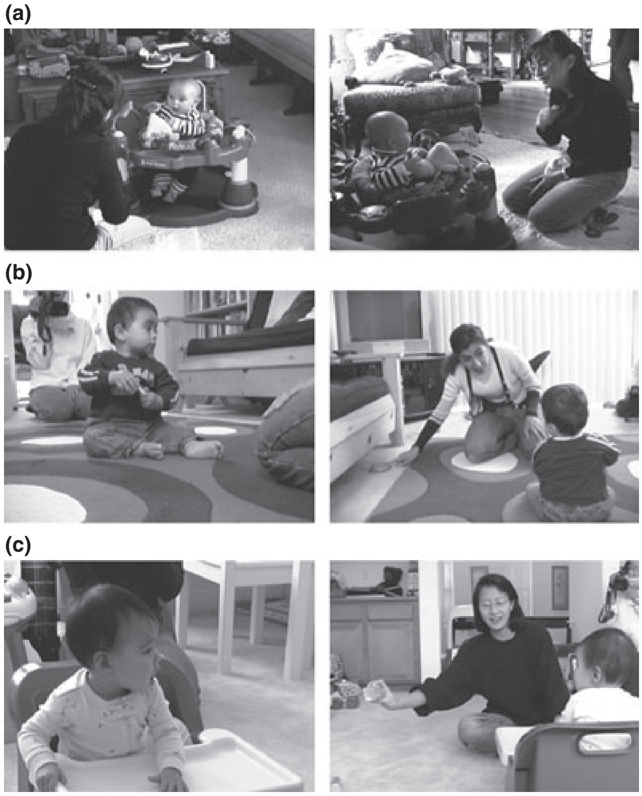

Figure 1 Examples of synchronized images of interactions between infants and caregivers. (a) Caregiver with empty hand looking at infant (5-month-old male) who is looking at caregiver (b) Caregiver touching object and looking at infant (12-month-old male) who is looking at object; (c) Caregiver holding an object to draw infant's attention (11-month-old female) who is tracking the object/hand.

faces, participant moving out of frame, or interruption (e.g. phone ringing; pet entering room). Exclusions were checked for accuracy by the first author. Synchronized videos were coded frame by frame (30 fps) by trained researchers, who marked all changes in any event category. This permitted temporally specific, time-locked queries for patterns of infant and CG actions. Action codes were continuous, mutually exclusive, and exhaustive. Due to the difficulty of coding very brief actions in ongoing unscripted social interaction, only events with durations of ≥0.1 sec were coded.

Coded behaviors included:

1 *CG direction and target of gaze*: Direction was coded in quantized horizontal arcs of ±36° centered on infant's midline: far right, near right, center, near left, far left, and back (±90-270°). Vertical direction was coded as up, center, or down (±60° each) relative to infant's face. *CG gaze target* categories were: static object, own hand, object-hand complex (when CG was holding or touching an object), infant's face, infant's body, or other/experimenter/camera. Fixations were timed from the first frame of static eye-direction to the first frame of the next fixation.
2 *Infants' gaze* direction was coded as above, except that horizontal direction was coded as center, left, right, or back. Targets could be: static object, object held/touched by CG, CG's face, CG's body, CG's hand, self (e.g. own hand or foot), or other (including researcher or camera).
3 *CG's manual actions* were coded as one of the above locations and one of the following states: empty hand, holding object, pointing, or 'extraneous motion' (e.g. tapping, waving, shaking).
4 Dyadic attention state was coded in continuous, mutually exclusive, exhaustive categories defined by the conjunction of independently coded current and previous infant actions, and CG gaze and hand actions. These definitions minimize subjective judgments of dyadic engagement. Dyadic state categories included three shared-attention states: infant-follows-CG, CG-follows-infant, and CG-imposed (i.e. object moved by CG to fill infant's visual field). Our analyses focused mainly on infant-follows-CG states, which were defined as the infant shifting gaze to the target of CG's gaze, point, or directed manual action. That is, even if CG was looking at the infant, if she pointed to or held up and shook or waved the object to attract the infant's attention, and the infant then looked at the object, this counted as CG-initiated shared attention. The scheme also included four non-shared attention states: infant-watching-CG while CG looked elsewhere, CG-watching-infant while infant looked elsewhere, infant and CG each looking at different targets, and mutual gaze. Details of most of these dyadic states will be described elsewhere.

For all episodes of infant-following-CG shared attention, we examined CG's gaze and manual action codes in the 5 seconds prior to the onset of the episode. This indicated the CG cues that could have promoted the infant's attention following. For example, if CG both looked and pointed toward a target within the 5 seconds before the infant shifted attention to it, then gaze and point were coded as possible cues. In order to check whether infants were gaze-following, we searched for infant-following events preceded by a CG gaze cue but no other cue.

A second independent coder re-coded a random sample of 23% of videos. Agreement, defined as selecting the same code within 0.1 sec (Cohen, 1960), was $\kappa = .75$ for CG's gaze, $\kappa = .75$ for infant's gaze, and $\kappa = .76$ for CG's manual actions. These values are high (Landis & Koch, 1977), especially given the large number of codes and the variability of unscripted naturalistic videos.

Data analyses

To determine *opportunity*, descriptive statistics on the distribution of CGs' gaze states over targets and locations were examined. Also, the distribution of infants' gaze and the range of attention-sharing states were examined. To determine *looking preferences*, infants' looking times to different types of stimuli were compared. Lastly, *cue validity* was tested by examining the relative frequency with which infants saw a rewarding stimulus just after seeing CG looking towards that stimulus. To determine whether these events could facilitate infants' learning of gaze-following, analyses were stratified by infant age. The sample was divided into three age groups: 3 to 5 months, 6 to 8 months, and 9 to 11 months. Any pattern seen in the youngest group would be a possible basis for learning, whereas any pattern seen only in the oldest group might be due to prior learning of gaze-following.

Results

To assess infants' opportunity to learn gaze-following from events in dyadic play, we classified caregivers' (CGs') gaze directions and actions relative to infants' attentiveness (i.e. whether or not they looked at CG). CGs averaged 5.1 gaze-shifts/min (SD = 2.1). However, infants spent only 12.8% of time looking at CG's face (SD = 9.5%; see below). If CGs' shifts and infants' looks to CGs were statistically independent, we would expect infants to have seen an average of only 0.65 CG gaze-shifts per min. In fact they saw an average of 2.15 per min (SD = 2.76), significantly more than expected, $t(34) = 3.22$, $p = .003$.

The distribution of CGs' gaze states is shown in Figure 2. CGs spent most of the time (mean = 80.6%, SD = 9.2%) looking at the infant's face. However, CGs also occasionally looked at static objects (mean = 10.8%, SD = 7.0%) or objects they were holding or touching (mean = 5.6%, SD = 4.4%).

Not only were CGs' fixations focused on the infant but their gaze shifts (i.e. changes in head pose) were spatially distributed in approximately a peaked Gaussian distribution around the infant's face, with the number of shifts averaging 9.2 (SD = 9.1) to (and from) far-right targets, 9.5 (7.7) to near-right targets, 42.7 (21.8) to center targets, 6.0 (6.8) to near-left targets, and 8.7 (9.5) to far left targets. They also made 1.4 (3.6) fixations above the infant, 28.1 (22.4) fixations to the level of the infant's head, and 17.7 (14.6) fixations below the infant. Thus, infants would have occasionally seen CG shifting gaze to one of those locations, but far more often would have seen CG turning back to look at them. From this evidence alone, it is not clear that infants received enough input to learn gaze-following.

To assess infants' looking preferences and patterns, we quantified their fixations to, and gaze shifts among, different types of targets. Infants engaged in considerable visual exploration, averaging 22.5 gaze-shifts/min (SD = 5.8). They showed a strong preference for watching their CG handle objects. Figure 3 shows infants' mean looking times to several target types: CG's face, CG's body, CG's hands, all static objects, objects held or touched by CG, the infant's own body (e.g. hands, feet), and other stimuli (e.g. experimenter, cameras). The data are divided into periods when CG was holding an object and periods when her hands were empty.[1] During the former ('holding') periods, infants preferred to look at the held object (mean = 46.9% looking time; SD = 20.0); their next strongest preference was for any static (non-held) object (mean = 29.4%; SD = 20.5%). This difference was significant, $t(34) = 2.6$, $p = .013$. Infants also spent more time looking at the held object than the sum of CG's face, CG's hands, CG's body, their own body, and 'other' targets (mean total = 14.4% of looking time, SD = 8.8): $t(34) = 8.8$, $p < .001$. Even when CG's hands were empty infants looked more at static objects (mean = 46.4%, SD = 21.1) than at the sum of CG's face, hands, and body (29.3%, 21.0), $t(34) = 2.6$, $p = .014$.

Age-related changes in looking times were tested in a mixed-model ANOVA with age group (3–5 months, 6–8 months, and 9–11 months) and hand state (empty or holding) between subjects, and object type (held/touched or static) within subjects. These conditions are represented in the 12 left bars in Figure 3. This analysis tests whether infants' strongest looking preferences, relative to whether CG was holding an object, changed with age. The interaction between age and object type (static vs. held/touched) was not significant, $F(2,32) = 1.3$. The interaction between age and CG's hand state was significant, $F(2,32) = 5.3$, $p = .010$ ($\eta^2_{\text{part.}} = .249$): older infants looked relatively more than younger infants at static objects when CG's hands were empty; however, older infants looked relatively less at held objects when CG was holding an object. However, neither of the simple age differences was significant: empty hands, $F(2,32) = 1.96$, $p = .157$; holding objects, $F(2,32) = 1.52$,

[1]Infants could look at touched objects even when CG's hands were empty because CGs sometimes touched objects without picking them up or actively manipulating them. We combined object holding and touching in this analysis because in either case infants were looking at a dynamic hand/object event.

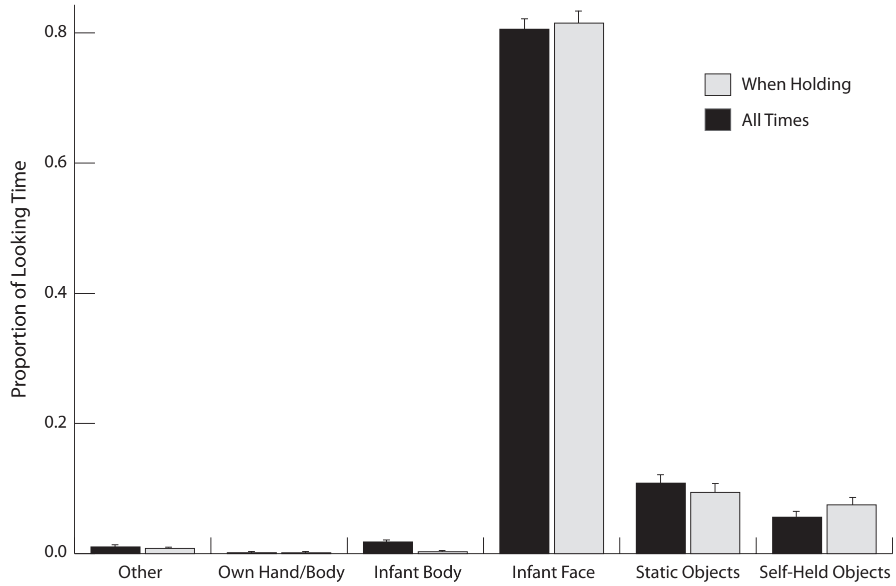

Figure 2 *Mean proportion of CG looking time (with SEs), for times when she was holding an object, and when her hands were empty. Targets: Other, Own Body/Hand, Infant's Body, Infant's Face, Static Objects, Held Object.*

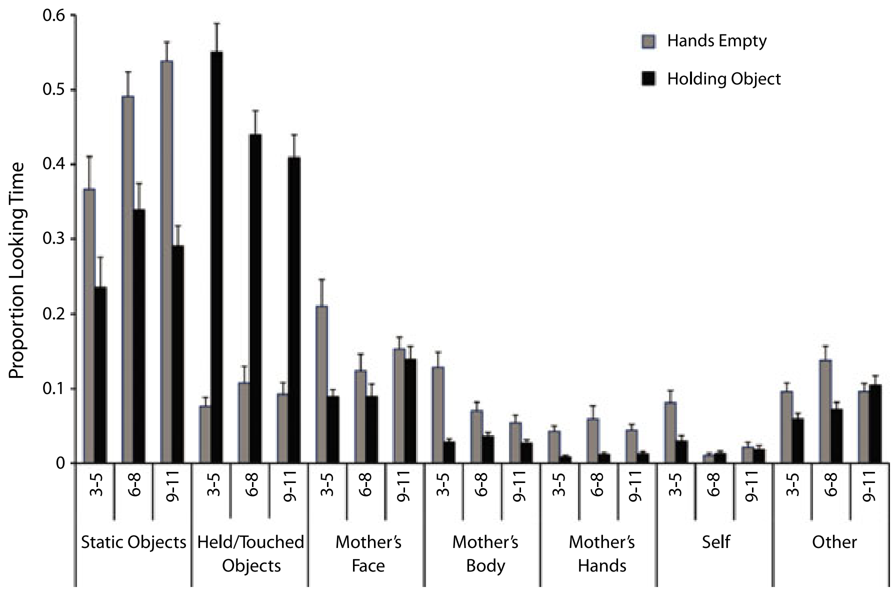

Figure 3 *Mean proportion of Infant looking time (with SEs), by age (3–5, 6–8, and 9–11 mo), for times when CG was holding an object, and times when CG was not ('Hands Empty'). Target categories: Held or touched object, all static objects (sum of 2–5 additional toy targets), CG's body, CG's Face, CG's hands, self (i.e. infant's own body/hands), Other (camera, researcher, environment).*

p = .234. The three-way interaction (age × hand state × object type) was not significant, $F < 1$.

To determine whether the youngest infants, 3- to 5-month-olds, had a reliable preference for watching CG-held objects, we examined their looking times when CG was holding objects. Three- to 5-month-olds looked longer at held objects than at all other static objects (means = 55.1% vs. 23.5% time), $t(10) = 2.3$, $p = .042$. They also looked longer at held objects than the sum of all other targets (i.e. CG, self, and 'other'; mean = 21.5%), $t(10) = 4.4$, $p = .001$. These effects are not corrected for the fact that there is only one held object at a time to see, compared to several available static objects and an innumerable number of other possible targets. Thus the preference for watching object handling was even more robust than the statistics indicate, even in the youngest group.

These looking-state statistics do not themselves indicate whether, and how much, dyads engaged in shared attention. In fact, dyads shared attention in some way for an average of 35% (SD = 10%) of the session. This includes time when CG was pointing to or holding an object for the infant to see, but watching the infant's face. (This is why shared-attention time exceeds the proportion of time CG spent looking away from the infant.) These intervals usually (mean = 84.5% of instances) occurred because infants followed CG's action (i.e. gaze, point, or object presentation), not because CG followed infants' gaze. Non-shared attention states averaged 51% of total time (SD = 12.0%), and mutual gaze states averaged 13% (SD = 9.0%). Dyadic attention states changed an average of 31.7 times/min (SD = 7.2). The rate of state-changes increased with age, r = .377 (p = .026), but even 3- to 5-month-olds and their CGs changed dyadic state an average of 27.6 times/min.

Given that dyads sometimes shared attention, but in many different particular states and following a wide range of possible prior states, it is unclear whether CGs' gaze served as a valid cue to her focus of attention. This is especially uncertain because CGs usually watched infants' faces and infants seldom looked at CG's face. Thus, infants might not have witnessed enough valid CG gaze cues to learn to follow gaze. To assess cue validity, we counted three types of sequences:

1 If the infant fixated CG's face as CG looked at her own hand holding/touching an object, and the infant looked towards the CG-held object in the next fixation, this was coded as a *high-reward outcome*.
2 If the infant fixated CG's face as she looked toward a static object, and the infant then looked towards that static object, this was coded as a *lower-reward outcome*. In these events infants presumably were rewarded for turning in the direction where CG's head was pointed. These were compared to another critical event type:
3 If the infant fixated CG's face and then turned in a *different* direction and saw a rewarding sight there (usually a static object), this was coded as an *error reinforcement outcome*. That is, the infant was rewarded for turning away from the gaze cue. Because the dyads were interacting in cluttered environments with many objects and features, such sequences were plausible.

All instances of infant-following-CG attention-sharing were classified as high-reward, low-reward, error-reinforcement, or ambiguous. Ambiguous events were those in which CG's and infant's gaze targets were not identical but were in nearby locations, or CG was looking at an object in the same region as the infant's head, or the infant's gaze shift was too small to assume that there was an action/reward contingency. The results, shown in Figure 4, reveal 7.3 times more reward outcomes than error reinforcement outcomes (discounting ambiguous events).[2] This demonstrates the validity of CG gaze cues after infants see their CG looking away from them, their next look is usually in the same direction, and their next target is what they find most rewarding: CG's manipulation of an object.

This regularity could be assumed to teach infants gaze-following behaviors only if the pattern holds true for the youngest infants. This is the case: 3- to 5-month-olds experienced 5.9 times more reinforcement events (0.35/min; 83% higher-reward) than error reinforcement events (0.06/min). Thus, the youngest infants experienced input that provided valid cues to learn gaze-following.

It could be argued that inter-individual variability in infants' preference for held versus static objects makes our definitions of 'higher' and 'lower' reward too general. However, the results suggest that reinforcement learning could have worked despite this variability. There was a moderate positive correlation between the magnitude of infants' preference for held versus static objects, and CGs' proportion of time looking at held versus static objects, $r(34)$ = .48, p = .003. Furthermore, 77% of weak reinforcement events were experienced by the 37% of infants who spent the most time looking at static objects. Thus, infants who were relatively highly rewarded by looking at static objects had relatively more opportunities to learn correlations between CG head poses and the locations of rewarding sights.

A reasonable question is whether the rewarded sequences were in fact instances of gaze-following. This is plausible but unlikely: gaze-following was infrequent,[3] occurring only 12 times (one in 3–5-month-olds, four in 6–8-month-olds, seven in 9–11-month-olds), or 0.08 times/min (see Deák *et al*., 2008, for converging evidence). Thus, the sequences were not gaze-following *per se*.

[2] The reported contingency results were based on a conservative 3-second window, but similar results are obtained using a broader 5-second window.

[3] Gaze-following was defined as the dyadic state changing from infant looking at CG while CG looked at something else, to infant looking at the same target as CG, if CG's gaze was the only cue in the last 5 sec. These cases were irrespective of the target's reward value. By contrast, reward outcomes were defined by CG's gaze direction and the reward value of infants' first target after CG's face, regardless of any other cues CG produced.

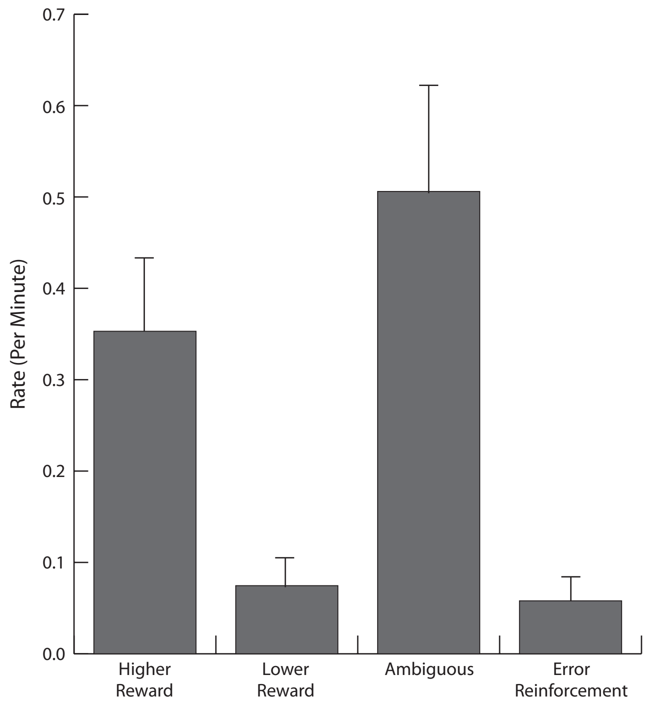

Figure 4 *Mean (and SE) rate of higher-reward (i.e. infant sees CG's face, then CG's held object), lower-reward (infant sees CG's face, then static object of CG's attention), error reinforcement (i.e. sees CG's face, then object in a different location), and ambiguous events.*

Discussion

The results suggest that infants could learn gaze-following in the course of event sequences that naturally occur during dyadic interactions with parents. This is one of the first demonstrations that naturalistic social events could provide a sufficient teaching signal for infants to learn a complex social skill. The conclusion rests on findings that the interactions provided adequate opportunity to observe different gaze states, that learning was supported by infants' looking preferences, and that infant-observed patterns of caregiver looking provided a teaching signal for gaze-following.

Regarding the question of opportunity, although infants spent only a small proportion of time looking at their caregiver's face, they still had sufficient opportunity to observe a wide range of caregiver gaze directions. Infants saw more gaze shifts than expected based on the amount of time they looked at CG's face and the amount of time CGs looked away from them. Although most of CG's gaze shifts were back to the infant (who was seldom looking at her), CGs also shifted gaze to various locations around the infant, to fixate on both held and static objects. When CGs were *not* looking at the infant, 91.2% of their time was spent looking at an object they were holding/touching (40.3%), or at a static object (50.9%) – usually one they were about to grasp or had just released. Also, when CGs were looking at the infant and holding an object, they moved the object to attract the infant's attention. Thus, CG's actions were structured, both intentionally and unintentionally, to increase infants' exploration of objects in various places in the dyad's shared environment.

Regarding the hypothesis that infants' looking preferences could support learning of gaze-following, infants clearly preferred watching adults handle objects, and had a secondary preference for static toys. This finding is at odds with the popular assumption that infants strongly prefer to look at faces, and particularly their mother's face. However, other recent findings support this: for example, older toddlers seldom fixate on their mother when playing with objects (Smith & Yu, 2008; Amano, Kezuka & Yamamoto, 2004). Moreover, individual differences in infants' preferences were reflected in their social environment: infants most interested in static objects had CGs who looked more at static objects. Thus, for these infants the outcomes defined as 'lower reward' were roughly as interesting as those defined as 'higher reward', but nonetheless their parents' looking patterns supported reinforcement learning.

There was also an interaction between infants' age and the strength of their preference for held/touched versus static objects: infants' preference for held/touched objects versus static objects attenuated with age when parents were holding the object, whereas their preference for static objects strengthened slightly when no object was held. Neither of the main effects was significant, however. Possibly the interaction stems from infants' increasing ability to retrieve and manipulate objects themselves (Lockman, 2000). That is, as infants' manual skill increases, they might focus attention on objects that they can grasp and explore. Consistent with this hypothesis, a recent longitudinal study showed that from 4 to 12 months infants attended progressively less to mother-held objects and more to self-held objects. This shift correlated with infants' increasing manual skills (de Barbaro, Johnson & Deák, 2013b).

The age-by-preference interaction notwithstanding, almost every infant of any age strongly preferred seeing their CG handling an object to any other sight, including static objects. This is apparent from the distribution of looking times when CG was holding an object (Figure 2). The tendency is even more robust if we consider that there was typically only one held object to look at, versus several static objects. Thus infants' 'per object' attention was strongly biased toward held objects. This attraction to held objects was related to CGs' manipulation of the objects: parents typically manipulated or at least moved (e.g. waved or shook) the objects they were holding. Although the effect might instead be due to the additive salience of the object and the parent's hand in proximity, this is unlikely. Infants almost never focused on mother's empty hands (Figure 2) even though the hands were almost always visible and were frequently moving. Thus, this explanation requires a post-hoc assumption that the salience of objects and hands combines by some non-additive function.

Although infants' preferences in this context were robust, we cannot infer from these results that infants are most interested in seeing adults handle objects in every context. In other settings infants might find other sights most interesting. For example, an infant watching two adults in conversation will see each adult looking at a moderately interesting stimulus – specifically the face of the other adult. The infant might obtain some reward by looking back and forth between the interlocutors. More broadly, by assessing infants' looking preferences in other contexts (e.g. conversations), we might learn whether and how those preferences can teach infants the relations between their caregivers' social actions (e.g. gaze shift; gestures) and the probabilistic structure of the environment.

We cannot infer from the current data whether infants' preference for watching adults handle objects extends beyond 3 to 11 months of age. However, there is some evidence that older infants and toddlers also are more interested in parents' hands than in their faces (Deák et al., 2008; Smith & Yu, 2008).

Our results indicate that infants received valid information to learn gaze-following in the form of adults' gaze direction cues. When infants happened to look at their mother's face while she was looking at what she was holding, infants' next look was usually towards that toy. By contrast, infants almost never looked from their mother's face to a rewarding stimulus in a different location. The difference in the frequencies of these events provided the critical information to learn gaze-following contingencies.

An alternative interpretation might be that in these situations infants were exercising gaze-following skills, rather than experiencing event-contingencies that would contribute to the acquisition of gaze-following ability. That interpretation is implausible, however. Few of the critical event sequences involved strict gaze-following. When CG's gaze was the only cue indicating an object, infants very seldom turned to that object. Rather, when infants turned from the CG to the held object, the CG was typically handling and moving the object.

Although infants very seldom followed gaze, this conclusion could be challenged if only older infants experienced the higher-reward sequences (Figure 4). That is, if only 9–11-month-olds look from CG's face to her object-handling while she is watching her hands, perhaps that sequence is a consequence, not an antecedent, of gaze-following skill. However, the sequence was found in 3- to 5-month-olds as well. Because 3- to 5-month-olds show at most only primitive, context-restricted, and infrequent responses to adult gaze direction (D'Entremont, 2000; Gredebäck *et al.*, 2010), it is unlikely that the sequences were by-products of gaze-following ability.

In short, the results show that face-to-face play episodes could provide sufficient information for infants to learn gaze-following contingencies, without any specialized or innate knowledge about mental states (e.g. 'intention' or 'attention'), or indeed any understanding of visual perception or other people's looking behaviors. The results confirm simulation results from Lewis *et al.* (2010). In those simulations, an infant agent learned rudimentary gaze-following patterns from 'watching' a virtual parent in a simulated 3D environment. The virtual parent reproduced the looking and object-handling actions of the parents in the present study, so the pattern of actions was somewhat realistic. Importantly, the virtual infant had no knowledge of mental states or of gaze or gaze direction. The infant agent learned to follow gaze through reinforcement learning and habituation mechanisms operating on quasi-realistic patterns of social input (e.g. 'parent' looking and object-handling). Thus, formal simulations support the claim that infants could learn to associate parents' gaze direction with different locations through non-specialized learning mechanisms and patterned input.

If infants have sufficient information in everyday interactions to learn gaze-following, why do they require months to develop gaze-following skills (Deák and Triesch, 2006)? Part of the answer is probably that learning is slowed by infants' 'sticky attention' in the first 3 months (Butcher *et al.*, 2000). A simulation showed that slow attention-shifting could cause a delay in reinforcement learning of gaze-following (Triesch *et al.*, 2006). Our findings are also consistent with this hypothesis: 3- to 5-month-olds in this study shifted dyadic attention states less frequently than older infants, and thus had fewer opportunities to learn spatial associations with parents' gaze direction. Also, because the reward outcome event sequences were rare, infants might need weeks or months to accrue enough input to learn the associations.

The results do not exclusively support a specific Temporal-Difference Reinforcement Learning model. Reinforcement learning is consistent with a range of models and algorithms that converge on similar results for a given learning problem (e.g. Ng & Russell, 2000; Sutton and Barto, 1998). Regardless, a range of documented attention-sharing behaviors can be simulated in a biologically and psychologically inspired model that incorporates reinforcement learning, habituation, attention-shifting processes, empirically verifiable stimulus preferences, and a structured social environment (Triesch *et al*., 2006; Jasso *et al*., 2012; Teuscher & Triesch, 2007), but no higher-level knowledge. The model also predicts the existence of mirror neurons that relate another animal's gaze direction to one's own gaze-shifting actions (Triesch *et al*., 2007) – a prediction that has since been experimentally confirmed (Shepherd, Klein, Deaner & Platt, 2009).

Gredebäck et al. (2010), however, report results that they interpret as falsifying a reinforcement learning account. Specifically, infants follow strangers' gaze more than parents' (Striano & Bertin, 2005), yet the authors argue that infants should be more reinforced by following the parent's gaze. Thus the results seem to disconfirm a reinforcement learning model. However, that argument rests on the assumption that infants are reinforced more by attending to parents than to strangers. That assumption is empirically untested. In fact the opposite prediction can be made: most organisms tend to seek out some degree of novelty (formalized as the temperature parameter in reinforcement learning models; Sutton and Barto, 1998). Exploring new stimuli (e.g. a stranger's face) can produce a reward signal in animals (Hollerman & Schultz, 1998). This might account for infants' greater attentiveness to the gaze direction of strangers versus caregivers. Alternately, if infants do find their parent's face more reinforcing than a stranger's, they would be less likely to look away from the parent's face, which could also explain Gredebäck *et al*.'s results. These alternative predictions underscore the importance of establishing the baseline reward value of various stimuli or outcomes to make predictions about learning.

Few developmental psychologists have considered whether infants' complex social skills can be explained by reinforcement learning frameworks or other frameworks that incorporate only biologically grounded mechanisms and processes. Instead, many researchers prefer a *social-cognitive* framework for conceptualizing infants' attention-sharing and other social skills (e.g. imitation). In the social-cognitive framework, infants follow gaze because they infer other people's communicative intentions or meanings (Csibra & Gergely, 2009; Csibra & Volein, 2008). The current results do not disprove this framework, but they show that it imposes unnecessary assumptions and still fails to predict specific results. For example, that framework does not predict that infants prefer to watch static objects more than their parent's face, or that infants are attracted by a hand simply waving or shaking an object. More generally the social-cognitive framework as it stands cannot predict or explain any particular gaze shift or change in dyadic state, or the relation of those changes to any specific neural process. Any framework that *can* make these specific predictions (e.g.Grossberg & Vladusich, 2010; Hoffman, Grimes, Shon & Rao, 2006; Triesch *et al*., 2006), and relate them to neural learning processes, is *de facto* a preferable theoretical alternative.

The current study has several limitations. First, the attrition rate was high. This was partly due to the difficulty of collecting two streams of uniformly high-quality video data in diverse home environments, and partly due to the duration of the session over which infants had to remain seated and engaged. Second, although we endeavored to create a naturalistic infant–parent play interaction, it is unusual for parents to be video-recorded at home by researchers, or to receive instructions – however minimal – about how to interact with their infant. We do not know how these factors affected parents' and infants' interactions. Although infants rarely fixated on the cameras or researchers (see Figure 2), and it is unclear how the experimenters' presence could have caused the main findings, it is possible that there was some effect. Also, infants usually followed parents into joint attention states, rather than caregivers following infants. This might be because parents had been asked to keep the infants engaged. A third limitation is that behavioral coding in home settings cannot attain the same temporal or spatial precision as state-of-the-art laboratory-based recording systems. For example, direction of gaze was divided into fairly large regions, and small eye movements were not coded. Ongoing improvements in mobile technology, machine coding, and coding software will allow greater precision in future studies. Nevertheless, the current results provide new information about how infants and caregivers explore and interact with one another and with objects, and how these patterns could support social learning.

Acknowledgements

The study was supported by grants to G. Deák, J. Triesch and J. Movellan from the MIND Institute and the National Alliance for Autism Research, and by grants to G. Deák and J. Triesch from the National Science Foundation (SES-0527756). Thanks to Hector Jasso, Javier Movellan, and Yuri You for assistance, advice and feedback.

References

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**, 64–80.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. London: Cambridge University Press.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, **23** (2), 223–229.

Botvinick, M.M., Niv, Y., & Barto, A.C. (2009).Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective.*Cognition*, **113** (3),262–280.

Butcher, P.R., Kalverboer, A.F., & Geuze, R.H. (2000). Infants' shifts of gaze from a central to a peripheral stimulus: a longitudinal study of development between 6 and 26 weeks. *Infant Behavior and Development*, **23**, 3–21.

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

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, **13**, 148–153.

Csibra, G., & Volein, A. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, **26**, 1–11.

Deák, G.O., Flom, R., & Pick, A.D. (2000). Effects of target and gesture on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, **36**, 511–523.

Deák, G.O., & Triesch, J. (2006).The emergence of attention-sharing skills in human infants.In K. Fujita & S. Itakura (Eds.), *Diversity of cognition* (pp. 331–363).Kyoto: University of Kyoto Press.

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

deBarbaro, K., Forster, K., Johnson, C.M., & Deák, G.O (2013a). Methodological considerations for bridging micro- and macro-behavioral data in studies of social interaction.*IEEE Transactions on Autonomous Mental Development*, **5** (3), 258–270.

deBarbaro, K., Johnson, C., & Deák, G.O. (2013b). Twelve-month 'social revolution' emerges from mother–infant sensorimotor coordination: a longitudinal investigation. *Human Development*, **56**, 223–248.

D'Entremont, B. (2000). A perceptual-attentional explanation of gaze-following in 3- and 6-month-olds. *Developmental Science*, **3**(3), 302–311.

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

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.

Foulsham, T., Cheng, J., Tracy, J., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: effects of social status and speaking. *Cognition*, **117**, 319–331.

Goren, C.C., Sarty, M., & Wu, P.Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, **56**(4), 544–549.

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: a longitudinal study of gaze-following during interactions with mothers and strangers. *Developmental Science*, **13**, 839–848.

Grossberg, S., & Vladusich, T. (2010). How do children learn to follow gaze, share joint attention, imitate their teachers, and use tools during social interactions? *Neural Networks*, **23** (8–9), 940–965.

Hayes, L.A., & Watson, J.S. (1981). Facial orientation of parents and elicited smiling by infants. *Infant Behavior and Development*, **4**, 333–340.

Hoffman, M.W., Grimes, D.B., Shon, A.P., & Rao, R.P. (2006). A probabilistic model of gaze imitation and shared attention. *Neural Networks*, **19**, 299–310.

Hollerman, J.R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, **1**, 304–309.

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

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

Johnson, M.H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, **40**(1), 1–19.

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

Lewis, J., Deák, G.O., Jasso, H., & Triesch, J. (2010).Building a model of infant social interaction. Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 278–283).

Lockman, J.J. (2000). A perception-action perspective on tool use development. *Child Development*, **71**, 137–144.

Macchi Cassia, V., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science*, **15**(6), 379–383.

Ng, A.Y., & Russell, S. (2000). Algorithms for inverse reinforcement learning. International Conference on Machine Learning (pp. 663–670). San Francisco, CA: Morgan Kaufmann.

Robledo, M., Deák, G.O., & Kolling, T. (2010). Infants' visual processing of faces and objects: age-related changes in interest, and stability of individual differences. Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 2482–2487). Austin, TX: Cognitive Science Society.

Rochat, P., Querido, J.G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of proto conversation in early infancy. *Developmental Psychology*, **35**, 950–957.

Shepherd, S.V., Klein, J.T., Deaner, R.O., & Platt, M.L. (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proceedings of the National Academy of Sciences, USA*, **106**(23), 9489–9494.

Siqueland, E.R., & Delucua, C.A. (1969). Visual reinforcement of nonnutritive sucking in human infants. *Science*, **165**(3898), 1144–1146.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, **106**, 1558–1568.

Striano, T., & Bertin, E. (2005). Coordinated affect with mothers and strangers: a longitudinal analysis of joint engagement between 5 and 9 months of age. *Cognition & Emotion*, **19**, 781–790.

Sutton, R.S., & Barto, A.G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.

Teuscher, C., & Triesch, J. (2007). To each his own: the caregiver's role in a computational model of gaze-following. *Neurocomputing*, **70**(13), 2166–2180.

Triesch, J., Jasso, H., & Deák, G.O. (2007). Emergence of mirror neurons in a model of gaze-following. *Adaptive Behavior*, **15**, 149–165.

Triesch, J., Teuscher, C., Deák, G., & Carlson, E. (2006). Gaze-following: why (not) learn it? *Developmental Science*, **9**, 125–147.

Wellman, H.M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology*, **44**, 618–623.

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, **69**, 1–34.

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

Received: 19 September 2011
Accepted: 30 July 2013