Developmental Science 9:2 (2006), pp 125–157

TARGET ARTICLE WITH COMMENTARIES AND RESPONSE

Gaze following: why (not) learn it?

Jochen Triesch,[1,2] Christof Teuscher,[3] Gedeon O. Deák[1] and Eric Carlson[1]

1. *Department of Cognitive Science, University of California, San Diego, USA*
2. *Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, Germany*
3. *Los Alamos National Laboratory, Los Alamos, USA*

For commentaries on this article see Csibra (2006), Moore (2006) and Richardson and Thomas (2006).

Abstract

We propose a computational model of the emergence of gaze following skills in infant–caregiver interactions. The model is based on the idea that infants learn that monitoring their caregiver's direction of gaze allows them to predict the locations of interesting objects or events in their environment (Moore & Corkum, 1994). Elaborating on this theory, we demonstrate that a specific Basic Set *of structures and mechanisms is sufficient for gaze following to emerge. This Basic Set includes the infant's perceptual skills and preferences, habituation and reward-driven learning, and a structured social environment featuring a caregiver who tends to look at things the infant will find interesting. We review evidence that all elements of the Basic Set are established well before the relevant gaze following skills emerge. We evaluate the model in a series of simulations and show that it can account for typical development. We also demonstrate that plausible alterations of model parameters, motivated by findings on two different developmental disorders – autism and Williams syndrome – produce delays or deficits in the emergence of gaze following. The model makes a number of testable predictions. In addition, it opens a new perspective for theorizing about cross-species differences in gaze following.*

Introduction

The capacity for shared attention is a cornerstone of social intelligence. It plays a crucial role in the communication between infant and caregiver (Brazelton, Koslowski & Main, 1974; Kaye, 1982; Adamson & Bakeman, 1991; Adamson, 1995; Moore & Dunham, 1995). By 9–12 months most infants can follow adults' gaze and pointing gestures, and monitor a caregiver's affect and use it to modulate their own response to an ambiguous stimulus. These behaviors emerge and coalesce on a predictable schedule (e.g. Butterworth & Itakura, 2000; Deàk, Flom & Pick, 2000), although specific milestones show considerable individual differences in age of attainment (Mundy & Gomes, 1998; Markus, Mundy, Morales, Delgado & Yale, 2000). Shared attention skills allow the young of our species to learn what is important in the environment, based on the patterns of attention in older, more expert individuals. In conjunction with a shared language, these skills allow children to communicate what they perceive and think about, and to construct mental representations of what others perceive and think about. Consequently, shared attention is crucial for language and communication (Bruner, 1983; Baldwin, 1993; Tomasello, 1999).

The term shared attention is typically used to denote a set of different skills comprising gaze following, pointing and requesting behaviors. While some authors use the terms *joint* and *shared* attention interchangeably to refer to the matching of one's focus of attention with that of another person, other authors make a subtle distinction between the two. 'Shared' attention is sometimes reserved for the more complex form of communication, wherein two individuals attend to the same object, and each have knowledge of the other's attention to this object (Tomasello, 1995; Emery, 2000). In this paper, we will be concerned with joint attention more broadly, which we view as an important precursor to the emergence of true shared attention. Our particular focus is on gaze following, which may be defined as looking where

Address for correspondence: Jochen Triesch, Department of Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA; e-mail: triesch@cogsci.ucsd.edu

somebody else is looking. Gaze following is a good starting point for investigations into shared attention, because it develops early in life and is a precedent for other shared attention skills.

How does gaze following emerge?

Starting with a pioneering study by Scaife and Bruner (1975), the emergence of gaze following has been investigated in many studies. There has been some debate about when gaze following emerges in human infants, with most estimates ranging from 3 to 12 months (e.g. Butterworth & Cochran, 1980; D'Entremont, Hains & Muir, 1997; Hood, Willen & Driver, 1998; Morales, Mundy & Rojas, 1998). The reasons for this wide range are threefold. First, researchers have used different criteria to define gaze following (Tomasello, 1995). Second, different levels of sophistication of gaze following can be distinguished. Third, different experimental paradigms may differ in sensitivity. The earliest signs or precursors of gaze following can be observed around 3 months of age, and some very rudimentary skills are even present in newborns (Farroni, Massaccesi, Pividori & Johnson, 2004). In particular, D'Entremont *et al.* (1997) showed that 3-month-olds will turn their eyes in the direction of an adult's head turn more frequently than in the opposite direction. Their observation requires rather ideal conditions, such as targets that are well within the infant's visual field. In addition, these demonstrations of 'gaze following' seem to rely on more basic visual tracking mechanisms that facilitate gaze shifts in the direction of motion of a centrally located stimulus. In fact, such motion cueing may initially be necessary, but by around 9 months static head pose alone can be sufficient for gaze following (Moore, Angelopoulos & Bennett, 1997).

Beyond these first signs of gaze following, Butterworth and Jarrett (1991) proposed three different *stages of gaze following* emerging around 6, 12 and 18 months, respectively (but also see Deàk *et al.*, 2000). These stages are defined by infants' new abilities, first to ignore distracting visual objects, and later to follow adults' gaze to locations outside of their visual field.

An important line of research is concerned with the specific features that infants use to establish the adult's direction of gaze. There is evidence that younger infants rely more on the caregiver's head pose than the eyes, whereas between 12 and 14 months there is a significant increase in sensitivity to eye orientation (Caron, Butler & Brooks, 2002). By 18 months, gaze following is reliably produced on the basis of eye movements alone (Butterworth & Jarrett, 1991). This body of work suggests that limitations of the infant's developing face processing skills may play an important role in their ability to follow gaze.

A rather difficult question is what gaze following skills imply about how infants at various ages conceptualize their caregivers' looking behavior. Although early accounts interpreted gaze following skills as indicating considerable social understanding or even a theory of mind, it has been argued that young infants may learn to follow gaze without such an understanding (Moore & Corkum, 1994; Corkum & Moore, 1995). More recently, Woodward (2003) demonstrated that infants need not have an understanding of the relation between a person who looks and the object of his or her gaze. In addition, early gaze following skills may not even require a representational strategy involving the identification of the caregiver as an intentional, perceiving individual (Leekam, Hunnisett & Moore, 1998). Certainly, such representations will emerge over time in older infants, but they might not be necessary to explain the emergence of gaze following behaviors.

Gaze following in other species

Humans are not the only species that exhibit gaze following. Gaze following has been demonstrated in a number of other species, including some (but not all) non-human primates (e.g. Itakura, 1996, 2004; Emery, Lorincz, Perrett, Oram & Baker, 1997; Tomasello, Call & Hare, 1997). Chimpanzees even seem to exhibit the more advanced level of gaze following that requires ignoring a distractor object along the scan path – Butterworth's *geometric stage* of gaze following (see above) (Tomasello, Hare & Agnetta, 1999). In addition, Hare, Call, Agnetta and Tomasello (2000) demonstrated that chimpanzees know what conspecifics can and cannot see. There has also been some work with non-primates. Domestic dogs, for example, are capable of following the gaze of humans at about the level of 6- to 9-month-old human infants (but are not capable of shared attention) (Hare & Tomasello, 1999; Agnetta, Hare & Tomasello, 2000). In contrast, wolves don't seem to follow the gaze of humans (Hare, Brown, Williamson & Tomasello, 2002). Why some species are able to follow gaze while other species are not is currently unclear. Behavioral research has been cataloging cross-species differences but little is known about the underlying reasons for cross-species differences.

The role of learning

Early attempts to explain gaze following postulated the existence of innate modules. Examples of strongly nativist theories have been articulated by Leslie and Baron-Cohen (Leslie, 1987; Baron-Cohen, 1995). Such approaches have marginalized the role of learning in the development

of cognitive skills. One line of critique against modular accounts is that they tend to have little predictive power, because it is typically not made explicit how the modules work internally and exactly what information is passed between them (see Deák & Triesch, in press, for detailed analysis). In principle, however, this criticism can be overcome, and recent computational and robotic modeling work has started to address this question (Scassellati, 2002).

An alternative view explains the emergence of gaze following by postulating that infants gradually discover that monitoring their caregiver's direction of gaze allows them to predict where interesting visual events will be. This idea was first articulated by Moore and Corkum (1994; Corkum & Moore, 1995). Note that while this view highlights the role of learning processes, it does not preclude an evolved propensity to follow gaze in certain situations, which depends only minimally or not at all on early social experiences. Such mechanisms may be important in jump-starting the learning process. There is substantial evidence consistent with a learning account. In particular, Corkum and Moore (1998) (C&M) demonstrated that 8-month-old infants can be trained to follow their caregiver's gaze in a contingent reinforcement paradigm, where an interesting visual stimulus was shown if the infant followed the adult's gaze to the stimulus location. C&M concluded that 'learning could be involved in the acquisition of gaze following' (p. 37). A second experiment by C&M, however, seems somewhat inconsistent with a pure learning account. Specifically, they found it more difficult to train infants to look to the location opposite of where the adult turned. This prompts C&M to claim that 'simple learning is not sufficient as the mechanism through which joint attention cues acquire their signal value' (p. 28). In our view, however, C&M's second experiment is quite difficult to interpret and the results appear still consistent with a learning account.[1]

The importance of learning is also supported by some evidence, albeit preliminary, that gaze following skills emerge gradually through social experience. Deák *et al.* (2000) found that 12- and 18-month-old infants' gaze following diminished less across trials if targets were novel and distinctive, than if targets were repetitive and identical. This suggests that even in a single interaction with as few as 12 trials, infants adjust their expectations about the validity of adults' social cues for predicting visual reward. Also, Deák *et al.* (Deák, Wakabayashi, Sepeta & Triesch, 2004) reported preliminary observational data showing that gaze and gesture following skills emerge somewhat gradually between 5 and 10 months of age, which is consistent with an ongoing learning process. In sum, then, there is intriguing evidence to suggest that learning models might explain how gaze following and other joint attention skills emerge in the first 18 months. However, existing models are too vague to specify the kinds of data that would help us sharpen a powerful, predictive account of how these skills emerge.

The need for computational models

Our ultimate goal is to explain how and why gaze following (in its different forms) emerges at a level that reveals the underlying mechanisms of change in the brain and their relation to changes in overt social behavior. A theory of the emergence of gaze following should account for the experimental findings obtained in behavioral experiments, be consistent with known neuroscience data, and make specific predictions that can be used to falsify it. It should offer plausible explanations for differences in populations with developmental disorders and in other species. All else being equal, it should be as simple and parsimonious as possible.

In this paper we propose an account of the emergence of gaze following and evaluate its plausibility through computational modeling. Like many others, we believe that computational models can be a great aid in theorizing about developmental phenomena. The benefits of such an approach have been adequately discussed in several places (e.g. Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996; O'Reilly & Munakata, 2002). For instance, computational models can be very helpful in bridging the explanatory gap between biological mechanisms and observed behaviors. Importantly, computational approaches can be useful in analyzing the *causal structure* of developmental processes, that is, which changes may be necessary or sufficient for developmental events like the emergence of a new cognitive skill. These questions cannot easily be studied experimentally because (1) changes to individual neural processes are not readily observable or manipulable, and (2) there are typically many processes changing at the same time, making it very difficult to answer questions about cause and effect relations. Computational modeling may be particularly helpful in studying such relations because

[1] There are at least two questions about the proper interpretation of Experiment 2 in Corkum and Moore (1998). First, it is unclear to what extent the participants could already follow gaze, because the exclusion measure was not very powerful. Corkum and Moore's interpretation rests on the assumption that the tested infants were incapable of any gaze following. Second, motion cues may have facilitated gaze shifts in the direction of the caregiver's head turn, but Corkum and Moore's interpretation rests on the assumption that turns in the opposite direction are equally likely *a priori*. This does not consider that motion cueing facilitates gaze shifts in the same direction, which is supported by current evidence (e.g. Farroni *et al.*, 2000).

one can easily monitor all changes in the model, and systematically prohibit or promote certain changes in order to study how this alters the developmental trajectory.

The specific approach described in the following is comparable to other modeling work in the area of cognitive development. To some extent our approach is inspired by connectionist models (Elman *et al.*, 1996) and dynamical systems approaches to development (Thelen & Smith, 1994). We share with connectionist modelers the desire to explain behavior in terms of underlying neural structures. In contrast to classical connectionist models of development, however, our approach emphasizes aspects of the embodied nature of cognitive development (Clark, 1997; Wilson, 2002). In particular, we consider the role of the learner's situated real-time interaction with its environment. A good understanding and careful modeling of this interaction is a central goal of our approach (see Schlesinger & Parisi, 2001, for another example of this approach). These issues have also been addressed to some extent within the dynamic systems approach (Thelen, Schöner, Scheier & Smith, 2000), but our approach emphasizes the role of biologically plausible reward-driven learning processes. It is surprising to us that reward-driven learning mechanisms such as Temporal Difference learning (see below) are rarely being used in computational models of infant development. For example, connectionist style models typically utilize supervised learning (often using the backpropagation learning mechanism) which is not applicable to many developmental learning contexts. Similarly, in dynamical systems approaches, goal-directed learning is frequently not addressed either. Instead, the transition from one (younger and less capable) developmental state to the next (older and more capable) state is often modeled by changing a control parameter of the dynamical system in order to account for different performance levels. What is not addressed is what forces may drive these changing control parameters in developing infants. We feel that computational models that aim to carefully capture the affect-driven learning during situated, real-time interactions with the environment hold much promise for advancing our understanding of early cognitive development. The account that follows is an attempt to evaluate the promise of such models in the context of gaze following.

The *Basic Set* account of gaze following

At the heart of our account lies the idea that infants learn gaze following because they discover that monitoring their caregiver's direction of gaze allows them to predict where interesting visual sights occur. Elaborating on this idea, we propose that gaze following (and other attention-sharing skills) emerge through the interplay of a *Basic Set* of structures and mechanisms. This set includes perceptual skills and preferences, reward-driven learning, habituation and a structured social environment (Fasel, Deák, Triesch & Movellan, 2002). In the following, we will briefly discuss each component of this Basic Set, and review evidence that each of these is functioning in normally developing infants before the time that the first solid gaze following skills emerge. This is crucial for establishing the viability of this set as a causal precursor for the emergence of gaze following skills. We will then describe how these components may interact to allow for the learning of gaze following.

Perceptual skills and preferences

Several perceptual skills and preferences that are in place by 3 months of age or earlier might be important for shared attention skills to develop. Even the youngest infants prefer human stimuli, especially their caregivers' faces and voices (Brazelton *et al.*, 1974; DeCasper & Fifer, 1980; Pascalis, de Schonen, Morton, Deruelle & Fabre-Grenet, 1995). One interpretation is that social stimuli have a higher salience than competing inanimate stimuli (Bates, 1979). Infants also generally enjoy social interaction. Around 2–3 months, infants begin responding in a more consistent and focused way to caregivers. At the same time most infants produce their first social smiles, and parents report greater engagement and 'presence' during interactions (Cole & Cole, 1996). Infants as young as 3 months prefer looking at the eyes of an approaching person, rather than the mouth (Haith, Bergman & Moore, 1979).

Attention-shifting skills (critical for following gaze or pointing cues) begin to mature around 3–4 months (e.g. Butcher, Kalverboer & Geuze, 2000; Farroni, Johnson, Brockbank & Simion, 2000; Johnson, Posner & Rothbart, 1994), but other, more complex perceptual skills will continue to undergo significant changes. A skill that is highly relevant to the development of gaze following and other attention-sharing skills is face processing, or more specifically, head pose and eye direction perception (i.e. discriminating the rotational angles of the face, and estimating the line of gaze). One study found that 1-month-olds prefer a photograph of their caregiver's face in frontal to profile poses, suggesting that even young infants can discriminate extreme differences in caregivers' head poses (Sai & Bushnell, 1998). But this finding has not been extended, so we do not know how well infants of different ages can discriminate different head poses. It appears that 8–10-month-olds use head pose, not eye direction, to estimate adults' gaze direction

(Moore *et al.*, 1997). Robust use of the eyes seems to emerge later, with significant improvement between 12 and 14 months (Caron *et al.*, 2002). Thus, by this age, face processing skills must be sufficiently well developed to allow for robust gaze following even in somewhat ambiguous circumstances. However, for gaze following to be successful, the ability to accurately encode the caregiver's head pose needs to be mapped to the proper motor behaviors, which requires additional learning processes.

Reward-driven learning

Reward-driven learning, we claim, is important for learning attention-sharing. Reward-driven or reinforcement learning occurs in 2- and 3-month-olds (Kaye, 1982) and may even be present at birth (Floccia, 1997).[2] Two-month-olds can, for example, learn within minutes to predict the locations of the next interesting event in a simple repeated sequence (Haith, Hazan & Goodman, 1988). We propose that the principal learning mechanisms used for acquiring attention-sharing behaviors are neurally plausible processes of Reinforcement Learning called *Temporal Difference* or *TD learning* (Sutton, 1988; Sutton & Barto, 1998). These processes are not merely Skinnerean, nor are they anti-mentalistic, but they have the goal of formalizing the relation between an agent's affect-laden experienced outcomes (positive or negative) and the agent's means of adapting behavior to increase positive outcomes and decrease negative ones. TD learning in particular has been tied to specific neuromodulatory systems (Schultz, Dayan & Montague, 1997), and recent models are neurally plausible (Montague, Hyman & Cohen, 2004). In particular, the firing of dopaminergic neurons in parts of the basal ganglia has been associated with the temporal difference signal from which TD learning methods derive their name. Although TD learning has previously played almost no role in developmental models, it holds promise for understanding the development of behaviors in all contexts that involve affectively valued outcomes. Reward-driven learning, however, may not be the only learning mechanism that is important for the emergence of gaze following.

Habituation

Habituation also plays an important role in our theory as a fundamental learning process. Habituation processes have complex dynamics that are in themselves challenging to understand and to model (Sirois & Mareschal, 2002). In most previous modeling attempts, habituation was related directly to the behavioral responses of the organism, e.g. the strength or probability of a motor response to a certain stimulus. Our view is somewhat different in that we relate habituation processes to changes in the internal evaluation or reward of a stimulus. Together, habituation and reward-driven learning (see above) will produce certain behavioral sequences and modify them adaptively. For example, when an infant looks at a caregiver's face, or at a toy held by the caregiver, habituation will systematically occur, which we interpret as a systematically declining reward value over time for looking at this object. Dishabituation, conversely, amounts to a recovery of this reward. Because TD learning predicts future rewards, habituation will facilitate attention shifts away from the current target so that a new, more rewarding target can be fixated. Dishabituation leads to a relative recovery of the reward value of an object when a different stimulus is attended. These processes, in conjunction with reward-driven learning of behavioral policies, will produce cycles of attention-shifting between interesting social objects in the visual environment, such as the caregiver, and various other objects with properties that infants find interesting. The utility of these cycles for learning to follow gaze will depend on predictable behavior patterns provided by the caregiver.

Structured social environment

We posit that the most relevant situations for learning shared attention skills include interactions such as face-to-face play, feeding, diaper changing and bathing, which make up a high proportion of infants' waking time. What is important about such interactions, we hypothesize, is their predictable event-contingency structure. This structure is learnable, by means of reinforcement learning and habituation, and infants can learn to maximize their positive engagement in such interactions. Studies on the statistical structure of infant–parent interactions generally show that each participant synchronizes his or her actions with the other, and selects actions based partly on the other's recent actions, emotions and messages (Watson & Ramey, 1985). We hypothesize that infants soon start to predict where interesting objects and events will be, based on their caregivers' gaze patterns. The caregiver's gaze is predictive of interesting sights because caregivers will tend to look at other people or at objects they are manipulating (Land, Mennie & Rusted, 1999), and infants are interested in such stimuli.

[2] Sometimes the term *contingency learning* is used in the developmental literature. We use reinforcement learning because it is more common in neuroscience, cognitive science and machine learning, and because it makes explicit an assumption that is implicit in the idea of contingency learning – specifically, that the learner is *motivated* or affectively driven to predict, and experience, certain outcomes.

The emergence of gaze following

How can the Basic Set elements (perceptual skills and preferences, TD learning, habituation, structured social environment provided by caregivers) act in concert to allow gaze following to emerge? Our claim is that infants (or other developing organisms, or even robots) with these 'ingredients' will learn to anticipate the locations of interesting visual stimuli based on caregivers' attentive behaviors, both intentional (e.g. pointing) and unintentional (e.g. reflexive looking). They will learn to parse social events into conditions and outcomes, each associated with a hedonic value. A typical social sequence that supports learning might include the following events:

1. Initially, the caregiver and infant are looking at one another, in part because the infant has a preference for looking at social stimuli (i.e. it is rewarding to do so).
2. The caregiver looks away toward an object (possibly while holding or pointing to it), causing, first, a reduction in the reward value of the caregiver's face (making the infant more likely to search for other stimuli); and second, producing directional motion of the head or eyes, which can trigger a same-direction attention shift by the infant (Farroni *et al.*, 2000). Also, the infant may start to habituate to the caregiver's face, further biasing the infant to make a gaze shift.
3. In some of these cases, due to 'noisy' action selection or random exploration of different behaviors (e.g. Sutton & Barto, 1998), the infant makes a gaze shift in the same direction as the adult. This can result in the infant looking directly at the rewarding sight, or it can bring the sight into the field of view so that a subsequent eye movement can bring it to the center of gaze.
4. In these cases, the infant on average receives a relatively greater reward (in terms of interesting sights) than if he or she had selected other actions. In a 'high-reward sequence', infants receive information about contingencies between the caregiver's head pose and the presence of interesting visual events in a certain location. This allows infants to learn that it is beneficial to follow caregivers' gaze shifts by shifting their own gaze to the same regions of space.

In summary, we propose that the Basic Set of structures and mechanisms outlined above allows infants to learn to follow gaze because they learn to exploit the caregiver's tendency to look at things that are interesting (rewarding) for the infant. This theory is geared to explain the basic phenomenon of gaze following, i.e. how the infant learns to associate the head pose of others with gaze shifts to certain locations inside or outside of its own visual field. Ultimately, the test of this theory will be whether it can be extended to explain many of the interesting subtleties such as the ordered sequence of the development of gaze following skills, or the value of different caregiver cues (eyes, face, body posture) for joint attention, or the later development of theory-of-mind-like representations. We are optimistic that our framework provides a good starting point for this endeavor, and that we will eventually be able to account for a large range of empirical phenomena, including 'higher' shared attention skills. We will return to this point in the discussion.

Computational model

We now present a simple computational model to test whether the mechanisms of the Basic Set can lead to the emergence of gaze following and to explore how alterations of model parameters can simulate some developmental disorders that are characterized by delays in the emergence of gaze following.[3] The goal of this inquiry is to determine under what conditions the Basic Set is *sufficient* for the emergence of gaze following. We do *not* suggest, however, that all of the Basic Set elements are strictly *necessary* – some might be replaceable by alternative mechanisms. Also, we do not claim that this set is sufficient for a comprehensive account of *all* human attention-sharing behaviors. It merely attempts to explain the basic gaze following behaviors that progressively emerge during the first year in typically developing infants, and, hopefully, disruptions of this progression that occur in certain developmental disorders. Future work will establish whether the model can also explain, for example, point-following behaviors.

The model was implemented in Matlab. The source code is available at http://mesa.ucsd.edu

Environment and caregiver model

The simulation comprises a model of the infant (referred to simply as 'infant', merely for expositional fluency), a model of the caregiver (the 'caregiver') and a model of the environment in which they interact. An overview of the model is given in Figure 1. As a simplification in the model, we assume that infant and caregiver are facing each other and remain in the same position. The space surrounding infant and caregiver is discretized into N distinct regions. The caregiver can look at any of these regions or at the infant. The infant can look at any of

[3] An initial account of the model was given in Carlson and Triesch (2003).

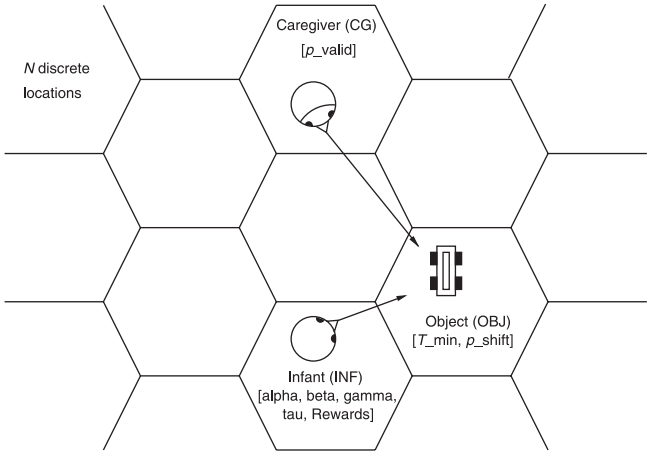

Figure 1 *Overview of the model showing infant, caregiver and interesting object. Corresponding model parameters are given in brackets. Note that while we draw the spatial locations as arranged in a hexagonal fashion, the model does not assume or use any specific topological relations between these locations.*

these regions or at the caregiver. The infant's and caregiver's shifting of gaze are the only ways they interact with each other and the environment. Time runs in discrete steps, each corresponding to roughly a quarter of a second. Each gaze shift is assumed to take one time step.

At any time there is one interesting object present or event occurring in one of the N regions of the environment. This could be an interesting toy, a third social agent, the caregiver's hand manipulating an object or performing a gesture, or other stimuli that the infant would find interesting. We will refer to this object or event as the *target*. (Below we will also consider environments with multiple targets.) After some minimum time at one location $T_{\min}$, the interesting target is relocated to a randomly chosen new location with some probability p_{shift} per time step.

Whenever the target moves, the caregiver model shifts its direction of gaze. There is a certain probability p_{valid} that the caregiver will be looking at the new location of the target. Otherwise, the caregiver's new direction of gaze is drawn from a uniform distribution over all of the other N locations (one for the infant plus $N - 1$ locations not containing the target). Thus, the parameter p_{valid} models how predictive the caregiver's direction of gaze is for indicating the location of the interesting target.

The parameter p_{valid} also has a second function. We can use it to model inaccuracies in the infant's head pose discrimination. Consider the case where the caregiver is always looking at the target. Even in this case, if the infant's head pose discrimination is inaccurate or noisy, the infant will not be able to correctly infer the caregiver's head pose and, as a consequence, the *estimated* head pose will not be very predictive of rewarding sights. Thus, a not-so-predictive caregiver whose head pose can be estimated accurately and a highly predictive caregiver whose head pose we can only infer correctly some fraction of the time will produce the same net effect, and we can model both situations with the same parameter p_{valid}.

Note that this environment and caregiver model is extremely simple. In particular, the caregiver is not responding to the infant in any way. This is obviously a gross simplification of the complex, reciprocal dynamics of infant–caregiver interactions (e.g. Kaye, 1982), but as we will demonstrate below, even this kind of social environment can be *sufficient* for gaze following to emerge. More complex, interactive caregiver models have also recently been investigated, and these show that the caregiver's behavior plays an important role (Teuscher & Triesch, 2004). In particular, the caregiver's behavior has to be properly matched to the parameters of the infant model for optimal learning speed, although gaze following will emerge under a wide range of caregiver behaviors.

Infant model

Our infant model is essentially that of a *pleasure-driven agent*. There are many ways of formalizing this idea but a particularly appropriate formal framework is reinforcement learning (Sutton & Barto, 1998). Besides being the basis for modern theories of learning under rewards and punishments, reinforcement learning is also an important subfield of machine learning with some impressive application successes (Sutton & Barto, 1998). In particular, our model uses temporal difference learning (TD learning) algorithms, which have been proposed as models for certain basal ganglia functions (Schultz *et al.*, 1997). A detailed description of the equations of the model is given in the Appendix.

We conceive the infant as a reinforcement learning system that learns to make two kinds of decisions. First, at any given time it decides whether to shift gaze or keep fixating the same location. Second, it decides where to look next, once the decision to shift the direction of gaze has been made. The information available to the infant includes the identity of its current object of fixation, its associated reward value, and the length of time the infant has been fixating this object. If and only if the fixated object is the caregiver, the infant will know the caregiver's current head pose.

Looking, reward and habituation

The infant model receives rewards for looking at interesting things. The amount of reward received depends

on the contents of the infant's gaze and how habituated the infant is to those contents. There are four possible things for the infant to see, (1) a frontal view of the caregiver (in case the caregiver is also looking at the infant), (2) a non-frontal view of the caregiver, which we simply refer to as a profile view (in case the caregiver is not looking at the infant), (3) the target or (4) nothing. Associated with these sights are the *base rewards* R_{frontal}, R_{profile}, R_{target}, R_{nothing}. The *actual reward* received by the infant is the base reward attenuated by habituation. As the infant looks at a location, the infant habituates to its contents in the sense that the actual reward for any object at this location will decrease over time. Similarly, dishabituation is modeled as a recovery of the actual reward for objects at other locations.

For each object in the environment, including the caregiver, the infant has a habituation value $h_{\text{fix}}(t) \in [0,1]$, indicating the fraction of the base reward the infant receives for looking at this object. A value of $h_{\text{fix}} = 1$ means that the infant is not habituated to the object, while a value of $h_{\text{fix}} = 0$ means that the infant is completely habituated to the object. As the infant continues to fixate on an object its habituation value decreases according to $h_{\text{fix}}(t) = h_{\text{fix}}(0)e^{-\beta t}$, where $h_{\text{fix}}(0)$ is the habituation level at the beginning of the current fixation, and t is the time since the start of the fixation, and β is the *habituation rate*. Thus, the actual reward received by the infant at time t is $r_{\text{actual}}(t) = R_{\text{fix}}h_{\text{fix}}(t)$, where $R_{\text{fix}} \in \{R_{\text{frontal}}, R_{\text{profile}}, R_{\text{target}}, R_{\text{nothing}}\}$ is the base reward. At the same time, the reward levels for objects at locations not being fixated recover in a corresponding fashion, modeling dishabituation. In particular, when the infant is not looking at an object it dishabituates according to $h_{\text{nofix}}(t) = 1 - h_{\text{nofix}}(0)e^{-\beta t}$, where t is the time since last looking at that object and h_{nofix} is the level of habituation of this object currently not being fixated.

One infant, two agents: when and where

Inspired by the proposal that the decisions of when to shift gaze and where to shift gaze are made in separate neural pathways (Findlay & Walker, 1999), the infant model consists of two separate agents. The state space of the *when-agent*, which decides whether to continue to fixate on the same location or shift gaze, has two dimensions. The first dimension represents the time the infant has been fixating at the same location, discretized as the number of time steps (0, 1, 2, . . . , 8, 9 or more). The second dimension is the actual reward received by the infant. This is the total reward the infant receives on that time step, taking habituation into account, discretized uniformly into ten bins between the maximum and minimum possible actual rewards.

If the when-agent makes the decision to shift gaze, the *where-agent* determines the target of the gaze shift. The state space of this agent has only a single dimension: the caregiver's head pose. Importantly, unless the infant is looking at the caregiver, the caregiver's head pose will be unknown to the infant. Concretely, this agent distinguishes $N + 2$ different states: N for the N different head poses observed when the caregiver looks at the N regions of space, plus one for the caregiver's head pose when the caregiver is facing the infant, plus one state to represent that the head pose of the caregiver is unknown to the infant. The where-agent learns to map these states onto $N + 1$ different actions: one action for looking at each of the N regions of space and one action for looking at the caregiver. Note that we assume a one-to-one correspondence between a caregiver head pose and the region of space the caregiver looks at. In reality, this mapping is ambiguous and the ambiguity can produce characteristic errors in gaze following (Butterworth & Jarrett, 1991). Modeling this ambiguity and how the infant learns to resolve it is the subject of a separate paper (Lau & Triesch, 2004).

Learning in both agents occurs through the SARSA algorithm (see Appendix), which was chosen because of its simplicity. Both agents balance exploration vs. exploitation by selecting actions with a softmax action selection mechanism (see Appendix). It should be noted that separating the infant model into two separate learning agents is not strictly necessary. We would expect similar results for a simpler model that uses a single reinforcement learning agent to model the infant, whose state space was the product space of the state spaces of the when and where agents, and whose possible actions are to shift gaze to any of the $N + 1$ locations. However, the learning time would be expected to increase because of the higher dimensionality of the resulting state space.

Experiments

Normal emergence of gaze following

In this section we describe a first analysis of the model and the effects of some model parameters on its learning behavior. For easy reference, all parameters, their default values, and their allowed ranges are listed in Table 1. In the following, default parameter values are used unless otherwise indicated. The effect of changing several parameters is discussed below. Generally speaking, the model is robust to changes in the parameters over wide ranges. The parameters T_{min}, p_{shift} and p_{valid} were set *ad hoc* but could eventually be set in accordance with data from an observational study of naturalistic

Table 1 *Overview of model parameters, their allowed ranges and default values*

Symbol	Explanation	Range	Default
N	number of spatial regions	1, 2, . . .	10
Δt	duration of one simulation step	arbitrary	~250 ms
α	learning rate	[0,1]	0.0025
β	habituation rate	[0,∞]	1
γ	discount factor for future rewards	[0,1]	0.8
τ	temperature (randomness of action selection)	[0,∞]	0.095
R_{frontal}	reward for looking at frontal view of caregiver	[−∞,∞]	1
R_{profile}	reward for looking at profile view of caregiver	[−∞,∞]	1
R_{target}	reward for looking at target	[−∞,∞]	1
R_{nothing}	reward for looking at other region	[−∞,∞]	0
T_{min}	minimum target stationary time (steps)	[0,∞]	4
p_{shift}	probability of target shift per time step	[0,1]	0.5
p_{valid}	predictiveness of caregiver gaze	[0,1]	0.75

infant–caregiver interactions that is currently under way (Deák *et al.*, 2004).

To quantify the emergence of gaze following in the model and its dependence on model parameters we use the following approach. At specific points during the learning process we temporarily 'freeze' the model and evaluate its behavior for 1000 time steps (which corresponds to slightly more than 4 minutes of simulated interaction), after which the learning process resumes. The model behavior at these stages of the learning process is analyzed by observing the infant model interacting with the environment and computing two statistics. The *caregiver index* CGI is defined as the frequency of the infant's gaze shifts towards the caregiver:

$$\text{CGI} = \frac{\#\text{ gaze shifts to caregiver}}{\#\text{ gaze shifts}}. \quad (1)$$

The *gaze following index* GFI is the frequency of gaze shifts that lead from the location of the caregiver to where the caregiver is looking:

$$\text{GFI} = \frac{\#\text{ gaze shifts from caregiver to correct location}}{\#\text{ gaze shifts}}. \quad (2)$$

An example run of the system with the default parameters is shown in Figure 2. The model first learns to alternate gaze between the caregiver and other locations. In terms of the model, the when-agent discovers that it is best not to continue staring at a single location for too long. At the same time, the where-agent discovers that if the infant is not looking at the caregiver it tends to be rewarding to make a gaze shift back to the caregiver. After this has been achieved, gaze following behavior slowly emerges. Here, the where-agent discovers that unexpectedly high rewards tend to follow gaze shifts to certain locations, depending on the caregiver's head

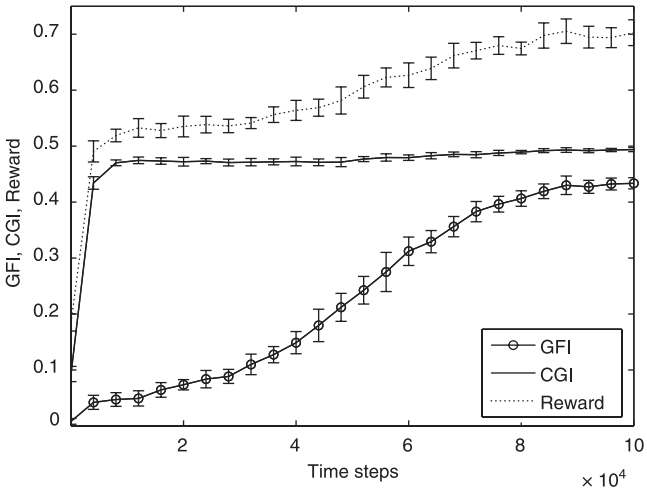

Figure 2 *Emergence of gaze following in simple environment with just one interesting target present at any time. The solid curve plots the caregiver index (CGI), the solid curve with circles plots the gaze following index (GFI) and the dotted curve plots average reward per time step, as functions of the number of learning iterations. Error bars indicate standard deviations across 15 simulations.*

pose. It learns to correctly map the caregiver's head pose to gaze shifts to the locations that the caregiver looks at. The increasing average reward the model obtains per time step during this phase confirms that gaze following is in fact beneficial for the model under these parameters. Note that for a model without habituating rewards it would be optimal to continually stare at the caregiver.

A microscopic view of the behavior of the infant model is shown in Figure 3 (top). It shows the fixation behavior of the infant during various stages of the learning process. Fixations on the caregiver are indicated by white pixels, target fixations by black pixels, and fixations on other regions of space by grey pixels. The quick

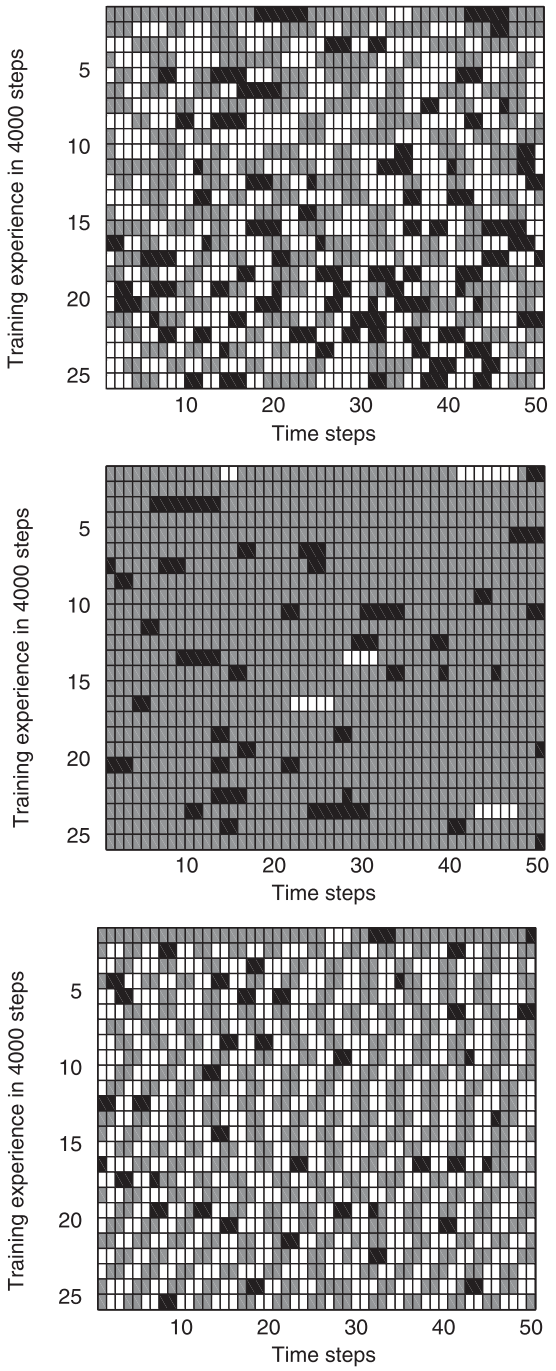

Figure 3 *Microscopic analysis of model behavior for normally developing (top), autism-like (center) and Williams-like (bottom) model. Each row of pixels shows the target of the infant's gaze as a function of time (for 50 steps). The gaze target is color coded, with white corresponding to the caregiver, black corresponding to the target, and grey corresponding to other regions of space. In particular, an instance of gaze following is represented by a black pixel lying to the right of a white pixel. Different rows show the behavior at different times during the learning process (every 4000 steps).*

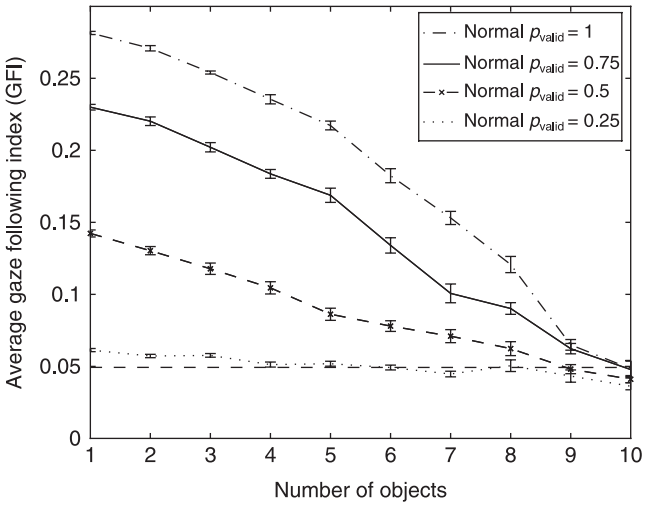

Figure 4 *Gaze following in the presence of multiple targets for various values of* p_{vaild}. *The gaze following performance averaged over 100 000 steps (y-axis) is plotted as a function of the number of targets that are present simultaneously (x-axis). Error bars indicate standard error across 15 simulations. Gaze following is diminished if significant ambiguities due to multiple targets exist. Also, a reduced predictiveness of the caregiver* p_{vaild} *has a negative impact on gaze following performance. The dashed horizontal line marks the 'chance level' of gaze following expected for an infant who first looks to the caregiver and then shifts gaze randomly to any of the* N *locations.*

development of a preference for looking at the caregiver is visible as the increase in the amount of white pixels (caregiver fixations) during the first few rows. The subsequent increase in target fixations (black pixels) is the effect of the emergence of gaze following. Gaze following episodes are shown by black pixels to the right of white pixels.[4] The increase in the number of such episodes during learning directly reflects the increasing GFI (compare Figure 2).

Figure 4 shows that gaze following will still be learned in more complex environments, where multiple interesting events occur simultaneously. In this case, the learning is somewhat slower because the infant may temporarily learn incorrect associations between a particular caregiver head pose and a gaze shift to a location not looked at by the caregiver but that nevertheless contains an interesting event.

[4] Note that there can be instances of black pixels to the right of white pixels that do not correspond to gaze following. This occurs when the infant looks away from the caregiver to a location not looked at by the caregiver that happens by chance to hold the interesting object. These instances are comparatively rare, however. More precisely, the probability of the infant finding the target this way is only $(1 - p_{valid})/(N - 1)$, where N is the number of locations in the environment.

We have also experimented with making $R_{profile}$ smaller than $R_{frontal}$ to capture infants' preference for frontal faces (Sai & Bushnell, 1998). We found that gaze following performance is largely determined by $R_{profile}$, with higher $R_{profile}$ values leading to faster learning. The value of $R_{frontal}$ plays a comparatively small role, because the current caregiver model only looks at the infant infrequently. A systematic analysis of learning speed as a function of caregiver reward is given below in the context of modeling developmental disorders.

Analysis of model parameters

Predictiveness of caregiver's gaze

An important parameter of the model is p_{valid} (see Figure 4). Unless p_{valid} is high enough, gaze following will not emerge. For $p_{valid} = 0.25$, the GFI remains very poor, even when there is only one interesting target in the environment. There are two interpretations of this result, corresponding to the two interpretations of p_{valid} (see above). First, a highly informative caregiver, i.e. one who frequently looks at the interesting target, facilitates the acquisition of gaze following. This confirms the importance of one component of the Basic Set: a structured social environment. Second, limitations of the infant's ability to discriminate head poses will delay the infant's acquisition of gaze following. Currently, little is known about how real infants' ability to discriminate head poses develops, but such data would be most useful in constraining the model (see also Lau & Triesch, 2004).

Speed of learning: learning rate and habituation

We hypothesized that the learning rate α and the habituation rate β might both influence the speed with which gaze following can be acquired. In the trivial case of $\alpha = 0$ no learning takes place at all, and gaze following obviously cannot emerge. However, too high a learning rate can also cause problems. This is illustrated in Figure 5, top. In general, an intermediate value for the learning rate seems to be optimal, which is common for reinforcement learning models.

Figure 5, bottom, shows the effect of the habituation rate β on the learning process. It shows that an infant that habituates faster (high β) learns to follow gaze more quickly. By contrast, slow habituation (low β) will result in less frequent gaze shifts between objects and therefore to fewer opportunities for the necessary learning experiences. Interestingly, however, even without any habituation ($\beta = 0$) gaze following is still learned, but very slowly. In this case, gaze shifts away from the most rewarding object occur only through the random selection of exploratory actions. The infant will spend most time looking at the caregiver, which is the optimal thing to do. Due to the random softmax action selection mechanism, however, which sometimes explores the consequences of seemingly suboptimal actions, the infant will look away from the caregiver, which creates an opportunity to discover the benefit of following gaze. We conclude that although habituation is not strictly necessary if there are

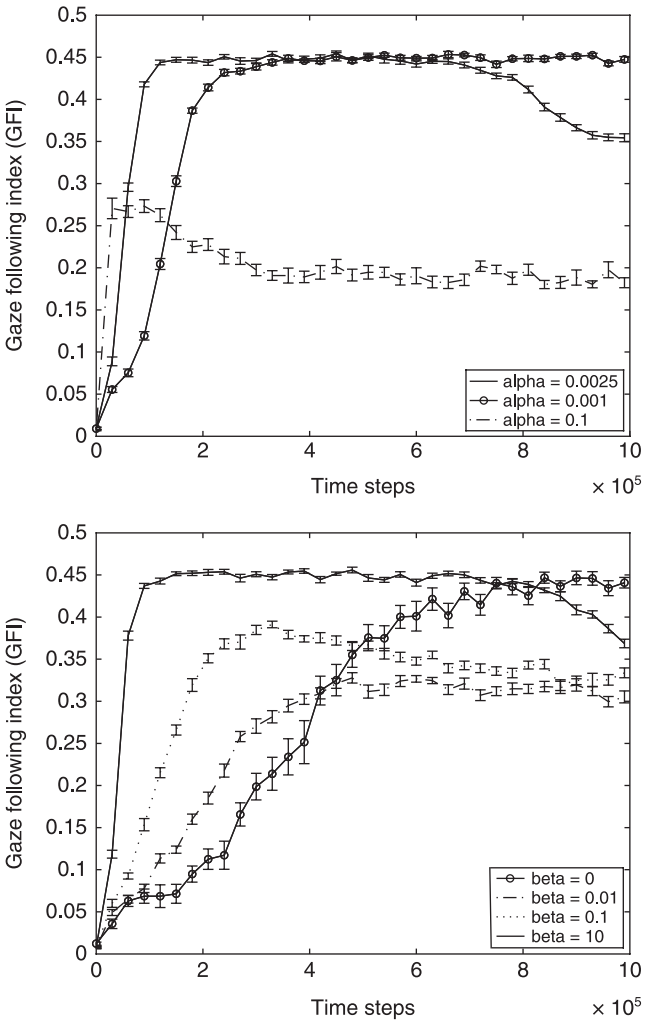

Figure 5 *Top: Effect of learning rate on emergence of gaze following. A higher learning rate α leads to accelerated initial learning as measured by the gaze following index (GFI). However, a high learning rate can lead to problems in the long run. The infant may never acquire a high level of gaze following. Error bars indicate standard errors across 15 runs. Bottom: Effect of habituation rate on learning of gaze following. Faster habituation leads to accelerated learning as measured by the gaze following index (GFI). Even without any habituation gaze following is still learned – albeit very slowly. Error bars indicate standard error across 15 simulations.*

other mechanisms for exploratory gaze shifting, learning may be very slow without it. The model thus predicts that infants who habituate quickly (in the sense of the model) may learn gaze following faster than their peers. This prediction is consistent with some evidence that infants who are 'fast habituators' at 5 months have better social and communicative skills at 13 months (Tamis-LeMonda & Bornstein, 1989), although care has to be taken because our notion of habituation as a decaying reward for a visual stimulus is not identical to the common behavioral measures of habituation.

In summary, both learning rate and habituation rate influence the speed of learning and may be related to individual differences in the emergence of gaze following in real infants. However, they act on the learning process in different ways. The learning rate α determines how much an individual learning experience changes the infant's future behavior. The habituation rate β determines how many relevant learning experiences the infant encounters during a fixed amount of time.

Modeling failures of the emergence of gaze following in autism and Williams syndrome

Any account of gaze following should answer why gaze following emerges, and why gaze following may not emerge under certain circumstances. An important line of research concerns differences in shared attention skills in developmental disorders such as autism and Williams syndrome. Autism is a *Pervasive Developmental Disorder* characterized by impairment in social interactions and communication (e.g. Dawson, Meltzoff, Osterling, Rinaldi & Brown, 2004), as well as atypical cognitive processing. Shared attention deficits are the most consistent early predictors of the social and language deficits of autism (Osterling & Dawson, 1994). Thus, a critical test of our model is its capacity to simulate autistic failure of gaze following.

A more subtle challenge is to test the model's capacity to simulate a disorder that is associated with less striking and more idiosyncratic differences in joint attention. Williams syndrome is a rare genetic disorder that is characterized by (among other things) hypersocial behavior, differences in face processing and deficits in learning and attention. Most importantly for us, there is also some evidence for deficits in triadic shared attention skills (Bertrand, Mervis, Rice & Adamson, 1993; Laing, Butterworth, Ansari, Gsödl, Longhi, Panagiotaki, Paterson & Karmiloff-Smith, 2002; Mervis, Morris, Klein-Tasman, Bertrand, Kwitny, Appelbaum & Rice, 2003), although more research is needed in this area.

While traditional nativist/modularist accounts typically propose broken or missing modules as the origin of developmental disorders (Baron-Cohen, 1995), our account prompts us to look for potential differences in the components of the Basic Set that may lead to different developmental trajectories. The goal here is not to provide a comprehensive model of these developmental disorders, but to show how *specific aspects* of these disorders may contribute to deficits in gaze following.

Changes in the reward structure

In the last section we have already seen how differences in learning rate or habituation rate can slow down or even prevent the emergence of gaze following. For autism spectrum disorders and Williams syndrome, however, a particularly interesting candidate is the reward structure of the model, because in both kinds of disorders the affective value of faces may be altered.

An intriguing attribute of autism is disinterest in faces. In general, the interest in or appeal of social stimuli is diminished in autism (Adrien, Lenoir, Martineau, Perrot, Hameury, Larmande & Sauvage, 1993; Chawarska, Klin & Volkmar, 2003; Maestro, Muratori, Cavallaro, Pei, Stern, Golse & Palacio-Espasa, 2002; Tantam, Holmes & Cordess, 1993; Klin, Jones, Schultz & Volkmar, 2003; Dawson, Meltzoff, Osterling, Rinaldi & Brown, 1998). For some (but not all) individuals with autism, direct eye contact even seems to be aversive, a phenomenon known as *gaze avoidance* (Hutt & Ounsted, 1966; Richer & Coss, 1976; Langdell, 1978). It has been proposed many times that a disruption in face processing may be an underlying cause for social deficits in autism (e.g. Trepagnier, 1996; Howard, Cowell, Boucher, Broks, Mayes, Farrant & Roberts, 2000; Klin *et al.*, 2003). Why faces are in some ways less salient or rewarding to individuals with autism is not clear. It may be that faces are too unpredictable for autistics, an idea consistent with the hypothesis that autistics prefer highly predictable stimuli (Gergely & Watson, 1999); it may also be that anatomical differences in the amygdala (which participates in processing facial affect displays) play a role (e.g. Howard *et al.*, 2000; Baumann & Kemper, 2005). Regardless of the cause, this symptom, and its long-term effect on social learning, bears more precise (ideally quantitative) specification.

In contrast to the disinterest in faces in autism, children with Williams syndrome show a high preference for looking at faces over looking at other objects (Bertrand *et al.*, 1993; Bellugi, Lichtenberger, Jones, Lai & St George, 2000; Mervis *et al.*, 2003). In addition, altered as well as delayed emergence of face processing skills has been reported (Karmiloff-Smith, Thomas, Annaz, Humphreys, Ewing, Brace, Van Duuren, Pike, Grice & Campbell, 2004).

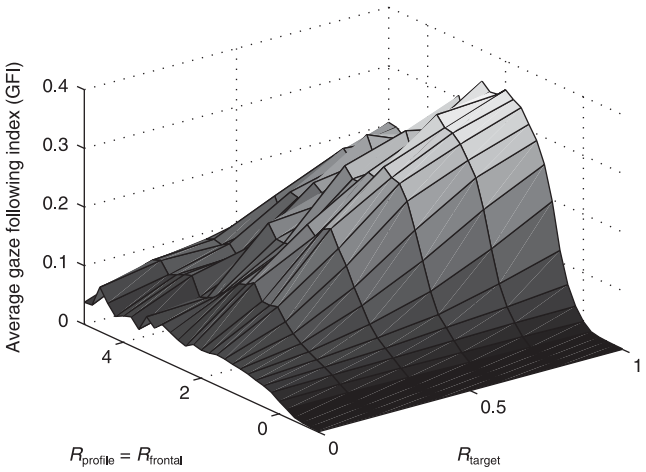

Figure 6 *Learning performance as a function of caregiver and target reward. For the caregiver reward we use* $R_{frontal} = R_{profile} \equiv R_{caregiver}$. *The z-axis corresponds to the GFI after* 10^5 *time steps of learning, averaged over 10 repetitions of the experiment.*

What would happen in the model if looking at the caregiver was made aversive, as for an atypical baby who finds faces unpredictable and overstimulating, or made highly positive, as for a hypersocial infant with an extreme preference for human faces over other sights?

To test the effect of different reward structures on the learning process, we systematically varied the reward parameters R_{frontal}, R_{profile} and R_{target} over a range of values. For simplicity we restricted ourselves to the case where $R_{\text{profile}} = R_{\text{frontal}}$. Figure 6 summarizes the results. For each combination of reward values we ran the simulation for 10^5 time steps and measured the GFI at the end of this time. Figure 6 plots the GFI averaged over 10 experiments as a function of R_{frontal} and R_{target}.

For $R_{\text{target}} \leq 0$ no gaze following behavior emerges. This makes intuitive sense because if the targets that the caregiver tends to look at are not rewarding for the infant, there is no benefit in gaze following behavior. That is, no additional reward can be obtained by following the caregiver's gaze. If R_{frontal} and R_{profile} are small or even negative, modeling reduced interest in or aversion to faces as seen in autism, gaze following behavior does not develop normally. Depending on the caregiver and target reward, the infant model will spend little time looking at the caregiver. For example, while the 'normal' model with a base reward of 1 for the caregiver (frontal and profile) and for the target spends 49% of its time looking at the caregiver and 14% of the time looking at the target (averaged over the entire learning period), the 'autistic-like' model with caregiver reward of −1 will spend only 1% of its time looking at the caregiver and 11% looking at the target (which it occasionally finds by chance without utilizing the caregiver's gaze). As a consequence, the learning process is slowed down or even prevented, and the GFI stays close to zero. The microscopic behavior of such a model is shown in Figure 3 (middle). Thus, a reduced reward for looking at the caregiver's face or aversiveness of the caregiver is sufficient to explain delays or complete failure in the emergence of gaze following.

It is interesting to note that an analysis of the model shows that even for negative caregiver rewards, the model will nevertheless slowly learn how to follow gaze, even if it does not exhibit the behavior on a regular basis. By analyzing the infant's action selection probabilities we found that the probability for following the caregiver's gaze once the infant is looking at the caregiver slowly but clearly rises above those for other actions. However, the model rarely executes a complete gaze following sequence because it is unrewarding to do so, due to first having to look at the aversive caregiver. This behavior of the model might explain a puzzling finding by Leekam, Baron-Cohen, Perret, Milders and Brown (1997) that autistic children can follow gaze if explicitly told to do so, though they may rarely do it spontaneously. This finding is very problematic for previous accounts of the emergence of gaze following. We know of no theory that offers a satisfactory explanation for it. Subsequent studies by Leekam and colleagues (Leekam *et al.*, 1998; Leekam, López & Moore, 2000) suggest that autistic children can be trained to follow gaze through contingent presentation of rewarding visual stimuli (Whalen & Schreibman, 2003), but that a lack of motivation to engage with the experimenter may impede learning. These findings are also consistent with our account. The association from caregiver head pose to regions in space is learned (although slowly) due to the constant low level of random exploration, but gaze following is simply not rewarding enough to be produced on a regular basis. If, however, an additional incentive for following gaze is present (e.g. being asked to look where another person is looking, or being trained via operant conditioning), the behavior can be elicited. Also, it is in line with the finding that gaze following in response to static pictures may be 'easier', if we make the additional assumption that static pictures of faces are not as aversive as dynamic displays (Klin *et al.*, 2003).

It should be noted that an infant who looks less at faces due to a diminished reward for faces can be expected to develop deficits in face processing skills such as fine discrimination of head poses or estimation of the direction of gaze. This will likely corroborate delays in the emergence of gaze following. The model could capture this by making the parameter p_{valid} a function of the total amount of time the infant has been looking at the caregiver.

We also tested what would happen if the reward for looking at the caregiver is much higher than the reward for looking at the target. This manipulation may be thought of as an attempt to model differences in Williams syndrome, where children exhibit an abnormally high preference for faces. Our experiments with the model show that in this case, somewhat surprisingly, the learning of gaze following can be substantially delayed (Figure 6). To give an example, a 'Williams-like' model with a base reward of 5 for looking at the caregiver and a base reward of 0.5 for looking at the target will spend 51% of its time looking at the caregiver but only 5% looking at the target. Thus, little gaze following will be observed, as illustrated in Figure 3 (bottom). The reason is that because the caregiver is relatively so rewarding to look at, it makes little difference to the infant where it looks in between fixations on the caregiver: the probability of looking at the target is only slightly higher than the probability of looking at any other region of space under the model's probabilistic action selection rule.

Deficits in attention-shifting

Another important aspect of autism spectrum disorders are deficits in shifting attention. For example, many studies have shown that people with autism are slower to shift attention between targets (e.g. Casey, Gordon, Manheim & Rumsey, 1993; Wainwright-Sharp & Bryson, 1993; Goldberg, Lasker, Zee, Garth, Tien & Landa, 2002; Landry & Bryson, 2004). This deficit might be related to cerebellar abnormalities (Harris, Courchesne, Townsend, Carper & Lord, 1999). Slow attention shifting can be incorporated into the model in the following way. Instead of gaze shifts taking effect immediately, we introduce a latency T_{lat} of 1 to 3 time steps. After the infant makes a decision to shift gaze, it has to wait T_{lat} time steps before the gaze shift takes effect. Figure 7 shows how this affects the emergence of gaze following. In these experiments all other parameters were set to their default values. The error bars indicate standard errors of 15 independent simulations per condition. As can be seen in the figure, the additional latency can slow down or even prevent the emergence of gaze following behavior, because there is a growing probability that by the time the infant shifts gaze, the rewarding sight has moved to a different location. This effect is clearly visible in infants with a normal, positive caregiver reward (Figure 7, top). However, it is more pronounced for a caregiver reward of zero, i.e. infants who find their caregivers uninteresting but not aversive (Figure 7, bottom), and it is even more pronounced for a model with negative caregiver reward (not shown). These results and the previous ones show that either different reward structures,

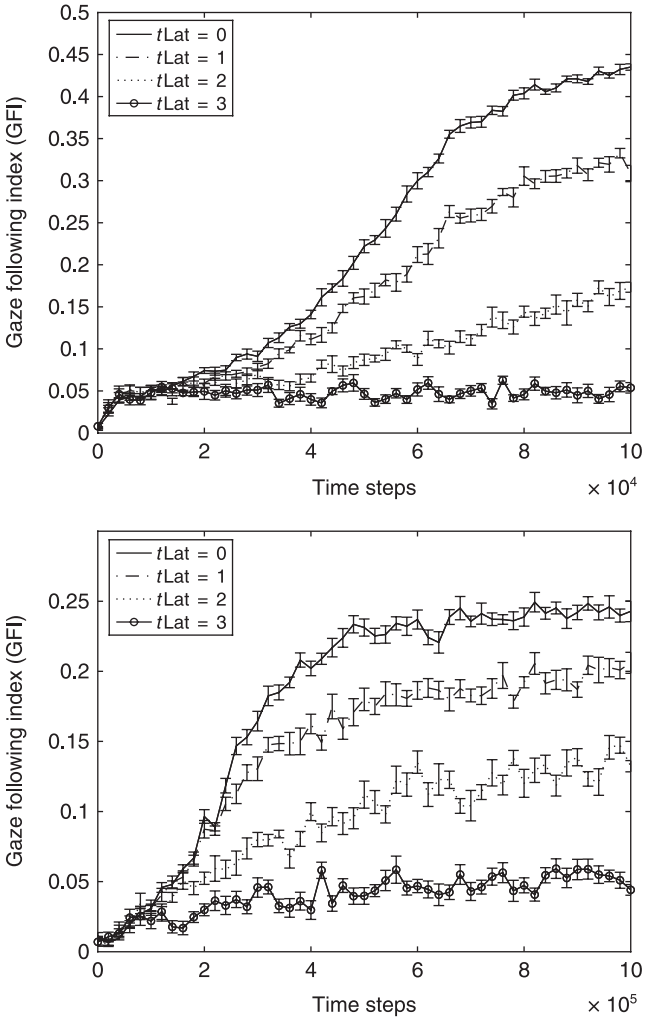

Figure 7 *Learning performance for infant models with attention shifting deficits of varying degree. Top: for normal, positive caregiver reward. Bottom: for zero caregiver reward. Note the different scales on the axes. Error bars indicate standard error across 15 simulations.*

or poor attention-shifting, or both, can explain gaze following deficits in autism within the proposed model.

Regarding Williams syndrome, a noteworthy recent report on the perception of faces in adults with Williams syndrome finds less accuracy in determining the direction of gaze, and significantly longer response latencies during face perception (Mobbs, Garrett, Menon, Rose, Bellugi & Reiss, 2004). Given our results above, we can conclude that both of these symptoms, if present in infants, would corroborate problems in the emergence of gaze following. Less accuracy in determining the direction of gaze will lower the predictiveness of the caregiver (smaller p_{valid}), while longer response latencies can be thought of as increasing T_{lat}. In a similar vein, recently observed inaccuracies of saccade targeting and a higher

number of corrective saccades in Williams syndrome (van der Geest, van Haselen, van Hagen, Govaerts, de Coo, de Zeeuw & Frens, 2004) may also contribute to longer latencies before the target of a gaze shift is reached, corroborating difficulties in learning to follow gaze.

Summary

To summarize, simple manipulations to the reward structure and attention shifting behavior of the model motivated by findings on two very different developmental disorders lead to deficits in the emergence of shared attention. What is needed for further constraining the model is more experimental data on how, for example, the accuracy of infants' head pose discrimination, or the preference for viewing frontal vs. profile faces develops for normally and atypically developing infants.

Summary of model predictions

Although our model is simple and incorporates only well-known and accepted infant skills, it makes a number of novel predictions, summarized below. The list is certainly not exhaustive, since there are many ways of manipulating the model (we invite readers to download the software from http://mesa.ucsd.edu and derive new predictions). Of course, not all predictions of the model will lend themselves to experimental investigation, and some manipulations would be unethical to do with real infants. Leaving these concerns aside, the model makes the following predictions.

1. *Fast habituation leads to quicker acquisition of gaze following.* The systematic variation of the habituation parameter β showed an advantage in learning speed for faster habituation. Fast habituation in the model leads to more gaze shifts per time interval on average, which produces more opportunities to learn the predictive value of the caregiver's direction of gaze, all else being equal.
2. *Face perception skills should correlate with gaze following ability.* One interpretation of the parameter p_{valid} was that it reflected accuracy of head pose estimation in infants. The model showed that without a sufficiently high p_{valid}, gaze following will not emerge.
3. *Infants with general learning deficits should also have an impairment in the acquisition of gaze following.* Choosing too small a learning rate in the model leads to delays in the emergence of gaze following. Not surprisingly, though, too high a learning rate was also found to be maladaptive.
4. *Infants whose visual preferences do not match their caregivers' should have deficits in gaze following.* The model shows that if the reward values associated with the objects/events that caregivers tend to look at are not higher than those for random locations, gaze following will not emerge. By the same token, infants whose caregivers produce few predictive gaze cues (e.g. due to visual deficits) should also learn gaze following more slowly.
5. *Infants who find faces too attractive should have deficits in gaze following.* Using a caregiver reward much higher than the target reward leads to deficits in gaze following in the model.
6. *Infants who find faces uninteresting or aversive should have deficits in gaze following.* Using small positive or negative rewards for looking at the caregiver leads to gradual deficits in the emergence of gaze following. This problem may be corroborated by a poor development of face processing skills caused by aversiveness (or even neutrality) of faces.
7. *Infants with deficits in attention-shifting should exhibit delays in learning gaze following.* The model shows that slow attention-shifting ($T_{\text{lat}} > 0$) leads to a sluggish emergence of gaze following behavior.
8. *Amount of caregiver contact should influence emergence of gaze following.* An infant who experiences few face-to-face interactions with caregivers may be slower to acquire gaze following because of a shortage of relevant learning experiences.
9. *Differences in caregiver behavior can aid or hinder the emergence of gaze following.* Varying the model parameters related to the caregiver behavior (p_{shift}, T_{min}) while keeping the parameters of the infant identical, leads to differences in learning speed. It is likely that 'optimal' caregiver behavior depends on particular infant parameters. Thus, the optimal caregiver behavior will generally be different for each infant – especially in the case of abnormally developing infants. More work is needed to understand these issues and their potential ramifications for therapeutic interventions (Teuscher & Triesch, 2004).
10. *Lesioning certain neural pathways should impair gaze following behavior.* We assume that information about the caregiver's direction of gaze is extracted from face processing areas including (but not necessarily limited to) the Fusiform Face Area (Kanwisher, McDermott & Chun, 1997). Control of gaze shifts is assumed to be mediated through areas such as the Frontal Eye Fields (Tehovnik, Sommer, Chou, Slocum & Schiller, 2000). Our temporal difference learning model assumes that pathways between these sites (direct or indirect) are modified during learning and lesioning these pathways may impair gaze following.

Discussion

We have proposed a model of the emergence of gaze following in situated infant–caregiver interactions. Our account is an elaboration of ideas that explain the emergence of gaze following as a learning process driven by hedonic principles (Moore & Corkum, 1994). Infants are viewed as *pleasure-driven agents*, who learn to exploit information about their caregiver's head movement and head pose (and, later, eye direction) to find interesting sights in their environment. More specifically, we have proposed a *Basic Set* of structures and mechanisms that allow the infant to succeed in learning in an appropriately structured environment where the caregiver tends to look at things that the infant will find interesting. The proposed Basic Set has a small number of elements but, as our computer simulations demonstrate, it is *sufficient* for gaze following to emerge. In particular, no additional specialized cognitive modules are *necessary* to explain the emergence of gaze following in infant–caregiver interactions. Note that all elements of our proposed Basic Set are established within days of birth (or, for attention-shifting, at around 3 months) in typically developing infants. This does not mean that we think all other mechanisms are unimportant for a comprehensive account of the emergence of gaze following. It merely means that other mechanisms are not required for explaining the basic gaze following phenomenon.

We have used the model to demonstrate how the Basic Set mechanisms are sufficient to allow an infant to learn to associate a particular head pose of the caregiver with a gaze shift to a location outside of the infant's field of view. This specific ability emerges rather late in normal development. Earlier signs of gaze following may be learned in a very similar way, however. The presence of the Basic Set mechanisms in even very young infants makes a learning account of any earlier gaze following competence plausible. For example, in the context of the present model it is easy to see that, say, gaze following to targets inside the infant's field of view may be learned with the same mechanisms – only more easily and faster/earlier. The only Basic Set element for which there is no evidence of its presence within days of birth is the ability to shift gaze away from a central stimulus. Indeed, all demonstrations of very early 'gaze following' have to remove the face stimulus after the gaze shift to facilitate a gaze shift to the periphery. Overall, we find it hard to envision an account of the progressive expansion of gaze following competence in infancy that is not based on a gradual learning process. Again, as stated in the introduction, this view does not at all preclude the presence of evolved rudimentary propensities that contribute to gaze following in specific situations, but it places a clear emphasis on learning, especially for the emergence of more advanced gaze following skills.

It has been noted that infants will follow not only the line of regard of humans, but also that of non-human objects with face-like features, or objects that behave contingently to them (Johnson, Slaughter & Carey, 1998). This suggests that infants' capacity for joint attention is a generalizable skill that is not tightly tied to specific situations with specific caregivers. Rather, it is a robust skill that extends flexibly to various social interactions. Our model readily accounts for these findings, if the additional assumption is made that such non-human objects may be able to activate some of the same head pose and gaze direction sensitive neurons in the infant's face processing areas that are utilized for following the gaze of humans.

Related work

A few related models have recently been proposed in the literature. The idea of using temporal difference learning to model the acquisition of gaze following was first mentioned by Matsuda and Omori (2001). They model a learning situation as used by Corkum and Moore (1998), where an experimenter monitors the infant's behavior and gives visual rewards to the infant when it follows the caregiver's gaze. Their paper lacks details, however, and does not explicitly model how the caregiver's direction of gaze becomes associated with certain gaze shifts. We consider explaining this process to be the central problem of learning gaze following.

A recent model by Nagai, Hosoda, Morita and Asada (2003) has been implemented in a robot. Their model, which was developed concurrently with ours, shares a number of aspects of our model (Fasel *et al.*, 2002; Carlson & Triesch, 2003). In Nagai *et al.*'s model the infant also learns to associate head poses of the caregiver with appropriate gaze shifts based on the success or failure of finding a visually appealing stimulus. To this end, a neural network is trained to map the robot's current gaze direction and an image of the caregiver's face onto the desired gaze shift. Their model, however, does not utilize temporal difference learning, but rather an *ad hoc* learning mechanism. Also, no attempts are made to explain failures of the emergence of gaze following in either developmental disorders or in other species. On the positive side, the authors do not make the simplifying assumption that caregiver head poses have a one-to-one correspondence with regions in space, which we have used here. Nagai *et al.* also attempt to explain the progressive development of gaze following skills as described by Butterworth and Jarrett (1991). However, a closer look at their model reveals that the most sophisticated

so-called *representational stage* cannot be achieved. In contrast, new models of our group correctly capture the sequential emergence of all skill levels described by Butterworth and Jarrett (Lau & Triesch, 2004; Jasso, Triesch & Teuscher, 2005). Interestingly, these models predict that limitations in head pose discrimination ability and/or depth perception ability may be the factors preventing younger infants from learning advanced gaze following skills (Butterworth's geometric and representational stages). Taken together, the current study and our more recent ones point to the possibility that simple perceptual limitations may limit the emergence of advanced gaze following skills. We think it is crucial for the field to carefully study how perceptual skills (head pose discrimination, gaze direction estimation, depth perception) and gaze following skills co-develop in the same individual, in order to test the predicted causal relation between these factors.

Developmental disorders

Our account of the emergence of gaze following offers new perspectives on failures of its emergence in developmental disorders. If a small Basic Set of 'ingredients' is demonstrably sufficient for the emergence of gaze following, in situations where the learning process does not succeed, one or several elements of the Basic Set, or their interaction, has been compromised. Elaborating on this idea, we showed how changes to the model motivated by two different developmental disorders (autism and Williams syndrome) can lead to delays or deficits in learning gaze following. In particular, our model is consistent with the idea that in autism an initial reduction in preference for faces might be at the root of a cascade of problems leading to deficits in gaze following and attention-sharing. Our account is also consistent with evidence of the success of therapeutic interventions where infants are explicitly rewarded for a desired behavior such as following gaze (Whalen & Schreibman, 2003). Finally, the model points to the possibility that various combinations of a few small alterations in the developing infant, none of which may be critical by itself, could conspire to produce severe deficits. This is consistent with the characterization of autism as a *spectrum disorder*.

While our accounts of deficits in gaze following in different developmental disorders may seem simplistic, it nevertheless offers important lessons. Most prominently, the model shows that very different causes can lead to deficits in the emergence of gaze following. These causes include (but are not limited to) parameters related to face perception, learning, habituation and value/reward systems. Given that several completely independent causes can all lead to deficits in gaze following, it appears ill-advised to use deficits in gaze following to *define* a disorder. This is still the case in autism, where deficits in social interaction skills such as gaze following are used to *define* the syndrome. Our hope is that computational modeling efforts like ours will help in understanding complex developmental disorders by helping to better differentiate symptoms and narrow down their primary causes. This, in turn, will suggest promising avenues for treatment and early diagnosis.

Cross-species differences

A good account of the emergence of gaze following should also explain differences in the emergence of gaze following behavior, or the complete absence of it, in other species. Since a simple Basic Set of structures and mechanisms is sufficient for gaze following to emerge, any species with the Basic Set should be able to acquire gaze following to some degree. Deficits or differences in the Basic Set may limit the emergence of gaze following, as seen in our discussion of developmental disorders.

Across vertebrate species some Basic Set elements such as habituation and reward-driven learning are essentially ubiquitous, suggesting that these are likely not the missing factors. This inference demands some caution, however, because the presence of, say, reward-driven learning does not mean that just any contingencies can be learned. Nevertheless, we feel that differences in other Basic Set elements are more relevant.

Regarding *perceptual skills and preferences*, the basic questions are how infants of other species might prefer to look at conspecifics, and how well they might distinguish different head or eye orientations. The first question can be studied with controlled preferential looking paradigms to evaluate visual preferences for looking at conspecifics (or humans) (e.g. Bard, Platzman, Lester & Suomi, 1992). Our model predicts that a (not too big) preference for looking at conspecifics' faces is beneficial (although not strictly necessary) for gaze following to emerge.

In terms of the ability to distinguish different head or eye poses of conspecifics, there is evidence that, for example, many primate species can do so to some extent (Itakura, 2004). Interestingly, eye direction may be particularly easy to discern for humans because of the white sclera (Kobayashi & Kohshima, 1997; Emery, 2000). We assume that gaze direction (orientation of the eyes) is more informative than just head pose, but it is also harder to perceptually discriminate, because the eyes are small. A first attempt to relate such differences to our model is as follows. If an animal with a weaker perceptual system can only inaccurately estimate a conspecific's head position, then this cue will be less predictive of

interesting sights compared to accurate knowledge of the conspecific's direction of gaze. Thus, as explained above, we can attempt to model limited perceptual skills by reducing the predictiveness of the caregiver's gaze p_{vaild}. As our experiments showed, reducing p_{vaild} slows the emergence of gaze following or prevents it altogether. Thus, some species may not learn to follow gaze at all or may only learn primitive forms of gaze following because their perceptual apparatus does not allow them to gather sufficiently accurate information about conspecifics' gaze direction to make gaze following worthwhile. A more detailed analysis of the perceptual requirements for higher gaze following skills specifically implicates depth perception abilities and accuracy of gaze direction estimation as possible culprits (Lau & Triesch, 2004). Generally speaking, we can expect advanced gaze following skills only in those species that have adequate perceptual abilities.

A related issue is foveation. The more foveation there is in an animal's visual system, the more important it is to look directly at the most relevant regions of the environment. Gaze following can help to identify such regions. At the same time, a more foveated vision system will be better at making fine discriminations, say, of a conspecific's direction of gaze, which benefits gaze following. Thus, we suspect that there may be a correlation between the degree of foveation of a species' visual system and its propensity to follow gaze.

Regarding a *structured social environment*, a first condition for the emergence of gaze following is that species must live in social groups. Further, the gaze of conspecifics must be predictive of informative events. Note that gaze can have a number of other meanings in social species that could potentially impact gaze following. For instance, *gaze aversion* is found in several monkey species (Argyle & Cook, 1976). In such species, direct eye contact is a gesture of aggression and it is particularly important for members of such species to be sensitive to direct versus averted gaze, as indicated by head and eye direction (Coss, 1978; Emery, 2000).

Point following

Although we have focused on gaze following in this paper, note that point following may be learned based on the same principles. Pointing with an outstretched and aligned arm, hand and finger is the most natural way to intentionally direct another's attention to a new target, and caregivers and (older) infants do produce pointing gestures to direct each other's attention (Bates, Camaioni & Volterra, 1975; Lempers, 1979; Leung & Rheingold, 1981). To model the emergence of point following, we could simply choose to identify different caregiver head poses in the current model with different pointing gestures performed by the caregiver. However, there are certain differences to consider. First, while the caregiver frequently shifts gaze, pointing gestures during naturalistic exchanges are rare by comparison (Deák *et al.*, 2004). Second, pointing gestures are likely to be more salient for infants because of the large amount of movement involved. Third, infants may be better at discriminating pointing direction than head direction because the extended arm provides a better directional cue (Deák *et al.*, 2000). Fourth, pointing gestures are likely to be more predictive of interesting events, because caregivers will tend to engage in this 'effort' only when a particularly relevant environmental stimulus is present. All but the first of these four points suggest that it might be easier for infants to learn point following. In fact, human infants by 9 months follow gaze much more reliably when it is accompanied by a point (Flom, Deák, Phill & Pick, 2003), and a quasi-naturalistic observational study shows that infants from 5 to 10 months are far more likely to follow a parent's point than a parent's gaze shift (Deák *et al.*, 2004).

Future work

Of course, our model and the ones discussed above must be seen as only first steps towards a full computational account of the emergence of gaze following. In many respects, these models are still overly simplistic. Examples of simplifications in our model are the restriction to a small set of discrete spatial regions, the absence of peripheral vision and the stereotypic, non-interactive behavior of the caregiver model, just to name a few. Recent work has started to address some of these issues (Lau & Triesch, 2004; Teuscher & Triesch, 2004; Jasso & Triesch, 2004). Another limitation is that the model currently does not address how higher attention sharing skills may emerge. Future work needs to demonstrate that models such as the present one can be scaled up to explain the emergence of more advanced attention sharing skills. Despite these shortcomings and limitations, we think our model is a useful step in theorizing about the emergence of gaze following and shared attention in general. In some respects, the simplicity of the model is a strength, since it brings the computational essence of the underlying learning mechanisms into focus.

Appendix

Model equations

We will follow the notation in Sutton and Barto, 1998. Time progresses in discrete steps ($t = 0, 1, 2, \ldots$). At any

time t the when- and where-agents of the model are each in a particular *state* s_t. In the following we will only consider a single agent (*when* or *where*). Upon observing the current state s_t, the agent decides to take an *action* a_t and potentially receives a *reward* r_t as a consequence. The probabilistic mapping from states to actions is the agent's *policy* (denoted π), which is adapted during learning. The goal of the agent is to learn a policy that maximizes the *future discounted reward* R_t defined as:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}, \tag{3}$$

where r_{t+k+1} is the reward received at time $t + k + 1$, and $0 \leq \gamma \leq 1$ is the so-called *discount factor*.

In order to improve its policy, the agent learns a so-called *state-action value function* $Q^{\pi}(s,a)$. These are estimates of the future discounted reward the agent will receive when choosing action a in state s and following the current policy π thereafter. Formally, the unknown state-action values are defined as:

$$Q^{\pi}(s,a) = E_{\pi}[R_t \mid s_t = s, a_t = a], \tag{4}$$

where $E_{\pi}[.]$ denotes the expected value with respect to the current policy $\pi(t)$.

We will denote the estimate of a state-action value at time t as $Q_t(s,a)$. Our agent estimates the $Q_t(s,a)$ with a *temporal difference learning (TD learning)* method, the SARSA algorithm (Sutton & Barto, 1998): On taking an action and receiving a reward, the *temporal difference error* is computed as

$$\delta_t = r_t + \gamma Q_t(s_{t+1},a_{t+1}) - Q_t(s_t,a_t), \tag{5}$$

where $Q_t(s_t,a_t)$ is the state-action value assigned to the state-action pair (s_t,a_t) at time t. The temporal difference is used to adjust the state-action value estimate with the learning step:

$$Q_{t+1}(s_t,a_t) = Q_t(s_t,a_t) + \alpha\delta_t, \tag{6}$$

where $\alpha > 0$ is a learning rate parameter.

The agent balances exploration and exploitation using a *softmax* or *Boltzmann action selection rule*. The probability of choosing action a in state s is given by:

$$p_t(a \mid s) = \frac{e^{\tilde{Q}_t(s,a)/\tau}}{\sum_{a'=1}^{N} e^{\tilde{Q}_t(s,a')/\tau}} \tag{7}$$

where $\tilde{Q}_t(s,a) = Q_t(s,a)/\max_{a'} \mid Q_t(s,a') \mid$ and τ is the so-called *temperature* parameter. Selecting actions based on the normalized $\tilde{Q}_t(s,a)$ instead of the $Q_t(s,a)$ has the advantage that the amount of exploration is stabilized in the presence of changes to other parameters.

In a neural implementation, the estimated Q-values can be thought of as the strength of synaptic connections between units coding for different environmental states (presynaptically) and possible actions (postsynaptically), such that increasing the estimate of a Q-value corresponds to strengthening a connection from the corresponding state to the corresponding action. In the context of gaze following these connections may be along a pathway from face processing areas such as the Fusiform Face Area to gaze control structures such as the Frontal Eye Fields.

Acknowledgements

This work would not have been possible without the support of the UC Davis MIND Institute and the National Alliance for Autism Research. The work described here is part of the MESA project at UC San Diego (Modelling for the Emergence of Shared Attention; http://mesa.ucsd.edu), a larger effort to understand the emergence of shared attention in normal and abnormal development through closely integrating observational studies and systematic experiments with computational modelling approaches. We thank all members of the MESA project for their continuing collaboration: Ian Fasel, Hector Jasso, Boris Lau, Javier Movellan, Leigh Sepeta and Yuri You. We also thank Shoji Itakura, Christine Johnson and Laura Schreibman for comments on earlier drafts.

References

Adamson, L.B. (1995). *Communication development during infancy*. Boulder, CO: Westview.

Adamson, L.B., & Bakeman, R. (1991). The development of shared attention during infancy. *Annals of Child Development*, **8**, 1–41.

Adrien, J.L., Lenoir, P., Martineau, J., Perrot, A., Hameury, L., Larmande, C., & Sauvage, D. (1993). Blind ratings of early symptoms of autism based upon family home movies. *Journal of the American Academy of Child and Adolescent Psychiatry*, **32**, 617–626.

Agnetta, B., Hare, B., & Tomasello, M. (2000). Cues to food locations that domestic dogs (*canis familiaris*) of different ages do and do not use. *Animal Cognition*, **3**, 107–112.

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

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

Bard, K.A., Platzman, K.A., Lester, B.M., & Suomi, S.J. (1992). Orientation to social and nonsocial stimuli in neonatal

chimpanzees and humans. *Infant Behavior and Development*, **15**, 43–56.
Baron-Cohen, S. (1995). *Mindblindness: an essay on autism and theory of mind*. Cambridge, MA: A Bradford Book/The MIT Press.
Bates, E. (1979). On the evolution and development of symbols. In E. Bates (Ed.), *The emergence of symbols: Cognition and communication in infancy* (pp. 1–32). New York: Academic Press.
Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly*, **21**, 205–226.
Baumann, M., & Kemper, T. (2005). Neuroanatomic observations of the brain in autism: a review and future directions. *International Journal of Developmental Neuroscience*, **23** (2–3), 183–187.
Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St George, M. (2000). The neurocognitive profile of Williams syndrome: a complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*, **12**, 7–29.
Bertrand, J., Mervis, C., Rice, C.E., & Adamson, L. (1993). Development of joint attention by a toddler with Williams syndrome. Paper presented at the Gatlinberg Conference on Research and Theory in Mental Retardation and Developmental Disabilities, Gatlinberg.
Brazelton, T.B., Koslowski, B., & Main, M. (1974). The origins of reciprocity: the early mother–infant interaction. In M. Lewis & L. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 49–76). New York: John Wiley.
Bruner, J. (1983). *Child's talk: Learning to use language*. 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.
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.E., & 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. (1991). What minds have in common in space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, **9**, 55–72.
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* (pp. 105–114). London: World Scientific.
Caron, A.J., Butler, S.C., & Brooks, R. (2002). Gaze following at 12 and 14 months: do the eyes matter? *British Journal of Developmental Psychology*, **20**, 225–239.
Casey, B., Gordon, C., Manheim, G., & Rumsey, J. (1993). Dysfunctional attention in autistic savants. *Journal of Clinical and Experimental Neuropsychology*, **15** (6), 933–946.
Chawarska, K., Klin, A., & Volkmar, F. (2003). Automatic attention cueing through eye movement in 2-year-old children with autism. *Child Development*, **74**, 1108–1122.
Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
Cole, M., & Cole, S. (1996). *The development of children* (3rd edn.). New York: Freeman.
Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P.J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 61–83). Hillsdale, NJ: Erlbaum.
Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, **34** (1), 28–38.
Coss, R.G. (1978). Perceptual determinants of gaze aversion by the lesser mouse lemur (*microcerbus murinus*): the role of two facing eyes. *Behaviour*, **64**, 248–267.
Csibra, G. (2006). Blind infants in random environments: further predictions. *Developmental Science*, **9** (2), 148–149.
Dawson, G., Meltzoff, A.N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, **28**, 479–485.
Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., & Estes, A. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental Psychology*, **40** (2), 271–283.
Deák, G.O., Flom, R., & Pick, A.D. (2000). Perceptual and motivational factors affecting joint visual attention in 12- and 18-month-olds. *Developmental Psychology*, **36**, 511–523.
Deák, G.O., & Triesch, J. (in press). The emergence of attention-sharing skills in human infants. In K. Fujita & S. Itakura (Eds.), *Diversity of cognition*. University of Kyoto Press.
Deák, G.O., Wakabayashi, Y., Sepeta, L., & Triesch, J. (2004). Development of attention-sharing from 5 to 10 months of age in naturalistic interactions. Paper presented at the International Conference on Infancy Studies, Chicago, IL.
DeCasper, A.J., & Fifer, W.P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, **208**, 1174–1176.
D'Entremont, B., Hains, S., & Muir, D. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*, **20** (4), 569–572.
Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness*. Cambridge, MA: A Bradford Book/The MIT Press.
Emery, N.J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, **24**, 581–604.
Emery, N.J., Lorincz, E.N., Perrett, D.I., Oram, M.W., & Baker, C.I. (1997). Gaze following and joint attention in rhesus monkeys (*macaca mulatta*). *Journal of Comparative Psychology*, **111** (3), 286–293.
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., Massaccesi, S., Pividori, D., & Johnson, M.H. (2004). Gaze following in newborns. *Infancy*, **5** (1), 39–60.
Fasel, I., Deák, G.O., Triesch, J., & Movellan, J. (2002). Combining embodied models and empirical research for

understanding the development of shared attention. In A. Jacobs (Ed.), *International Conference on Development and Learning* (pp. 21–27). Los Alamitos, CA: IEEE Computer Society.

Findlay, J., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, **22**, 661–674.

Floccia, C. (1997). High-amplitude sucking and newborns: the quest for underlying mechanisms. *Journal of Experimental Child Psychology*, **64**, 175–198.

Flom, R., Deák, G., Phill, C.G., & 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.

Gergely, G., & Watson, J.S. (1999). Early social-emotional development: contingency perception and the social biofeedback model. In P. Rochat (Ed.), *Early social cognition* (pp. 101–137). Hillsdale, NJ: Erlbaum.

Goldberg, M., Lasker, A., Zee, D., Garth, E., Tien, A., & Landa, R. (2002). Deficits in the initiation of eye movements in the absence of a visual target in adolescents with high functioning autism. *Neuropsychologica*, **40** (12), 2039–2049.

Haith, M.M., Bergman, T., & Moore, M.J. (1979). Eye contact and face scanning in early infancy. *Science*, **198**, 853–855.

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., Brown, M., Williamson, C., & Tomasello, M. (2002). The domestication of social cognition in dogs. *Science*, **298** (5598), 1634–1636.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behavior*, **59**, 771–786.

Hare, B., & Tomasello, M. (1999). Domestic dogs (*canis familiaris*) use human and conspecific social cues to locate hidden food. *Journal of Comparative Psychology*, **113**, 173–177.

Harris, N., Courchesne, E., Townsend, J., Carper, R., & Lord, C. (1999). Neuroanatomic contributions to slowed orienting of attention in children with autism. *Cognitive Brain Research*, **8** (1), 61–71.

Hood, B., Willen, J., & Driver, J. (1998). Adults' eyes trigger shifts of visual attention in human infants. *Psychological Science*, **9** (2), 131–134.

Howard, M.A., Cowell, P.E., Boucher, J., Broks, P., Mayes, A., Farrant, A., & Roberts, N. (2000). Convergent neuroanatomical and behavioral evidence of an amygdala hypothesis of autism. *Neuroreport*, **2** (13), 2931–2935.

Hutt, C., & Ounsted, C. (1966). The biological significance of gaze aversion with particular reference to the syndrome of infantile autism. *Behavioral Science*, **11** (5), 346–356.

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

Itakura, S. (2004). Gaze following and joint visual attention in nonhuman animals. *Japanese Psychological Research*, **46** (3), 216–226.

Jasso, H., & Triesch, J. (2004). A virtual reality platform for modeling cognitive development. In J. Triesch & T. Jebara (Eds.), *Proceedings of the ICDL'04 – Third International Conference on Development and Learning*, San Diego, CA (pp. 229–236). The Salk Institute for Biological Studies.

Jasso, H., Triesch, J., & Teuscher, C. (2005). A reinforcement learning model explains the stage-wise development of gaze following. Proceedings of the 12th Joint Symposium on Neural Computation (JSNC 2005), Los Angeles, CA, 14 May.

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

Johnson, S., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze following in 12-month-olds. *Developmental Science*, **1** (2), 223–238.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, **17** (11), 4302–4311.

Karmiloff-Smith, A., Thomas, M., Annaz, D., Humphreys, K., Ewing, S., Brace, N., Van Duuren, M., Pike, G., Grice, S., & Campbell, R. (2004). Exploring the Williams syndrome face-processing debate: the importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry*, **45** (7), 1258–1274.

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

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society London, B*, **358**, 345–360.

Kobayashi, H., & Kohshima, S. (1997). Morphological uniqueness of human eyes and its adaptive meaning. *Nature*, **387**, 767–768.

Laing, E., Butterworth, G., Ansari, D., Gsödl, M., Longhi, E., Panagiotaki, G., Paterson, S., & Karmiloff-Smith, A. (2002). Atypical development of language and social communication in toddlers with Williams syndrome. *Developmental Science*, **5** (2), 233–246.

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.

Landry, R., & Bryson, S. (2004). Impaired disengagement of attention in young children with autism. *Journal of Child Psychology and Psychiatry*, **45** (6), 1115–1122.

Langdell, T. (1978). Recognition of faces: an approach to the study of autism. *Journal of Child Psychology and Psychiatry*, **19** (3), 255–268.

Lau, B., & Triesch, J. (2004). Learning gaze following in space: a computational model. In J. Triesch & T. Jebara (Eds.), *Proceedings of the ICDL'04 – Third International Conference on Development and Learning*, San Diego, CA (pp. 57–64). The Salk Institute for Biological Studies.

Leekam, S., Baron-Cohen, S., Perret, D., Milders, M., & Brown, S. (1997). Eye-direction detection: a dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, **15**, 77–95.

Leekam, S., Hunnisett, E., & Moore, C. (1998). Targets and cues: gaze-following in children with autism. *Journal of Child Psychology and Psychiatry*, **39** (7), 951–962.

Leekam, S., López, B., & Moore, C. (2000). Attention and joint attention in preschool children with autism. *Developmental Psychology*, **36** (2), 261–273.

Lempers, J.D. (1979). Young children's production and comprehension of nonverbal deictic behaviors. *The Journal of Genetic Psychology*, **135**, 93–102.

Leslie, A.M. (1987). Pretense and representation – the origins of theory of mind. *Psychological Review*, **94** (4), 412–426.

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

Maestro, S., Muratori, F., Cavallaro, M., Pei, F., Stern, D., Golse, B., & Palacio-Espasa, F. (2002). Attentional skills during the first 6 months of age in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 1239–1245.

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

Matsuda, G., & Omori, T. (2001). Learning of joint visual attention by reinforcement learning. In E.M. Altmann & A. Cleeremans (Eds.), *Proceedings of the fourth international conference on cognitive modeling* (pp. 157–162). Mahwah, NJ: Lawrence Erlbaum Associates.

Mervis, C.B., Morris, C.A., Klein-Tasman, B.P., Bertrand, J., Kwitny, S., Appelbaum, L., & Rice, C.E. (2003). Attentional characteristics of infants and toddlers with Williams syndrome during triadic interactions. *Developmental Neurpsychology*, **23** (1–2), 243–268.

Mobbs, D., Garrett, A., Menon, V., Rose, F., Bellugi, U., & Reiss, A. (2004). Anomalous brain activation during face and gaze processing in Williams syndrome. *Neurology*, **62** (11), 2070–2076.

Montague, P., Hyman, S., & Cohen, J. (2004). Computational roles for dopamine in behavioural control. *Nature*, **431**, 760–767.

Moore, C. (2006) Modeling the development of gaze following needs attention to space. *Developmental Science*, **9** (2), 149–150.

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., & Corkum, V. (1994). Social understanding at the end of the first year of life. *Developmental Review*, **14**, 349–372.

Moore, C., & Dunham, P.J. (Eds.) (1995). *Joint attention: Its origins and role in development*. Hillsdale, NJ: Erlbaum.

Morales, M., Mundy, P., & Rojas, J. (1998). Gaze following and language development in six-month-olds. *Infant Behavior and Development*, **36**, 325–338.

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

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

O'Reilly, R.C., & Munakata, Y. (2002). *Computational explorations in cognitive neuroscience*. Cambridge, MA: A Bradford Book.

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: a study of first birthday home video tapes. *Journal of Autism and Developmental Disorders*, **24**, 247–257.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenent, M. (1995). Mother's face recognition by neonates: a replication and an extension. *Infant Behavior and Development*, **18**, 79–85.

Richardson, F.M., & Thomas, M.S.C. (2006). The benefits of computational modelling for the study of developmental disorders: extending the Triesch *et al.* model to ADHD. *Developmental Science*, **9** (2), 151–155.

Richer, J.M., & Coss, R.G. (1976). Gaze aversion in autistic and normal children. *Acta Psychiatrica Scandinavica*, **53**, 193–210.

Sai, F., & Bushnell, W.R. (1998). 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.

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, **12**, 13–24.

Schlesinger, M., & Parisi, D. (2001). The agent-based approach: a new direction for computational models of development. *Developmental Review*, **21**, 121–146.

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

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences*, **6** (7), 293–298.

Sutton, R.S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, **3**, 9–44.

Sutton, R.S., & Barto, A.G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: A Bradford Book/The MIT Press.

Tamis-LeMonda, C., & Bornstein, M. (1989). Habituation and maternal encouragement of attention in infancy as predictors of infant language, play, and representational competence. *Child Development*, **60**, 738–751.

Tantam, D., Holmes, D., & Cordess, C. (1993). Nonverbal expression in autism of Asperger type. *Journal of Autism and Developmental Disorders*, **23**, 111–133.

Tehovnik, E.J., Sommer, M.A., Chou, I.-H., Slocum, W.M., & Schiller, P.H. (2000). Eye fields in the frontal lobes of primates. *Brain Research Reviews*, **32**, 413–448.

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 ICDL'04 – Third International Conference on Development and Learning*, San Diego, CA (pp. 9–16). The Salk Institute for Biological Studies.

Thelen, E., Schöner, G., Scheier, C., & Smith, L.B. (2000). The dynamics of embodiment: a field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, **24** (1), 1–86.

Thelen, E., & Smith, L.B. (1994). *A dynamics systems approach to the development of cognition and action*. Cambridge, MA: A Bradford Book/The MIT Press.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P.J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.

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

Tomasello, M., Call, J., & Hare, B. (1997). Five primate species follow the visual gaze of conspecifics. *Animal Behavior*, **55**, 1063–1069.

Tomasello, M., Hare, B., & Agnetta, B. (1999). Chimpanzees, *pan troglodytes* follow gaze geometrically. *Animal Behavior*, **58**, 769–777.

Trepagnier, C. (1996). A possible origin for the social and communicative deficits of autism. *Focus on Autism and Other Developmental Disabilities*, **11**, 170–182.

van der Geest, J.N., Lagers-van Haselen, G.C., van Hagen, J.M., Govaerts, L.C.P., de Coo, I.F.M., de Zeeuw, C.I., & Frens, M.A. (2004). Saccade dysmetria in Williams-Beuren syndrome. *Neuropsychologica*, **42**, 569–576.

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

Watson, J.S., & Ramey, C.T. (1985). Reactions to response-contingent stimulation in early infancy. In J. Oates (Ed.), *Cognitive development in infancy* (pp. 219–227). Hillsdale, NJ: Erlbaum.

Whalen, C., & Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry*, **44** (3), 456–468.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, **9** (4), 625–636.

Woodward, A.L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, **6** (3), 297–311.