Chapter 16: Origins of shared attention in human infants

Gedeon O. Deák and Jochen Triesch

From Shared Attention to Shared Language in Human Infants

Homo sapiens possess a unique behavioural system for social action and response, namely, language. Language permits action at a distance by transmitting messages with specific meanings from one individual's mind to that of another. It is a peculiar system as compared with other structures in the environment, because the information in language that specifies meaning is rather abstract and arbitrary. Despite—or perhaps due to—these characteristics, language is the prime medium for 'cultural ratcheting' (Tomasello, 1999) among humans. In cultural ratcheting, behavioural innovations (e.g. tools) spread through a group and are sustained and elaborated upon across generations. For a group to maintain a system of linguistic behaviours, each individual must be able to learn and adapt to the prevailing information structure.

Typically, most of the structure of language is learned within a few years of birth, when the human infants are dependent on and in near-constant contact with caregivers. One account for this is genetic determinism: the structure of the human genome makes the acquisition and use of language inevitable. However, there is ample evidence, too complex to summarize here, that nativist views of language development (e.g. the Chomskian 'Language acquisition device,' poverty-of-the-stimulus claims, and mass-media reports of a so-called "language gene") are either inadequate or blatantly incorrect (see, e.g. Elman et al., 1996; Pullum & Scholz, 2002). There is no doubt that some species-specific products and processes of the human genome are necessary for human language learning; however, these products are not sufficient to explain early language development (MacWhinney, 1999). Most developmental scientists agree that a more complete account of human language must carefully consider infant's social experience. Somehow the structure of social information acquired by infants facilitates language acquisition. Yet, this interdependency is incredibly complex. Social interaction in infancy reflects a complex and nuanced interplay between infants' neural learning processes, their perceptual-motor limitations and affective/motivational traits, and the many-layered structure of the social environment (Cole & Cole, 1996). How exactly does infant social experience support language learning?

There are multiple answers to this question. For example, some linguistic knowledge is acquired through the acoustic structure of utterances heard by infants (Jusczyk, 2000). In addition, there seems to be a causal relation between non-verbal social information and toddlers' assumptions about the meaning of others' language acts (Tomasello, 1999). The latter evidence suggests that infants' ability to *share attention* helps them achieve shared meaning. In other words, the tendency for infants and the adults they are communicating with to attend to the same things seems to help infants correctly infer what adults are talking about, and thereby enter the language community. How do shared attention skills emerge in infancy? How do they contribute to early language development? In the remainder of this chapter we address these questions by considering evidence from typically developing infants, infants with disabilities, juvenile nonhuman animals, and computer simulations.

Shared Attention in Human Social Cognition

Shared attention is defined as redirecting attention to match another's focus of attention, based on the other's behaviour. If, for example, you are at a café and your companion turns away from you to look towards the door, you might feel compelled to look and see who has just entered. If, on a hike, your guide points excitedly towards a distinct tree, you might look for an unusual animal or plant in that area. Such responses do not merely enhance social interaction. They reveal a peculiarity of human interaction. Human infants will subjugate their own interests to another person's apparent interest in some other stimulus. This early interest in external signs of others' mental states seems to be species-specific.

However, attention-sharing skills are not unique to humans. For example, adult members of several nonhuman primate species will turn around to see what another animal is looking at (Tomasello, Call, & Hare, 1998). Such data suggest that shared attention is not sufficient for human social intelligence (i.e. ability to represent mental states) and language, though it might be necessary. Perhaps evolution of the capacity for shared attention skills occurred independent of (and prior to) the evolution of language. A separate-evolution account is feasible because attention-sharing has multiple functions. Organisms with limited directional visual fields (e.g. primates) might benefit (in acquiring resources and avoiding danger) by using the behaviour of conspecifics (e.g. responses to an approaching predator or a delectable meal) as proxy information about seen and unseen information in the environment. Thus, shared attention skills such as gaze-following (discussed below) compensate for limitations in the primate visual system (i.e. limited visual field). These skills also reveal an ability to learn secondary associations between (or make inferences about) others' behaviours and events in the environment. These associations or inferences can be subtle. For example, human infants use their caregivers' emotional expression (joy or fear) towards an ambiguous object (e.g. remote-controlled robot) to modulate their approach-avoidance behaviours to the object. This phenomenon is called *social referencing* (Walden & Ogan, 1988). For those who do not consider this skill impressive, we point out that the most sophisticated machine face-processing systems (e.g., Bartlett, Movellan, & Sejnowski, 2002) can find faces in cluttered environments, or identify categorical facial expressions, but cannot approach the combination of these functions seen in typical human infants' social referencing.

Another function of attention sharing is to help infants learn what is important in their social environment based on the distribution of attention of older, more knowledgeable group members (Kaye, 1982). Attention-sharing will eventually help infants infer mental states (e.g. interests and attitudes) of other people, and facilitate shared understanding or common ground. These functions underscore the connection between attention-sharing and language. Even as early as the second year attention sharing is an integral part of language use (Tomasello, 1999). In language production, young children use attention-sharing to shape their messages based on inferences about what others can perceive (O'Neill, 1996). In comprehension, toddlers use others' attention-specifying behaviours, such as gaze and gestures, to interpret utterances. The idea that language and attention-sharing are closely integrated is compelling. Yet, we should not overlook the first function described above: using others' behaviours as secondary cues to events in the environment. Keeping this function in mind raises questions about how and why infants acquire attention-sharing skills. These skills might be acquired through learning processes, perceptual processes, and affective dispositions that are found in a wide range of species. This does not imply that attention-sharing is independent of language; on the contrary, human children will use any available skills and information to communicate with their conspecifics. However, attention-sharing and language might have evolved separately, with language bootstrapping off of existing attention-sharing capacities and subsequently refining them.

Table 1: Varieties of shared attention in human infant-caregiver interactions and its theoretical relevance to the social ecology of infancy

Variety of shared attention	Theoretical & ecological relevance
Event attracts the child's & the adult's attention	Does the coincidental shifting of attention to the same focus moderate ongoing attentiveness?
The adult joins in the child's attention	important for language learning (Dunham et al., 1993)
Child requests adult's attention	A lack of this behaviour is indicative of ASD. (Mundy et al., 1990)
Child monitors adult's attention; joins in on occasion	important for word learning (Baldwin, 1993) and interpreting events based on others' emotional displays (Walden & Ogan, 1988)
Adult recruits child's attention	crucial for teaching; possibly more frequent in non-Western cultures (Bakeman et al., 1985)

In outlining these theoretical concerns, we have referred to several specific forms of attention-sharing. These and others are explicitly described in Table 1 with some relevant questions or findings about each form. All the phenomena are described as occurring between an infant and a caregiver, although attention-sharing certainly is utilised by groups of various ages and relationships.

The following section explores the question of how attention-sharing emerges during infancy. We shall concentrate largely on the most-studied form of attention-sharing, namely, infants following an adult's visual attention or gaze. Adults also follow the gaze of infants, here our primary concern here is to explain how infants acquire the skill.

The Emergence of Shared Attention: Data and Theory

Overview; Survey of Ecological Factors

Although newborns lack the visual acuity to perceive faces in detail, by 9 to 12 months of age they can respond to adults' gaze shifts and pointing gestures by shifting attention to the indicated region. Thus, within the span of a year, attention-following skills develop from modest beginnings. Recent research has begun to outline the intermediate achievements in this process. For example, the probability of an infant following an adult's gaze increases between 6 and 12 months (Butterworth & Jarrett, 1991).

It must be noted that virtually all of the findings reviewed here come from experimental paradigms, which are characterized by unusual environments and interactions. For example, besides the adult model (either a parent or a stranger), the testing environment is usually stripped of interesting objects and organisms. We emphasize this point because it is known that ecological factors influence attention-sharing, including some factors that differ between everyday and experimental settings. Thus, caution must be exercised when interpreting the findings and age norms reviewed here. Experimental results might differ systematically from those obtained in natural settings and interactions.

In most experimental studies of gaze- and point-following, infants and adults are seated facing each other. Upon receiving a signal, the adult produces a cue or cues such as turning his or her head and eyes away from the infant to look directly at the target for 5–10 s. Pointing gestures are modeled by the adult lifting and extending the arm in a smooth movement to point directly and continuously at the target. Typically, trials are initiated by the adult calling the infant to draw the latter's attention. The room layout used in most studies (Butterworth & Cochran, 1980; Deák et al., 2000; Flom, Deák, Phill, & Pick, 2003; Scaife & Bruner, 1975), schematized in Figure 1 as an overhead view, has one or more targets on each side. (Early studies used only one target per side, but this yields ambiguous results). Correct gaze- or point-following requires infants to ignore an object that is closer to the front of their visual field (F in the figure) and to scan their periphery (P) or the area of the room that is behind them (B). One drawback with this arrangement is that target location is confounded with the size of the adult's gaze shift: a very small head turn is required from the adult to look at targets behind the infant. Deák et al. (2000) rotated 12- and 18-month-old infants 90° to correct this confound, as shown in Figure 2, and found significant independent effects of target location (i.e. less following to back targets) and magnitude of the adult's head turn (i.e. less following of small head turns than large ones).

A major determinant of infants' attention-sharing is the form of behavioural cues produced by the adult. It is far more effective for caregivers to point while looking rather than to merely shift gaze (Butterworth & Jarrett, 1991; Deák et al., 2000; Morissette, Ricard, & Gouin Décarie, 1995). Deák et al. noted several possible reasons for this: pointing is more noticeable than a simple gaze shift, possibly because the hand and arm motion subsumes a larger proportion of the visual field. Also, the pointing arm provides a more specific and salient directional cue (Butterworth & Itakura, 2000). Finally, pointing is intended to direct another's attention, whereas head pose is an incidental consequence of visual attention and is not always intended to direct another's attention. Any or all of these factors might contribute to the effectiveness of pointing, and there is some evidence for at least the first two explanations. Recently, You, Deák, Jasso and Teuscher (2005) reported preliminary quasi-naturalistic data suggesting that when parents pick up, wave, or tap objects, these actions are as likely to elicit

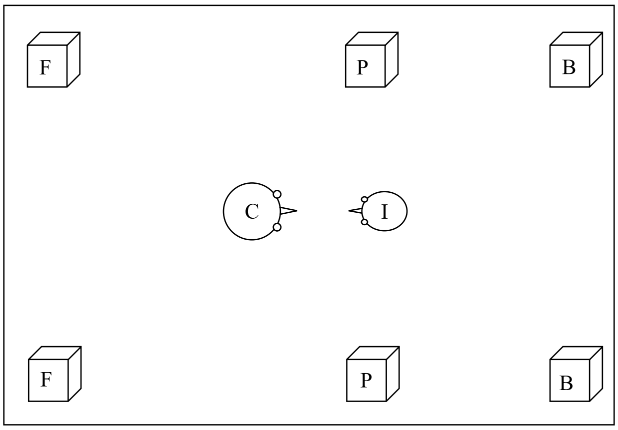

Figure 1.
Schematic overhead layout of the room used in studies by Butterworth and Jarrett (1991), Deák et al. (2000) and others. C = caregiver; I = Infant; F = frontal target; P = peripheral target; B = back target (all relative to the infant). In most studies, only two pairs of targets (e.g. left and right F and P) are present in a given trial.

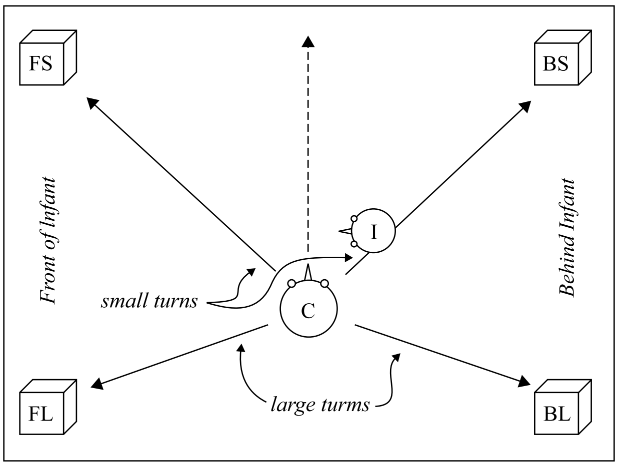

Figure 2.
Schematic overhead layout of the room used for Deák et al. (2000), Experiment 2, and Flom et al. (in press), Experiment 2. Note that caregiver C makes a small head turn (from midline) to bring one of the front (FS) targets or one of the back (BS) targets into view; the other targets (FL and BL) each require a larger head turn. Front and back are relative to the infant's midline visual field.

young infants' attention as when parents point at objects. This suggests that it is not pointing gestures per se, but the motion of an adult's raised or outstretched arm/hand, that captures infants' visual attention.

There is converging evidence that motion is an important ecological cue for infants to follow another person's actions or attention. The gaze-following of infants younger than 12 months depends on observing the adult's head movement rather than the final head pose (Moore, Angelopoulos, & Bennett, 1997). As might be expected, older infants and children learn to use a static head pose to infer another's direction of attention.

Other important ecological variables include object locations (relative to the infant) and features of the visual targets. Infants are more inclined to follow an adult's cues to targets in front of them than to targets in the periphery or behind them (Butterworth & Cochran, 1980; Butterworth & Jarrett, 1991; Deák et al., 2000; Flom et al., 2003; Morissette et al., 1995). Further, they are more inclined to follow an adults' cues to distinctive, complex targets than to repetitive, simple targets (Deák et al., 2000; Flom et al., 2003).

A less-studied but possibly critical ecological factor is the amount of competing information in the infant's environment. Most experimental studies have used stripped-down environments. However, one experimental study has examined conditions that are more realistically distracting (Walden, Deák, Yale, & Lewis, under review). One-year-old infants played with toys while the parents (seated such that their heads were always visible to the infants) periodically turned to look at a target, turned and pointed or used verbal cues to capture the infant's attention. One-year-olds rarely (<10% of the trials) followed adults' gaze, if that was the only cue. They did, however, follow gaze coupled with either pointing or verbalizations. Thus, when an informative environment competes with social information, infants' attention-following is reduced in predictable ways. The finding that infants rarely follow adults' gaze shifts in more naturally 'busy' settings has been replicated in observational study by You et al. (2005). However, the results obtained by Walden et al. also reveal that adult caregivers can compensate by producing more elaborate attention-getting behaviours: ongoing research will clarify how elaborate combinations of adult behaviours can recruit and redirect infants' attention in various circumstances.

Age-Related Changes in Attention-Sharing

Apart from ecological variables, the infant's own maturational status is a major determinant of attention-sharing behaviours. The qualitative change outlined above—from extremely immature vision at birth to sophisticated attention-following skills by 12 months—has been elaborated upon by experimental studies.

Some researchers believe that gaze-following begins very early in infancy; however, this rests on a definition of gaze-following that is too broad to be useful. An adult's horizontal gaze shifts can weakly trigger same-side attention-shifts in 3- to 5-month-olds (Hood, Willen, & Driver, 1998); however, this is apparently due to motion cueing (Farroni, Johnson, Brockbank, & Simion, 2000). No well-controlled studies have conclusively demonstrated gaze-following even in 6-month-olds, although in stripped-down experimental settings, it appears likely that some 6-month-olds do respond to adults' gaze shifts (Butterworth & Itakura, 2000) by turning to the same side of the visual field (Morales, Mundy, & Rojas, 1998). However, this simple same-side turning is hard to interpret, also because of motion cueing. A better method, as explained earlier, is to position multiple targets on either side of the infant, and test whether infants prefer to shift attention to the precise same-side target as the adult. In multiple-target designs, it is not until 9 months of age that infants tend to reliably follow an adult's gaze or pointing to targets in their frontal visual field. Some 9-month-olds will follow a combination of gaze and pointing gestures to targets in their periphery while ignoring same-side distracter objects. However, they do so reliably only when targets are distinctive and interesting (Flom et al., 2003). So far, no condition has been observed under which 9-month-olds will follow an adult's gaze to targets behind them.

There have been claims that point-following emerges later than gaze-following and that 9-month-olds are as likely to look at a pointing hand as they are to follow it. However, this claim is not well documented. No study has adequately investigated the separate and joint efficacy of pointing and gaze shifts in different ecological contexts for infants between 6 and 12 months. In such a study, controlling for (or experimentally manipulating) motion salience would be vital.

By 12 months of age, infants tend to follow an adult's gaze or pointing hand to targets behind them, even if there are same-side distracters nearby (Deák et al., 2000). This ability continues to improve through the second year (Butterworth & Jarrett, 1991; Deák et al.)

By 14–15 months, infants are sensitive to a line-of-sight constraint on others' visual attention. If there is a barrier between an adult's eyes and a target (Butler, Caron, & Brooks, 2000) or the adult closes his or her eyes (Butler & Meltzoff, 2002), 14- to 18-month-olds are less likely to follow the adult's gaze shift to a target on the would-be line-of-sight. However, this achievement should not be overstated because knowledge of line-of-sight constraints is limited in children even as old as 3 years (Flavell, Green, Herrera, & Flavell, 1991).

At approximately 15 months, most infants will point to interesting sights in order to recruit an adult's attention and will look at the adult as if to determine whether he or she is joining in. Even 12-month-olds have been observed to occasionally point (Leung & Rheingold, 1981). What does pointing and following others' gaze signify? A rich interpretation of these capabilities is that the infant is aware that others cannot see everything, and sometimes attend to different things than the infant. A sparser interpretation would be that infants' pointing enhances and prolongs social interactions, which they enjoy (Moore & D'Entremont, 2001). However, no evidence exists to favour one of these interpretations.

Many studies reveal that by 18 months, infants reliably follow gaze or pointing to in-sight and out-of-sight locations, and use verbal cues to moderate attention-shifts and take into account the adult's line-of-sight. This age milestone is relevant for two reasons. First, at approximately 18–22 months, the word learning rate of some infants accelerates, and most begin producing multiple-word utterances (Fenson et al., 1993). It seems as if they 'break the code' of predicate-object language. Second, there is converging evidence that 18–24-month-olds possess an explicit conceptual understanding that behaviours are caused by unseen feelings and mental states. At this age, for example, infants begin talking about feelings and mental states as precedents of observable behaviours (Bretherton, Beeghly-Smith, & McNew, 1981). Further, between 12 and 18 months, infants begin to represent other peoples' preferences as persistent traits (Repacholi & Gopnik, 1997). Finally, by 2 years, infants modify their requests based on their memory of what a parent has or has not seen (O'Neill, 1996). These suggest that toddlers relate a person's knowledge to his or her personal experiences, and use these inferences to form messages. In summary, between 18 and 24 months, infants' attention-sharing skills achieve greater sophistication, their language skills are consolidated, and their social cognition begins to incorporate mentalistic inferences.

The developmental changes in the attention-sharing skills outlined above are summarized as a timeline in Figure 3. Along with these changes, we list a few concurrent and possibly related traits or capacities, which are discussed next.

Development of Related Capacities

It is difficult to interpret changes in attention-following skills between 3 and 18 months without considering other concurrent developmental changes. Some of these changes appear especially relevant. For example, between 3 and 6 months, many fundamental visual capacities, including eye movements and accommodation, attention-shifting, visual field size and acuity, approach adult levels (Atkinson, 2000). It is possible, however, that the efficiency of some of these capacities continues to develop for several months and subtly affects the development of attention-sharing. Perception of pictorial depth cues, for

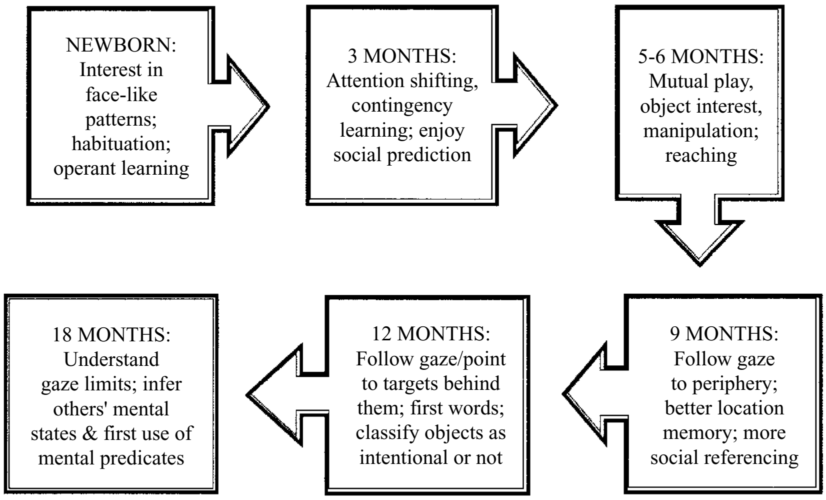

Figure 3.
Summary of developmental timeline featuring some age-related changes hypothesized in the current theory to be important for the emergence of attention-sharing skills. See text for explanation.

example, develops until at least 7 months (Arterberry, Craton, & Yonas, 1993), and might in some limit infants' ability to determine the target of an adults' gaze (e.g. if the target is fairly distant). However, this cannot explain the age differences in the aforementioned experiments, wherein kinetic and binocular distance cues are available (even very young infants can use these cues). Face processing is another critical perceptual skill gaze following, and some aspects of face processing continue to develop into late childhood (Carey & Diamond, 1994). The nature of face processing in very young infants remains controversial (e.g. Turati, 2004); however, most researchers agree that by 2–3 months, infants prefer faces to control stimuli (Johnson, 1997). A critical, but little-understood aspect of face processing for gaze-following is the discrimination of different head poses and eye positions. This is important because caregivers' gaze shifts to different locations will produce predictable changes in their head pose, and infants somehow learn to use this information for gaze-following. Sai and Bushnell (1988) found that 1-month-olds prefer seeing their caregiver's face in frontal pose rather than in profile, suggesting an early ability to discriminate extremely different head poses (see also Hains & Muir, 1996). This means rudimentary sensitivity to head pose precedes gaze-following by several months, although fine-tuning the discrimination of face pose and eye direction might continue into the second year.

Infants' representational capacity, particularly spatial working memory, might be important for establishing shared attention. For example, in gaze-following, when an adult focuses on a target behind the infant, the infant cannot keep both the adult and the target in view. Some memory trace of the adult's action, the object location, the infant's previous head-turning action (i.e., motor feedback), or a combination of these, is needed. Does the development of spatial memory limit the development of attention-following skills? No study has investigated this. However, it is known that by 9 months of age infants can sometimes remember the location of an object for long periods (Ashmead & Perlmutter, 1980). Also, spatial memory shows protracted development, with brief, partial and fragile representations as early as 4 months and far more robust spatial representations by the second year (see Haith & Benson, 1997, for review). This connection therefore remains viable, but unexplored.

Infants' learning processes play a critical role in the development of attention-sharing skills. We have yet to establish what the learning processes are involved in this development, and how they contribute to it. Triesch, Teuscher, Deák and Carlson (in press) argue that *contingency learning* is critical to the acquisition of attention-sharing skills. Contingency learning is defined as using experienced sequences of events to generate representations (i.e., predictions) of likely ongoing and future event sequences. Contingency learning occurs in 2-month-olds (Haith & Benson, 1997; Kaye, 1982). For example, after viewing a sequence of alternating lights, infants will shift their gaze in anticipation of the location of the next light. Thus, contingency learning in infants involves (a) predictions about the locations of upcoming events; and (b) behavioural reactions to these predictions, in the form of visual attention-shifts. Stated in this way, the relevance of contingency learning to attention-sharing is evident. Learning to follow gaze can be described as a multidimensional matrix of contingencies between changes in the other person's head pose (or eye direction), locations of expected interesting sights, and learned motor responses (i.e., head turns and saccades) to the generated prediction (e.g. If parent turns 90° to the left, then scan to the left and expect interesting sight approximately 90° from the midline).

Another critical learning process, although seldom considered important for attention-sharing (or for social development in general) is habituation (Sirois & Mareschal, 2002). The role of habituation in infants' attention-sharing abilities will receive only a brief mention in this chapter since it has been addressed in detail by Triesch et al. (in press). Infants' looking and gaze-shifting sequences depend in part on their diminishing interest over time in a given sight. This might help to resolve conflicts between choosing different interesting or appealing sights. For example, human infants are interested in faces, especially their caregivers' faces. They are also interested in high-contrast, moderately complex objects (e.g. toys). Visual habituation must work in tandem with these perceptual preferences by modulating interest (or reward values of stimuli) over time. Habituation thereby prevents repetitive or persistent interest in a given sight, which would suppress the attention-shifting required for attention-sharing. The role of habituation and perceptual preferences in attention-sharing goes even further. Human infants and adults have similar perceptual interests and similar habituation processes (although adults process information faster). This rough match in interest, and in the process of losing interest, could facilitate the interpersonal coordination of attention-shifts—especially if caregivers slow their attention shifts and restrict their interests to more closely match infants. By one account (Tomasello, 1999), willingness to make modifications such as these is at the heart of humans' unique capacity for attention-sharing and for cultural evolution.

Theories of Infant Attention-Sharing

Against the backdrop of ecological and developmental factors surveyed above, how do researchers explain the development of attention-sharing skills? Most theories historically have not explicitly taken into account ecological structure or learning, perceptual, and affective factors. In the following section we evaluate two influential theories that proposed specialized mechanism(s) by which infants learn attention-sharing skills. We assert that these theories are viable only if they explain phenomena that cannot be explained by known ecological and developmental factors (e.g., contingency learning). We will then backtrack, in a sense, by outlining an alternate theory that explains attention-sharing skills as the processing of a structured social environment through early-emerging learning, perceptual, and affective capacities (Triesch et al., in press), with no additional specialized mechanisms.

Previous Theories of the Development of Shared Attention

Two influential theories of attention-sharing skills in human infants are Baron-Cohen's (1995) and Butterworth's (1995). Each attempts to explain some of the facts described above. Each succeeds to some extent by proposing specialized attention-sharing mechanisms. However, each has significant shortcomings. Since both theories continue to be widely cited, it is worthwhile to evaluate each one in detail. We shall do so in this section, and in the following section we shall present an alternative theory of how attention-sharing skills develop. We argue that the alternative theory accounts for more data, is more explicit (and thereby falsifiable) and is more parsimonious than the two theories reviewed here.

Baron-Cohen's theory proposes that humans have special-purpose modules for detecting and processing social information. Some modules begin working before others, which explains developmental changes. First, humans and many other species have an evolutionarily primitive eye direction detector (EDD). In addition, human infants have an intentionality detector (ID). These two mechanisms feed input to an evolutionarily new shared attention mechanism (SAM), which is capable of inferring others' attention. The output from this mechanism serves as input to two theory of mind modules that draw causal inferences about unseen mental states. The evidence for this theory is that although many species are sensitive to the eye direction of other organisms, few use gaze to infer others' attention. In addition, Baron-Cohen explains autism spectrum disorder or ASD, a developmental syndrome characterized by social and language deficits, as a selective 'knock-out' of some modules that control attention-sharing and attention–inferring processes.

Baron-Cohen's theory has limitations. First, there exists no neurological evidence of these different modules; on the contrary, comparative brain studies have not revealed unique hominid brain features that underlie theories of mind, for example. Also, comparative studies of autistic brains have not identified specific deviant features that explain, for example, deficits in theory-of-mind or joint-attention skills. To the contrary, autism (ASD) is polymorphous, with a constellation of behavioural, cognitive, social and communicative deficits and numerous brain differences (Gillberg, 1999), including differences in cellular and anatomical structure in frontal and temporal cortex, basal ganglia, hemispheric connections and cerebellum. Thus, neuropsychological data on ASD fail to support Baron-Cohen's typology.

Baron-Cohen's theory also is disconfirmed by attention-sharing behaviours of children with autism. Leekham and colleagues (e.g. Leekham, Hunnisett, & Moore, 1998; Leekham, López, & Moore, 2000) found that children with ASD are capable of detecting eye direction and following others' gaze, but they spontaneously apply this ability only in limited circumstances. Similarly, children with ASD can be explicitly trained to share attention (Whalen & Schreibman, 2003). Such findings cannot be explained by Baron-Cohen's theory, which makes no predictions about learning. However, delays in attention-sharing in ASD might be explained by general perceptual or cognitive problems, such as motion perception deficits (Bertone, Mottron, Jelenic, & Faubert, 2003).

Empirical findings regarding the higher-order (i.e. post-EDD) modules also do not support Baron-Cohen's theory. Autism studies provide evidence that the failure of theory of mind is neither universal nor unique to autism, and that social information processing in ASD is related to attention and executive cognitive difficulties (Gillberg, 1999). Also, evidence from other psychiatric disorders does not support Baron-Cohen's modular scheme. For example, orbitofrontal cortical damage tends to impair social cognition and theory-of-mind test performance, but not specifically, completely, or universally (Grattan & Eslinger, 1991). More problematic for Baron-Cohen's theory is psychopathy, characterized by absence of some theory-of-mind functions (e.g., empathy) but great competence in another: deception (Hare, 1993). This pattern of mixed competence is only in some cases associated with frontal damage (Damasio, Tranel, & Damasio, 1990; Hart, Forth, & Hare, 1990).

Finally, comparative studies show that attention-sharing—not just eye-direction detection—occurs in nonhuman primates (Hare, Call, Agnetta, & Tomasello, 2000; Johnson, 2001). At least some ape species use information on other animals' mental states to choose social responses (Tomasello, Call, & Hare, 2003). Baron-Cohen's theory does not specify whether chimpanzees and perhaps all apes possess all, some or none the same modules as humans.

To summarize, there is little direct evidence for, and much evidence against, Baron-Cohen's theory. No explanation is given for a host of developmental findings, and no account is given of whether and how the putative modules interact with known learning, perceptual and affective processes.

A different account involving multiple mechanisms was proposed by the late George Butterworth (1995). Butterworth observed that shifts in adults' gaze cause 6- to 9-month-olds to search along the same direction until they see something interesting. His theory attempts to explain this and other interesting phenomena through a developmental sequence of mechanisms for attention-sharing. The first is a primitive, ecological mechanism that is no more sophisticated than many kinds of responses to gaze information made by nonhuman vertebrates (Chance, 1967). However, by 12 months, a geometric mechanism emerges which lets infants use head pose (or arm direction) to compute locations or directions. At this age, infants can ignore a frontal target to follow gaze to another, less-central target within their visual field; however, the theory predicts that infants still cannot follow gaze to locations outside their visual field. Finally, at approximately 18 months, infants are able to follow gaze to targets behind them. Butterworth attributed this new ability to a new representational mechanism that interprets adults' attention as directed to targets that might be anywhere within a viewer-specific Euclidean spatial frame. In other words, the infant knows that other people can attend to things he or she cannot see. This last mechanism emerges around the same age as other behaviors that show mental-state inferences (i.e., 18- to 24-months).

Butterworth's theory accounts for the intriguing finding that infants do not follow gaze to out-of-sight locations until a relatively late age. Younger infants sometimes begin scanning in the right direction; however, they get 'stuck' on the first target they see (Butterworth & Cochran, 1980; Butterworth & Jarrett, 1991). In this regard, Butterworth's theory is an excellent attempt to integrate several intriguing behavioral findings.

However, the theory has several problems. First, it does not explain how the three mechanisms develop; second, there is no account of how the three mechanisms interact; third, some findings are inconsistent with the theory.

The first problem is that the manner in which the mechanisms develop is not clearly specified. To be specific, we can ask (a) what brain, perception and motor patterns were available before the mechanisms emerged; (b) what relevant experiences were available and (c) how were these experiences processed by the infant? These questions are important because in the context of other developmental changes, difficult questions arise about Butterworth's mechanisms. For example, why does the ecological mechanism emerge around 6 months, though many relevant perceptual and motor abilities (e.g. attention-shifting, motion-cueing, face saliency and visual search) are in place by 3 months or earlier (Johnson, 1997)? By 2–3 months infants have had much experience with adults' attention-shifts. Thus, many critical components are in place for several months before the ecological mechanism emerges. Presumably this lag reflects a lengthy learning process; however, the nature of this process is not specified. Similarly, once the ecological mechanism emerges, it is several more months before the next mechanism emerges. Does each mechanism emerge as the product of a different learning process, or of the same learning process operating on different information, or perhaps as the timed unfolding of some gene-regulated processes? This is unspecified. Another problem is the theory does not explain why the ecological mechanism works on gaze shifts before pointing gestures, though pointing is a more predictive and salient cue to adults' attention. To test the theory it would be important to know what predictions it makes concerning these questions. Similarly, with regard to the geometric mechanism, infants receive plenty of input between 3 and 12 months pertaining to directional vectors and spatial relations. Why does this bear fruit only around 12 months? Similarly, how does the representational mechanism emerge from infants' experience? The underlying learning and/or maturational processes that give rise to the past mechanisms should be specified.

To be fair, concerns such as these apply to many theories of cognitive development, and do not disconfirm the theory. However, other problems remain. The second concern is that Butterworth's theory does not discuss how the mechanisms interact. This is an important consideration because the ecological mechanism will certainly remain active throughout the lifespan, even as spatial processing continues to improve. The output from these mechanisms must somehow be integrated by attention-shifting and target-selecting networks.

Once the representational mechanism comes 'on-line,' how is the output from all three mechanisms integrated to produce better-regulated responses to other people's social and perceptual behaviors? Butterworth's theory does not specify how the attention-shifting system integrates output from the mechanisms, or how integration develops during the first few years. Again, this does not disconfirm the theory, but it highlight the need for elaboration.

The third concern is that some findings are not consistent with Butterworth's theory. Some findings disconfirm Butterworth's age estimations. For example, infants follow adults' gaze to targets behind them by 12 months, not 18 months, and they can ignore a frontal target and follow an adult's gaze/point to a peripheral target by 9 months, not 12 months (Deák et al., 2000; Flom et al., 2003). In other tasks infants also use a non-egocentric representation of space before 18 months (Presson & Ihrig, 1982). These findings might just mean the mechanisms are acquired earlier than Butterworth proposed, which is not necessarily a major problem. However, other data suggest the need for additional modifications. Specifically, some behavioural benchmarks (e.g. following gaze to back targets) are sensitive to contextual factors. For instance, the degree of adults' head-turns influence infants' gaze-following to targets behind them (Deák et al., 2000; Flom et al., 2003); thus, some factor like the 'amount of motion' interacts with the geometric and/or representational mechanism (or both). This requires an elaboration of the theory. Further, when complex, distinctive targets are used instead of simple, repetitive targets, 9- and 12-month-olds follow adults gaze and pointing more. This also requires a modification of at least two of the mechanisms, because the complexity effect is found for targets in front of *and* behind the infant. Another unexplained finding is that 15-month-olds follow gaze less often if the adult's line-of-sight is obstructed (Brooks & Meltzoff, 2002; Butler et al., 2000). This requires another modification of the geometric mechanism. As such modifications accumulate, they undermine the theory's parsimony. It would be preferable to have a theory that explains many of these findings through a simpler framework. We now turn our attention to a framework that does not require any mechanisms beyond a set of factors that are known to exist in young infants and in their social environments, which can explain a wider range of results.

Shared Attention Emerges from Early Perceptual, Learning and Affective Processes

The Modeling the Emergence of Shared Attention (MESA) project at the University of California, San Diego (Carlson & Triesch, 2003; Fasel, Deák, Triesch, & Movellan, 2002), has been developing a theory of how joint-attention skills develop in infancy. This theory focuses on the hypothesis that complex social skills might emerge from the interaction of basic perceptual, cognitive and affective processes that begin operating early in infancy. The theory proposes the following elements as sufficient (and in most cases necessary) for attention-sharing behaviors to emerge: a *Basic Set* of affective-motivational tendencies, perceptual capacities, and learning processes including *Temporal Difference* (TD) reinforcement learning and habituation, and a *Structured Social Environment* (SSE) that provides input for learning attention-sharing. Each of these components is described below. A basic assumption is that infants' learning processes are tuned to structured patterns of information (i.e., caregivers' behaviors) in the social environment. By virtue of this tuning, infants can learn to predict and respond to regularities in others' behaviours.

The theory considers infant development from 3 to 12 months. Three months is a relevant starting point because many basic visual and attentive capacities have emerged by this age. Also infants around 2-3 months become more socially oriented and spend more time awake and alert, so they can receive more social input. However, 3-month-olds do not yet engage in episodes of attention-sharing, or not use adults' behaviours as cues to the locations of interesting sights. Nor do they intentionally attract or re-direct the attention of adults. As reviewed above, these more sophisticated behaviors are operational by 12 months; hence this is the upper range of our theory. Of course attention-sharing skills are not fully developed by 12 months, and we hope that future work will extend the theory to infants' later accomplishments.

Basic Perceptual and Affective Processes

We postulate that certain perceptual capacities of infants by 3 months of age are vital for acquiring attention-sharing skills. Some of these involve attention-shifting. By 3 months certain attention-regulating brain pathways, specifically projections from the visual cortex to the frontal eye fields (FEF), are maturing. These projections are critical for visual planning, anticipation and learning (Johnson, 1990). Consequently, several attention-shifting behaviours change at around 3 months: (a) stimuli appearing outside central vision elicit attention shifts (Butcher, Kalverboer, & Geuze, 2000); (b) directional cues (arrows or motion) facilitate directional attention shifts (Farroni et al., 2000) and (c) infants can inhibit attention shifts based on a directional spatial cue, if that cue predicts a stimulus elsewhere in the visual field (Johnson, Posner, & Rothbart, 1994). The third ability, which depends on FEF functions, is particularly relevant to gaze- or point-following, which sometimes require looking away from a social cue (e.g. caregiver's face) to find a distal target.

Attention-sharing also depends on affective and motivational traits. From a very young age, infants exhibit preferences for human social stimuli, including faces, voices, odours and tactile stimuli, particularly those of caregivers. These preferences suggest a pervasive motive to engage in and prolong social interaction, which facilities language and communication development (Locke, 1993). This motivation is a developmental product. Around 2–3 months the social responsiveness of infants becomes more consistent and focused. Infants produce their first social smiles, and parents describe them as being more engaged and responsive during interactions (Cole & Cole, 1996). This shift might be an early consequence of reinforcement learning (see the following section) based on dopaminergic activity in the basal ganglia and cortex (Schultz, 2000). This speculation presumes that basic reinforcement learning mechanisms in adult human and other species are also functioning in human infants within several weeks or months of birth. This is a reasonable assumption because even newborns exhibit operant learning. In more general terms, we assert that any satisfactory theory of the emergence of attention-sharing skills must consider infants' affective predilections, including preferences for certain stimuli, as well as their motives to seek out certain hedonic social situations.

Learning Processes: Reinforcement and Habituation

We hypothesize that reinforcement learning processes, a subset of which are observed in traditional operant conditioning, are critical for later acquisition of attention-sharing skills. This connection was first explored by Moore (1996). Reinforcement learning, and TD (i.e., temporal difference) learning in particular, is a family of neurally plausible algorithms that model reward- and punishment-based learning in the brain. We propose this as the fundamental process by which attention-sharing skills are acquired. Models of reinforcement learning involve value-based reward (or punishment) signals, but are not restricted to Skinnerean, anti-mentalistic frameworks or assumptions. They do share a goal of Skinnerean models: to understand the relation between experienced outcomes (positive or negative), affect, and adaptive behaviour. TD learning (Sutton & Barto, 1998) formalizes how agents (e.g. infants, undergraduates, rats, computer or robotic agents) learn to maximize reward over time through a trade-off between exploitation (i.e. choosing actions most likely to garner the highest future reward in a given situation) and exploration (i.e. choosing less-rewarded actions). A balance of exploration and exploitation can eventually generate behavioural policies that yield some short-term rewards but ultimately higher average long-term rewards. For example, constantly consuming chocolate because it has a large immediate reward is an exploitation-based policy, whereas a more balanced approach with some exploration of different foods (some with less immediate reward) will ultimately constitute a more healthful diet (example provided by Ian Fasel). TD learning agents register both short- and long-term consequences of specific action choices, to gradually shape more adaptive action policies. Exploration enables agents to adapt their policies to changing environments.

TD learning algorithms are plausible formalizations for the process by which infants learn shared attention skills. As mentioned above, infants are rewarded by social stimuli (faces, voices) as well as non-social stimuli (e.g. colourful objects). Thus, such stimuli yield short-term rewards. In addition, infants learn to predict regular event sequences and to respond (e.g., shift attention) in anticipation of future events (Haith, Hazan, & Goodman, 1988; Watson & Ramey, 1985). This implies that infants acquire action policies for predictable event sequences. Further, TD learning algorithms have been related to specific neuromodulatory systems (Doya, 2000; Schultz, Dayan, & Montague, 1997). Thus, although TD learning models have heretofore played almost no role in theories of infant and child development (but see Schlesinger & Parisi, 2001), we believe they hold great promise for explaining and predicting how young humans develop action policies in 'hot' contexts, that is, situations in which stimuli or outcomes are affective-laden. Interactions with caregivers are good examples of such situations.

Habituation also plays a critical role in our theory, not just as a methodological tool but as a critical learning process (Sirois & Mareschal, 2002) that works in concert with attention-shifting and reinforcement learning processes. How does habituation facilitate the development of attention-sharing skills? When an infant views a rewarding (i.e. interesting) stimulus, such as a caregiver's face or a toy, habituation begins. This can be modelled as a systematic decline in the reward value of a stimulus over time. This decline, in turn, affects the output of the TD learning algorithm. The decline of an anticipated reward will affect the agent's behaviour by increasing the probability of choosing another action (e.g. shifting attention) that has in the past yielded rewarding (i.e. interesting) outcomes (e.g., sights). This process can produce cycles of attention between the caregiver's face and toys or interesting objects that the caregiver is holding or manipulating.

Structured Social Environment

These processes will not function without patterned input from an SSE. We hypothesize that the most critical input for infants to learn attention-sharing skills is a category of everyday structured interactions in which caregivers and infants are in close proximity (within 1 m), positioned so that each can see the other. Each participant's attention may be on the other or on some prop of the activity, be it a toy, a tool held by the parent (e.g. hairbrush, spoon, washcloth) or the caregiver's hand. Activities in this category include face-to-face play, feeding, diapering and bathing. Since a large proportion of infants' waking time is spent in such activities, they constitute an important source of social input (Bruner, 1983; Watson, 1972).

We hypothesize that these activities are important because they provide structured information to the infant. This is not a new idea: a sizeable literature indicates that infants and caregivers reciprocally adjust to the statistical structure of their interactions, for example by synchronizing action (Kaye, 1982). Caregivers' actions are predictable enough that by 9–10 months of age infants can predict the locations of interesting sights based on their parents' head poses. This is the basis of gaze-following. Through the same TD/habituation-based learning process, infants could also learn to find interesting objects from their caregivers' pointing gestures or other manual actions (e.g., reaching with a open hand shape). Manual actions like touching, moving and reaching for objects might also provide predictive information to infants. Finally, the same learning mechanisms might explain how infants come to use caregivers' emotional expressions to regulate exploration (i.e. social referencing). Although this idea has not been tested, the point is that this framework might eventually explain a number of phenomena.

How Shared Attention Emerges

How exactly does shared attention emerge from the combination of a basic set of perceptual and affective traits, reinforcement learning and habituation, and a structured social environment? How can we test the claims that these elements are necessary and sufficient for skills like gaze- and point-following? We propose that given the basic perceptual and affective traits described above, TD learning (with fairly high levels of exploration) and habituation will produce cycles of attention to different interesting stimuli (caregiver's face, toys, and tools). Eventually (through TD learning), infants will anticipate parents' predictable gaze shifts and manual actions (within structured activities) and exploit these as a source of information about the locations of interesting targets. In shared attention, the reward value of different stimuli fulfils an important function. For example, infants prefer a parent's face in direct gaze to a face that is looking slightly away from them (Hains & Muir, 1996) or turned to the side (Sai & Bushnell, 1988). This implies that although infants are rewarded by seeing their parents' faces, the reward diminishes when the parent looks away from them. This creates a trade-off between the potential reward value of an (anticipated) peripheral object and the reward value of the parent's face. Moreover, habituation will decrease the face's reward value as a function of time. Together, these dynamics gradually increase the probability of a gaze shift away from the parent's face. Further, when the parent shifts gaze, the directional motion of his or her head might trigger a same-direction attention shift by the infant (Farroni, Mansfield, Lai, & Johnson, 2003). Collectively, these factors may result in the following sequence:

1. Caregiver and infant are looking at one another
2. Caregiver looks away towards an object
3. Infant begins a scan in the same direction

From this sequence, the infant obtains time-locked information about contingencies between the caregiver's head pose and the location of interesting sights. Over time, infants should learn how caregivers' head poses (or direction of pointing) relate to different directions or regions of space. Further, since habituation begins with each new fixation, infants will tend to shift back to their caregivers' faces, producing gaze alteration sequences of the kind sometimes observed in attention-sharing (Tomasello, 1999). Extending this idea, we can explain how infants could learn to follow gaze as a result of their interest in adults' manual actions. People tend to look at their own hands while manipulating objects (Land, Mennie, & Rusted, 1999), so caregivers' manual actions will often specify their gaze direction. This provides an excellent source of information for infants to associate specific locations of interesting events with people's head poses.

Other Predictions

This framework does not add any specialized evolved mechanism for attention-sharing. The critical elements are mostly available to agents other than humans; therefore, some attention-sharing skills might emerge in nonhuman primates or other vertebrates, and even in artificial agents (e.g. robots). Thus, it is not surprising that in some contexts, adult chimpanzees and other nonhuman primate species can use gaze direction of conspecifics or trainers to shift attention (Itakura, Agnetta, Hare, & Tomasello, 1999; Tomasello, Call, & Hare, 1998). There is also evidence that chimpanzees can learn to point (Leavens, Hopkins, & Bard, 1996) and use trainers for social referencing (Russell, Bard, & Adamson, 1997). This suggests that in some SSEs, chimpanzees' perceptual, affective and learning processes are sufficient to learn a range of shared-attention behaviours. However, their skills emerge later than young children's, and are limited (e.g. Itakura, 1996; Povinelli, Bierschwale, & Cech, 1999; Povinelli & Eddy, 1996). (Information on attention-sharing in other apes is sparse, but comparable skills seem to be present in gorillas, for example; Gómez, 2004.) Perhaps many primates are capable of learning spatial cued associates for social stimuli (e.g. faces) because this rests on fairly general visual attention and reinforcement learning processes, plus a pervasive interest in social events. However, more advanced joint attention functions (described above) might require greater interest in faces, more efficient learning process, and/or more supportive and informative social environments.

Other predictions can be made about infants with atypical perceptual, affective or learning processes, or social environments. As noted above, attention-sharing deficits are common in ASD (Baron-Cohen, Allen, & Gillberg, 1992; Sigman, Mundy, Ungerer, & Sherman, 1986; Tager-Flusberg, 1996). In ASD several elements in our theory may be disrupted. Children with autism prefer events that are very predictable, and they might have trouble learning to predict human social events, which are only moderately predictable (Gergely & Watson, 1999). This is consistent with evidence that children with ASD find it difficult to predict or infer others' behaviours or emotions (e.g. Baron-Cohen, 1991). People with ASD also have perceptual and attention-shifting deficits (e.g. Bertone et al., 2003; Gepner & Mestre, 2002; Wainwright-Sharp & Bryson, 1993); thus an important component of their perceptual abilities appear to be compromised. Finally, children with ASD find face-to-face interactions (and perhaps faces in general) less rewarding than normal children (Hutt & Ounsted, 1966). This lower reward value would affect reinforcement learning. Thus, our theory suggests at least three alternative possible causes of gaze-following deficits in ASD.

Testing the Theory: Computational Model and Future Questions

We have recently developed computational simulations to test our theory (Carlson & Triesch, 2003; Triesch et al., in press). The primary goal is to test whether the basic, TD learning and habituation are sufficient for learning gaze-following (and probably point-following) when given structured social input. The success of the model could provide an existence proof that these elements are adequate to account for gaze-following, rendering other mechanisms unnecessary (e.g. modules proposed by Baron-Cohen, 1995).

In the initial model the infant has been modelled as an artificial neural network that uses reward-driven TD-reinforcement algorithms (Sutton & Barto, 1998). The learning environment is structured as a limited visual field with a caregiver in the centre and 10 different locations in space represented as possible output vectors. At quasi-random intervals the pattern of the caregiver's face changes, to simulate changes in face pose. These changes correlate with the onset of moderately reinforcing input vectors (i.e. objects) in different spatial locations. The infant model at every time step must decide (a) whether to shift gaze; and (b) where to look, specifically, either at the caregiver's face or one of the ten locations. This decision constitutes the output, and will receive an immediate reward value r ($-1 \leq r \leq 1$). In our first simulation equal positive reward values were assigned for looking at the caregiver's frontal or profile face poses, or for looking at the location with an object. Looking at empty locations receives no reward. Recent simulations have used multiple object locations and more differentiated reward values (e.g. for frontal poses vs. profiles), partly based on naturalistic data on real infants' looking times (You et al., 2005).

Our analysis of learning in the preliminary model (Carlson & Triesch, 2003) under different system parameters revealed that gaze-following emerges robustly across a wide range of parameter values. Most generally, the model first learns to look at the caregiver's face, and then gradually learns to look away from the caregiver's face to the location specified by the caregiver's head pose. This gradual acquisition of looking away roughly maps onto developmental changes. This could not be be modeled by just any automated learning agent; there must be a balance of exploitation and exploration. In addition, the performance of the model varies in a graded manner with changes in two parameters: learning rate and habituation. Briefly put, the model's learning degrades with either very slow learning or very low rates of habituation.

In follow-up tests, the reward structure of the model was manipulated to test one theory of how attention-following deficits might develop in ASD. By reducing the reward value of the caregiver's face, gaze-following was learned more slowly, though the reward value of finding objects was unchanged (Triesch et al., in press). Thus, the model has the potential to simulate disorders in social attention. In addition, Teuscher and Triesch (2004) found that systematic variations in structured social input (e.g. a more- vs. less- responsive caregiver) caused predictable changes in the infant-agent's learning.

Note that every element of our theory is incorporated in the model. Since even this first version of the model simulated a number of behavioural findings, we consider it a useful tool. Currently, it is one of two computational learning models of gaze-following under active development (see also Nagai, Hosoda, Morita, & Asada, 2003). Differences between the models are reviewed by Triesch et al. (in press).

This model uses a number of simplifying abstractions to keep computational complexity at a manageable level, and to focus on essential aspects of the theoretical problem. Nevertheless, it is merely a first step towards a complete and formal account of the emergence of attention-sharing behaviors, and its explicit elements are grounded in solid developmental and neuroscience findings. However, to eventually account for all relevant behavioural results (e.g. Butterworth & Jarrett, 1991; Deák et al., 2000), the model must be extended, and this is an active goal of our current research. One major goal is to fine-tune the input to the model, based on real behavioral data from quasi-naturalistic play between infants aged 3 to 12 months and their parents (You et al., 2005). In this study parents and infants engage in several kinds of interactions, including parents showing infants objects in different locations. Where (and at what) infants and parents look, and what parents touch, manipulate or point to, are variables that are coded with precise timing. This study provides details of social input to infants, and these details will eventually drive the simulated caregiver's behavior in the computational model. Eventually, we will derive a model that uses averaged probabilities of sequences of caregivers' actions. We also hope to differentiate the action sequences of parents of younger infants versus parents of older infants. In this manner, the input to the simulated infant can be systematically changed from early to later social experiences, for a more accurate training regimen. This will let us precisely and realistically test the process of learning and the behavioral input of the virtual infant.

How Shared Attention Helps Infants Enter a World of Symbolic Communication

The attention-sharing behaviours acquired in the first year have their greatest impact when coordinated with language. Integrating shared attention and language in a single theory remains a serious challenge. Further, both are associated with the ability to represent mental states and their relations to overt behaviours (Tomasello, 1999). No current theory has been able to explain the relations between social attention, mental-state representations, and language as all three emerge during infancy and childhood.

One means to begin addressing this is to examine how infants respond to different linguistic and non-linguistic actions by parents that can promote attention-sharing. Walden et al. (under review) investigated how 1-year-olds share attention in semi-naturalistic conditions. On some trials, parents produced an attention-eliciting utterance ('Max, Max!') or an attention-directing utterance ('Look at the bunny!') while shifting gaze to a target. On other trials, the parent produced a pointing gesture while shifting. As compared with simple gaze shifts by the parents, any of these behaviours (i.e. either utterance or pointing) led to an increase in 21-month-olds' attention-following. In contrast, the attention-following of 15-month-olds increased with an attention-directing utterance or pointing, but not with an eliciting utterance. This suggests that 21-month-olds are better at using indirect cues (i.e. being called by name) as a signal to shift attention to, and then away from, the parent. Such an indirect action policy could be acquired by a TD learning system. Did 15-month-olds show *any* response to being called by name? These younger infants looked at parents more often on eliciting-utterance trials, but only the older infants used this cue as both a direct *and* an indirect message (i.e. to shift attention away from parents after checking their gaze direction).

Walden et al. (under review) also found that infants did not respond mechanically to parents' utterances. On some trials the parent made attention-directing utterances while looking towards the named target, but held a hand in front of their eyes (as if playing peek-a-boo). In this peculiar conjunction of an attention-directing utterance and blocked line-of-sight, infants frequently looked at the parent but infrequently shifted gaze to the target. Thus, 1-year-olds had sufficient knowledge about line-of-sight constraints to modulate their response to the utterance according to what the parent could or could not see (Baldwin, 1993). This indicates an ability to integrate verbal and non-verbal social cues in order to decide whether to shift attention to or away from parents. As infants approach their second birthday, they learn to make non-literal interpretations of specific verbal cues (e.g. being called by name) based on the social context, including the caregiver's looking behaviors (see also Baldwin, 1993). This suggests that caregivers' verbalizations initially serve as orienting or attention-prolonging signals with limited specific meaning, and gradually acquire specific literal and non-literal meaning to infants (see also Fernald, 1993).

The theory outlined above might eventually explain these findings. If parents' verbalizations predict interesting events (similar to, for example pointing gestures), infants might learn to use verbalizations in order to increase long-term social rewards. Moreover, since TD learning tracks reward states over multiple changes in the environment, the infant might learn to combine linguistic and non-linguistic cues to maximize reward. However, we are by no means advocating a Skinnerean account of language acquisition. Rather, these speculations concern only how infants' might learn the pragmatic force of certain types of speech acts. Beyond this, the theory cannot explain how infants acquire specific linguistic meaning of lexical forms and semantic relations, morphological paradigms, syntactic structures, etc. Moreover, it is not clear how the theory could explain how children and adults generate elaborate mental representations of mental states and processes (in self and others). Thus, our theoretical framework, including general perceptual, learning and affective traits and a structured social environment, is not meant to explain the development of the full range of attention sharing, linguistic, and theory of mind abilities. Nevertheless, it might explain many early phenomena in the development of a set of specialized skills by which infants communicate with caregivers and predict their behaviors.

Summary

Patterns of shared attention in humans are diverse (in form and function), early-emerging and critical for normal social, cognitive and language development. Currently there is considerable experimental and observational evidence about the development of attention-sharing skills in neurologically intact infants. There is also growing evidence about how these skills develop in infants and children with developmental disabilities. Finally, there has been much recent research on social attention in nonhuman primates and other vertebrates. These areas of research reveal great complexity and variability in primate social attention. Fortunately, advances in experimental and theoretical neuroscience, and computational modeling techniques, offer new possibilities for building and testing theories to explain the behavioral evidence. For example, simulations using embodied systems like virtual and robotic agents will allow testing of more complex and realistic models. Embodied simulations also enable ethical manipulation of different information-processing parameters to simulate developmental disabilities or cross-species differences. Although such efforts are in their initial stages (Triesch et al., in press; Nagai et al, 2003), early results reveal the importance of formally describing basic perceptual, learning and affective processes hypothesized to be critical, as well as structured information in the social environment. The major questions in coming years will be whether new theories and ways of testing them, including comparative studies, experimental and observational studies of human infants, and computational simulations will explain the relation between attention-sharing and symbol-using skills (e.g., language) in typical infants, infants with developmental disabilities, and nonhuman primates.

References

Atkinson, J. (2000). *The developing visual brain*. Oxford University Press.

Arterberry, M., Craton, L., & Yonas, A. (1993). Infants' sensitivity to motion-carried information for depth and object properties. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 225–234). Hillsdale, NJ: Erlbaum.

Ashmead, D., & Perlmutter, M. (1980). Infant memory in everyday life. In M. Perlmutter (Ed.), *New directions for child development: Children's memory* (Vol. 10, pp. 1–16). San Francisco: Jossey-Bass.

Bakeman, R., Adamson, L. B., Konner, M., & Barr, M. G. (1985). !Kung infancy: The social context of object exploration. *Child Development, 61,* 794–809.

Baldwin, D. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20,* 395–419.

Baron-Cohen, S. (1991). Do people with autism understand what causes emotion? *Child Development, 62,* 385–395.

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared

attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 41–59). Hillsdale, NJ: Erlbaum.
Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *British Journal of Psychiatry, 161,* 839–843.
Bartlett, M. S., Movellan, J. R, & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks, 13,* 1450-1464.
Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2003). Motion processing in autism: A complex issue. *Journal of Cognitive Neuroscience, 15,* 218–225.
Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a theory of mind? In M. Lamb & L. Sherrod (Eds.), *Social cognition in infancy* (pp. 333–373). Hillsdale, NJ: Erlbaum.
Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology, 38,* 958–966.
Bruner, J. (1983). *Child's talk.* New York: Norton.
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.
Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant understanding of the referential nature of looking. *Journal of Cognition and Development, 1,* 359–377.
Butterworth, G. (1995). Origins of mind in perception and action. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Hillsdale, NJ: Erlbaum.
Butterworth, G. E. & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioral Development, 3,* 253–272.
Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology, 18,* 25–50.
Butterworth, G. E., & Jarrett, N. L. M. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology, 9,* 55–72.
Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1,* 253–274.

Carlson, E. & Triesch, J. (2003). A computational model of the emergence of gaze following. In H. Bowman & C. Labiouse (Eds.), *Connectionist Models of Cognition and Perception II*. London: World Scientific.

Chance, M. R. (1967). Attention structure as the basis of primate rank orders. *Man, 2,* 503–518.

Cole, M., & Cole, S. (1996). *The development of children* (3rd ed.). New York: Freeman.

Courchesne, E. (1997). Brainstem, cerebellar and limbic neuroanatomical abnormalities in autism. *Current Opinion in Neurobiology, 7,* 269–278.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioral Brain Research, 41,* 81-94.

Deák, G. O., Flom, R., & Pick, A. D. (2000). Perceptual and motivational factor affecting joint visual attention in 12- and 18-month-olds. *Developmental Psychology, 36,* 511–523.

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology, 10,* 732–739.

Dunham, P., Dunham, F., & Curwin, A. (1993). Joint-attentional states and lexical acquisition at 18 months. *Developmental Psychology, 29,* 827–831.

Elman, J., Bates, E., Johnson M. H., Karmiloff-Smith A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge: MIT Press.

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.

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests and an evolutionary hypothesis. *Journal of Experimental Child Psychology, 85,* 199–212.

Fasel, I., Deák, G. O., Triesch, J., & Movellan, J. (2002). Combining embodied models and empirical research for understanding the development of shared attention. *Proceedings of the International Conference on Development and Learning, 2,* 21–27.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development*. Monographs of the Society for Research in Child Development, 59*(5, Serial No. 242).

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64,* 657–674.

Flavell, J. H., Green, F. L., Herrera, C., & Flavell, E. R. (1991). Young children's knowledge about visual perception: Lines of sight must be straight. *British Journal of Developmental Psychology, 9,* 73–87.

Flom, R., Deák, G. O., Phill, C., & Pick, A. D. (2003). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development, 27,* 181–194.

Gepner, B., & Mestre, D. (2002). Rapid visual-motion integration deficit in autism. *Trends in Cognitive Science, 11,* 455.

Gergely, G, & Watson, J. S. (1999). Early socio-emotional development: Contingency perception and the social-biofeedback model. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 101–136). Mahwah, NJ: Erlbaum.

Gillberg, C. (1999). Neurodevelopmental processes and psychological functioning in autism. *Development and Psychopathology, 11,* 567–587.

Gómez, J. C. (2004). *Apes, monkeys, children, and the growth of mind.* Cambridge, MA: Harvard University Press.

Grattan, L. M., & Eslinger, P. J. (1991). Frontal lobe damage in children: A comparative review. *Developmental Neuropsychology, 7*, 283-326.

Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development, 67,* 1940–1951.

Haith, M. M., & Benson, J. B. (1997). Infant cognition. In W. Damon (Series Ed.) & D. Kuhn & R. Siegler (Eds.), *The handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 199–254). New York: Wiley.

Haith, M. M., Hazan, C., & Goodman, G. S. (1988). Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development, 59,* 467–479.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behavior, 59,* 771–785.

Hare, R. D.(1993). *Without conscience: The disturbing world of the psychopaths among us.* New York: Guilford Press.

Hart, S. D., Forth, A. E., & Hare, R. D. (1990). Performance of male psychopaths on selected neuropsychological tests. *Journal of Abnormal Psychology, 99,* 374-379.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9,* 131–134.

Hutt, C., & Ounsted, C. (1966). The biological significance of gaze aversion with particular reference to the syndrome of infantile autism. *Behavioral Science, 11,* 346–356.

Itakura, S. (1996). An exploratory study of gaze-monitoring in nonhuman primates. *Japanese Psychological Research, 38,* 174–180.

Johnson, M. H. (1990). Cortical maturation and the development of visual attention in early infancy. *Journal of Cognitive Neuroscience, 2,* 81–95.

Johnson, M. H. (1997). The neural basis of cognitive development. In W. Damon (Series Ed.) & D. Kuhn & R. Siegler (Eds.), *The handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 1–49). New York: Wiley.

Johnson, M. K., Posner, M. I., & Rothbart, M. K. (1994). Facilitation of saccades toward a covertly attended location in early infancy. *Psychological Science, 5,* 90–93.

Jusczyk, P. W. (2000). *The discovery of spoken language.* Cambridge: MIT Press.

Kaye, K. (1982). *The mental and social life of babies*. University of Chicago Press.

Land, M. F., Mennie, N., & Rusted, J. (1999). Eye movements and the roles of vision in activities of daily living: Making a cup of tea. *Perception, 28,* 1311–1328.

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (1996). Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 110,* 346–353.

Leekam, S., Hunnisett, E., & Moore, C. (1998). Targets and cues: Gaze-following in children with autism. *Journal of Child Psychology and Psychiatry, 39,* 951–962.

Leekam, S., López, B., & Moore, C. (2000). Attention and joint attention in preschool children with autism. *Developmental Psychology, 36,* 261–273.

Leung, E. H., & Rheingold, H. L. (1981). Development of pointing as a social gesture. *Developmental Psychology, 17,* 215–220.

Locke, J. (1993). *The child's path to spoken language*. Cambridge, MA: Harvard.

MacWhinney, B. (Ed.). (1999). *The emergence of language.* Mahwah, NJ: Erlbaum.

Moore, C. (1996). Theories of mind in infancy. *British Journal of Developmental Psychology, 14,* 19–40.

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.

Moore, C., & D'Entremont, B. (2001). Developmental changes in pointing as

a function of attentional focus. *Journal of Cognition & Development, 2,* 109–129.
Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development, 21,* 373–377.
Morissette, P., Ricard, M., & Gouin Décarie, T. G. (1995). Joint visual attention and pointing in infancy: A longitudinal study of comprehension. *British Journal of Developmental Psychology, 13,* 163–175.
Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20,* 115–123.
Nagai, Y., Hosoda, K., Morita, A., & Asada, M. (2003). Emergence of joint attention based on visual attention and self learning. In *The 2nd international symposium on adaptive motion of animals and machines, SaA-II-3.*
O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development, 67,* 659–677.
Povinelli, D. J., Bierschwale, D. T., & Cech, C. G. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology, 17,* 37–60.
Povinelli, D. J., & Eddy, T.J. (1996). Chimpanzees: Joint visual attention. *Psychological Science, 7,* 129–135.
Presson, C. C., & Ihrig, L. H. (1982). Using mother as a spatial landmark: Evidence against egocentric coding in infancy. *Developmental Psychology, 18,* 699–703.
Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review, 19,* 9–50.
Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33,* 12–21.
Russell, C. L., Bard, K. A., & Adamson, L. B. (1997). Social referencing by young chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 111,* 185–193.
Sai, F., & Bushnell, W. R. (1988). The perception of faces in different poses by 1-month-olds. *British Journal of Developmental Psychology, 6,* 35–41.
Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature, 253,* 265–266.
Schlesinger, M., & Parisi, D. (2001). The agent-based approach: a new direction for computational models of development. *Developmental Review, 21*, 121–146.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience, 1,* 199–207.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Nature, 275,* 1593–1599.

Sigman, H., Mundy, P., Ungerer, J., & Sherman, T. (1986). Social interactions of autistic, mentally retarded, and normal children and their caregivers. *Journal of Child Psychology and Psychiatry, 27,* 647–656.

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Science, 6,* 293–298.

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

Tager-Flusberg, H. (1996). Brief report: Current theory and research on language and communication in autism. *Journal of Autism and Developmental Disorders, 26,* 169–171.

Teuscher, C., & Triesch, J. (2004). To care or not to care: Analyzing the caregiver in a computational gaze following framework. In J. Triesch & T. Jebara (Eds.), *Proceedings of the Third International Conference on Development and Learning*, (pp. 9–16). San Diego, USA: Salk Institute for Biological Studies.

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard.

Tomasello, M., Call, J., & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behavior, 55,* 1063–1069.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states—the question is which ones and to what extent. *Trends In Cognitive Sciences, 7,* 153–156.

Triesch, J., Teuscher, C., Deák, G., & Carlson, E. (in press). Gaze-following: Why (not) learn it? *Developmental Science*.

Turati, C. (2004). Why faces are not special to newborns: An alternative account of the face preference. *Current Directions in Psychological Science, 13,* 5–8.

Wainwright-Sharp, J., & Bryson, S. (1993). Visual orienting deficits in high-functioning people with autism. *Journal of Autism and Developmental Disorders, 13,* 1–13.

Walden, T. A., Deák, G. O., Yale, M., & Lewis, A. [Manuscript under review]. *Eliciting and directing 1-year-old infants' attention: Effects of verbal and non-verbal cues*.

Walden, T. A., & Ogan, T. (1988). The development of social referencing. *Child Development, 59,* 1230–1240.

Watson, J. S. (1972). Smiling, cooing, and "the game." *Merrill-Palmer Quarterly, 18,* 323–339.
Watson, J. S., & Ramey, C. T. (1985). Reactions to response-contingent stimulation in early infancy. In J. Oates et al. (Eds.), *Cognitive development in infancy.* Hove, England: Erlbaum.
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.
You, Y., Deak, G., Jasso, H., & Teuscher, C. (2005, April). *Emergence of shared attention from 3 to 11 months of age in naturalistic infant-parent interactions*. Biennial meeting of the Society for research in Child Development, Atlanta, GA.

Acknowledgment

The authors would like to thank Ian Fasel, Ross Flom, Javier Movellan, Anne Pick and Christof Teuscher for enlightening conversations, and also Karen Au, Anna Krasno and Yuri You for their helpful comments on an earlier version of this chapter. The work described here was supported by the Nicholas Hobbs Foundation, the M.I.N.D. Institute of UC-Davis and the National Alliance for Autism Research (NAAR).