

Deák, G.O., Triesch, J., Krasno, A., de Barbaro, K., & Robledo, M. (2013). Learning to share: The emergence of joint attention in human infancy. In B. Kar (Ed.), *Cognition and Brain Development: Converging Evidence from Various Methodologies* (pp. 173-210). Washington, DC: American Psychological Association.

# 9

## LEARNING TO SHARE: THE EMERGENCE OF JOINT ATTENTION IN HUMAN INFANCY

GEDEON O. DEÁK, JOCHEN TRIESCH, ANNA KRASNO,  
KAYA DE BARBARO, AND MARYBEL ROBLEDO

The social sciences are struggling to understand the dynamics of social groups as complex systems (Strogatz, 2001). How do individuals adapt their behaviors in the presence of others? How do they learn patterns of social information? Answering such questions requires theoretical frameworks and predictive models that are closely fitted to empirical behavioral, ethnographic, and physiological data. A major topic is how humans develop social behaviors and social knowledge during infancy. There is evidence that the foundations of social behaviors are laid in infant-caregiver interactions (Sroufe, 1996). What remains to be established is how the structure of infants' social and physical environment interacts with changes in their neural, sensorimotor, and body structures to yield new social knowledge and behavior.

---

Support for the research reported in this chapter was provided by the National Science Foundation (Contracts SES-0527756 and BCS-0827040); the M.I.N.D. Institute at the University of California, Davis; and the National Alliance for Autism Research. The authors thank numerous colleagues and former students, including Andrea Chiba, Jordan Danly, Kaya de Barbaro, Ian Fasel, Joanne Jao, Hector Jasso, Anna Krasno, Josh Lewis, Yu Liao, Javier Movellan, Cindy Nam, Jackie Overton, Yuri You, and Corrine Zavala.

DOI: 10.1037/14043-010

*Cognition and Brain Development: Converging Evidence From Various Methodologies*, B. R. Kar (Editor)  
Copyright © 2013 by the American Psychological Association. All rights reserved.

The infant-caregiver dyadic system has intricate dynamic properties. Infants' brains and bodies rapidly develop as they acquire vast, varied experience in the family socioecosystem. Parents are themselves complex systems that select from a large behavioral repertoire while interacting with and in the presence of infants. New infant social skills emerge within this complex metasystem. These emerging skills are, despite the high-dimensional complexity of the metasystem, stochastically predictable within some broad parameters of variance. These parameters have been partly outlined by vigorous research efforts spanning the past 4 decades.

In spite of these efforts, we still lack a viable framework that can integrate existing descriptive findings. Such a framework should address questions like the following: How do parents' behaviors contribute to specific infants' social skills? What learning processes does the infant's developing brain bring to bear? How do affect and arousal systems modulate the expression of social behaviors? There is, to be sure, a lot of descriptive evidence that can be brought to bear and a growing if vague acknowledgment that explanations must be framed in terms of dynamic physical systems. What we lack is a powerful theoretical model or, better yet, alternative models that are biologically plausible, ecologically plausible, and capable of generating specific predictions. However, new research incorporating more rigorous, high-dimensional behavioral and physiological experiments, microethnographic studies, insights from computational and basic neuroscience, and computational simulations have begun to yield plausible models of how infant social skills develop.

In this chapter, we describe one such effort: the MESA (Modeling the Emergence of Shared Attention) Project, an interdisciplinary collaboration begun at the University of California, San Diego, by Jochen Triesch, Javier Movellan, and Gedeon Deák. The guiding framework of the project is a theory of the development of attention sharing and other social behaviors. The PLeASES theory starts with the assumption that complex behaviors like gaze following could emerge from the complex interplay of infants' early phenotypes—Perceptual routines, Learning mechanisms, and Affective traits—and their environment, or Social Ecology Structures.

In the following pages, we summarize some MESA research that was done to test and refine the PLeASES theory. First, we summarize descriptive findings on the development of social attention in the first 2 years. This selective review focuses on phenomena that are relatively challenging to integrate within a plausible theory. Next, we outline our approach to formulating the theory. This approach has general implications: It can serve as a template for formulating a viable theory of any developing system of behavior. Next, we explain some of the main assumptions, claims, and predictions of the PLeASES theory. We then describe several empirical efforts to test and refine the PLeASES theory. The first is a naturalistic videoethnographic study that

shows how infants can learn to follow gaze as an incidental by-product of their own reward systems and attention shifting and of their parents' visual and manual actions. The second is a series of computational simulations showing that even a very simple adaptive agent with biologically inspired learning mechanisms and input that replicates real parents' actions can acquire many of the attention-shifting behaviors that we identified as challenging to integrate. The last is a series of results from experiments on infants' looking patterns and affective responses to social and nonsocial stimuli. All three strands of research speak to claims of the PLeASES theory and suggest aspects of the theory that can be refined or expanded.

#### INFANT ATTENTION SHARING: WHAT PHENOMENA SHOULD A THEORY EXPLAIN?

*Attention sharing* is deliberately shifting attention to the focus of another individual's attention because of seeing the individual seemingly attending to that focal stimulus. Social activities—like people watching at a café, visiting a museum or zoo, playing cards, or discussing a blueprint—all require shared attention. Yet, attention sharing is no mere social lubricant. It facilitates entire categories of interactions that are critical to humans, notably teaching and learning. Monitoring others' attention can provide useful information about uncertain, novel, and even dangerous environments. It is a key element for learning difficult or complex procedures.

Attention sharing is a critical skill for infants and children (Bakeman & Adamson, 1984). It is part of the behavioral system by which infants forge socioemotional bonds (Stern, 2000). It helps infants learn what is important in a complex environment and will eventually help children infer what information they do or do not share with another person (O'Neill, 1996). Attention sharing also facilitates language learning (Baldwin, 1993; Tomasello, 1999). By early childhood, attention-sharing skills are presumed in all educational settings (Rogoff, 1990).

Yet, attention-sharing skills vary widely across infants and even children. At one extreme, attention-sharing deficits are predictors of social and language deficits in autism (Mundy, Sigman, & Kasari, 1990). A plausible theory, then, should account for both typical variability and atypical developmental paths. Within this developmental path, gaze following and point following are typically the first behaviors to emerge during infancy. *Gaze following* is shifting visual attention to match another person's gaze target as a result of encoding and reacting to that person's looking behaviors. *Point following* is shifting attention to the distal target of another person's outstretched arm and (typically) finger or fingers.

Gaze following, point following, and other attention-sharing skills emerge in a semipredictable sequence from 3 to 24 months of age (Butterworth & Cochran, 1980; Butterworth & Itakura, 2000; Butterworth & Jarrett, 1991; Deák, Flom, & Pick, 2000; Flom, Deák, Phill, & Pick, 2004). At 4 or 5 months of age, infants do not respond to parents' looking or pointing bids to redirect attention (Robledo, Danly, Acuña, Ramundo, & Deák, 2009); some 9-month-olds occasionally, in stripped-down laboratory settings, follow gaze to targets already in their visual fields (Flom et al., 2004). A few 6-month-olds rarely show this response, again, in impoverished laboratory settings (Butterworth & Jarrett, 1991; Morales, Mundy, & Rojas, 1998). When 6- to 9-month-olds do turn in the direction of the adult's gaze, they tend to focus on the first thing they see, even if it is not the adult's focus of attention (e.g., Butterworth & Cochran, 1980). This "premature capture" declines from 9 to 12 months (Deák et al., 2000).

There have been claims that infants follow gaze by 3 months or younger (Hood, Willen, & Driver, 1998). However, those studies show limited, weak effects that are attributable to directional motion cueing (Farroni, Johnson, Brockbank, & Simion, 2000; Farroni, Massaccesi, Pividori, & Johnson, 2004). Recently, Robledo et al. (2009) followed infants from 4 to 12 months, testing gaze and point following monthly. In a critical test, infants learned that six monitors spaced around a room would sometimes play reinforcing videos. Infants received more immediate video rewards if they followed the adult's cue (gaze, point, or both) to the specified target. The conditional reinforcement design rules out the possibility that young infants do not follow gaze or point because they are simply unmotivated (see Deák et al., 2000). Motivational factors are a confound in virtually all previous experimental studies of infant attention sharing. Preliminary results are shown in Figure 9.1: Even 6-month-old infants did not follow gaze, even to front targets. Not until 9 to 10 months did some infants reliably follow gaze to targets in the periphery. Even at 12 months, infants rarely followed gaze to targets behind them. This is strong evidence that gaze following per se (i.e., not just motion cueing) emerges around 9 to 10 months.

The development of point following is in some ways similar, emerging around 9 to 12 months. From this period forward, infants are more likely to follow points than gaze shifts (Deák et al., 2000; Deák, Walden, Yale, & Lewis, 2008; Desrochers, Morissette, & Ricard, 1995). Point following is affected by some of the same variables as gaze following: For example, infants are more likely to follow points to targets in front of them than behind them (Deák et al., 2000; Flom et al., 2004). This finding suggests that some common factors underlie these behaviors—there might not be, for example, a narrowly specialized system that mediates gaze following behaviors.

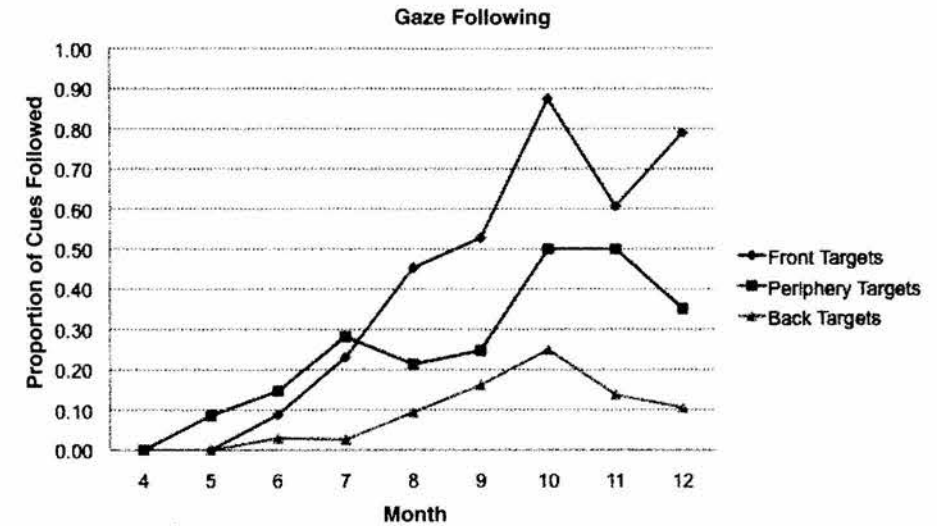


Figure 9.1. Proportion of gaze following responses to an 8-s cue (i.e., an adult turning head and eyes to one of six video targets), from the MESA longitudinal study (preliminary results;  $n = 32$ ).

Figure 9.2 shows some major developmental changes in infancy that we postulated might relate to the emergence of attention-sharing skills in the first 16 months. These changes include learning, perceptual, and traits; we return to these below in our description of the PLeASES theory. However, what remains unanswered by this list of traits, or by the phenomena described above, is why infants eventually follow gaze and pointing. These skills follow months of social experience, maturation, and learning, but we do not know how the experiences of those weeks and months cause change. A possible partial explanation is that parents' gaze and pointing actions help infants predict the location of future rewards—that is, the actions serve as basic reward cues (Schultz, Dayan, & Montague, 1997). If this was so, and if gaze and point perfectly predicted high-value stimulus locations, infants might quickly learn to use them. But adults' gaze shifts and pointing gestures are not perfect predictors of reward locations: For example, adults occasionally look at things that are boring for infants (e.g., rectangles of paper, small handheld blocks). They also roll their eyes, stare at the wall, blink, look at the infant, and otherwise produce uninformative or confusing fixations. Similarly, parents sometimes outstretch their arm or extend their finger to point out things to other adults, or gesticulate, or stretch their arms. All such behaviors are confusing "noise" for the infant who is learning which gestures are informative. Moreover, even if infants reliably attained a desirable outcome whenever they looked in the direction of adults' cues, the infant's



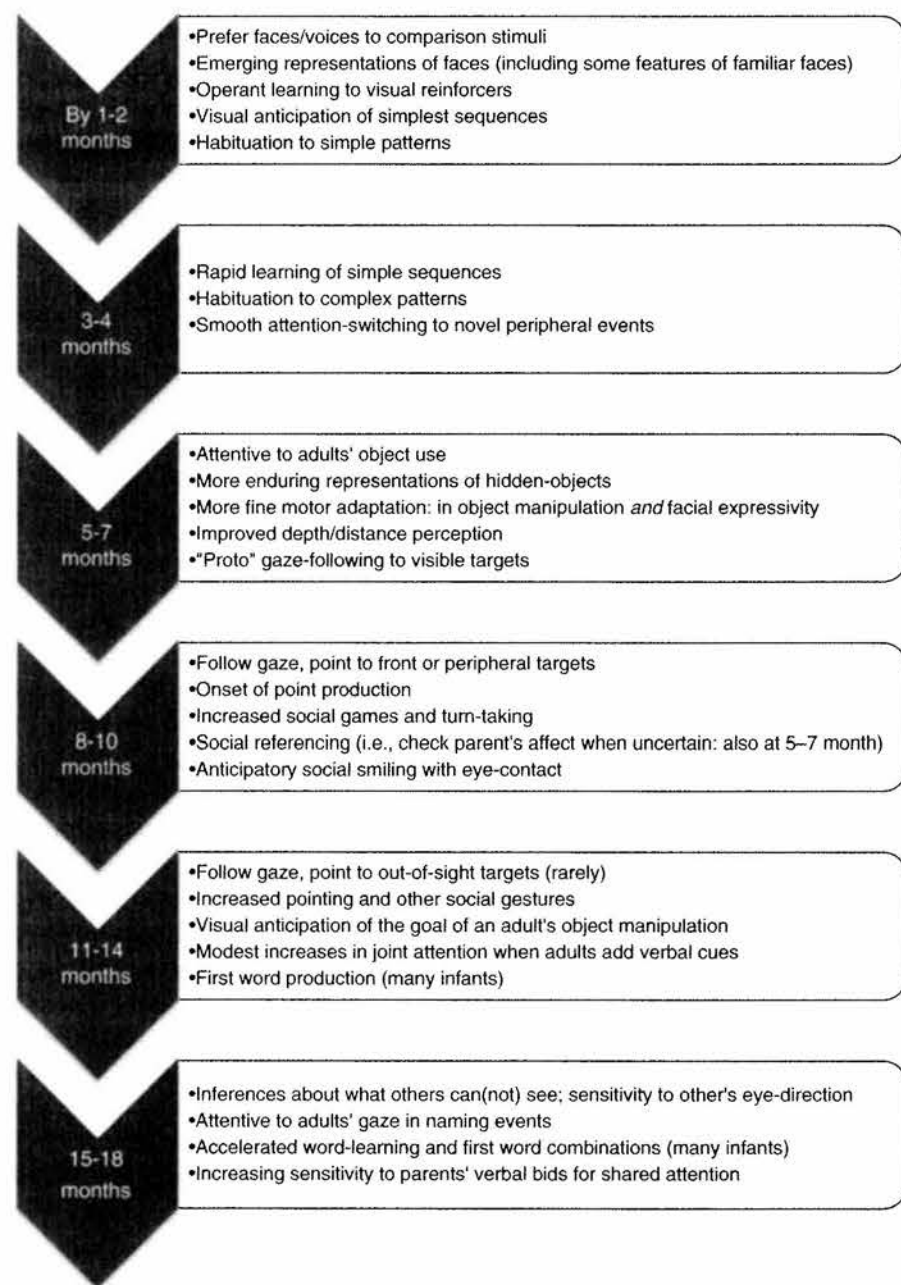


Figure 9.2. Schematic of some key developmental changes in, and related to, the development of gaze and point following skills in infants during the first 18 months of life. Some of the traits are key elements of the PLeASES theory.

learning task would not be trivial. First, to use adults' attention-specifying actions as predictive cues, infants must parse relevant events from a stream of highly variable social actions. Yet adults are often in continuous motion when interacting with infants (Deák, Wakabayashi, Jasso, & Triesch, 2004), and there is no guarantee that infants will notice, much less segment, the adults' critical head turn or eye or arm movement as a unit of significance. Even if they do, the infants must still induce what aspect of the action (e.g., final vector of the index finger) predicts the rewarding outcome. This is not inevitable. Also, the adult's actions are executed in variable contexts, and infants cannot be expected to know which contextual factors matter and to what extent. For example, the direction of light falling on the adult's face generally does not matter. However, the adult's head angle must be calculated relative to the infant's own heading, the triangulated distance and angle of various possible target objects, and the configuration of the shared space—for example, the location of opaque barriers between the adult and various objects. The infant must learn which of these things matter and which do not. The variability of all of these factors, across all environments where the infant interacts with adults, complicates the learning problem. In sum, any theory must consider the high dimensionality and continuousness of adults' actions, the nonobvious association between social actions and outcomes, and the many differences across shared environments.

Despite the challenges posed by these considerations, Deák and Triesch (2006) and Triesch, Teuscher, Deák, and Carlson (2006) proposed that reinforcement learning is an important element of infants' acquisition of attention-following behaviors. Naturally, although instrumental functions are critical aspects of social learning, they are the only critical elements (as we explain later). Also, learning attention-sharing policies is not a discrete goal for the infant; it is a category of states that sometimes occur in the context of various other motivated actions and motivating events. During social interactions, for example, infants babble, attempt to prolong games, point to things, seek proximity with parents, or try to separate and explore. They watch people use tools, watch the family pet, watch TV, and watch other children playing games. The reward function of attention sharing is dynamically nested within an extensive, ever-growing, and dynamically changing hierarchy of costly and rewarding responses to an often-unpredictable environment. At any moment, the status of this cost-reward hierarchy is unknown. Moreover, what infants learn from any given instance of attention sharing is unknown. Finally, caregivers simultaneously have their own range of predicted costs and benefits and related motivations. Parents work to bond with their infants, teach them, elicit smiles and chuckles, show off their infant to peers, or quickly finish a feeding or diaper change and get back to work (or sleep). In sum, attention-sharing episodes emerge within a fluid panorama of



concurrent motivational states, goals (convergent and conflicting), affective changes, and learned expectancies and responses. All of these states, changes, and so forth will have been shaped in a rich history of social experiences. Our challenge as researchers is to infer how these factors interact to generate, extend, and terminate episodes of shared attention. Indeed it is, *mutatis mutandi*, the challenge of understanding any emergent category of social behavior in infants.

#### IF IT PLEASES: AN APPROACH TO FORMULATING THEORIES OF EMERGENT BEHAVIORAL SYSTEMS

What experiences and changes in the first 6 to 12 months after birth lead to gaze following and related skills? To answer this, we should consider neural, behavioral, and ecological factors that might contribute to social behaviors. In addition, the answer should include a model of how those factors cause change. When the MESA Project began in 2001, there were no alternative theories and few, if any, examples of developmental theories that synthesized detailed behavioral, biological, and ecological factors in an explanatory model. For that reason, some of our early discussions focused on the basic question of how to generate a developmental theory (see Fasel, Deák, Triesch, & Movellan, 2002). Thus, although the strategy outlined here came from formulating a theory of attention-sharing development, the approach can be used to generate a theory of any class or system of social behavior. In broad sketches, the approach is as follows:

- Specify an age range when the behavioral system does not yet occur. Ideally, this is far enough back to allow enough time for input and growth processes but not so early that the explanatory problem is intractable (e.g., do not start at the blastula stage).
- Do a task analysis or “reverse engineering” of the behavioral system at some later, more mature state. Ideally, the behavioral phenotypes at this period are well documented and robust across population samples and context (however, that is not always known).
- Hypothesize a minimal set of established precursor phenotypes that would be necessary to acquire that phenotype. The set must be constrained by biological facts not only about age-specific brain physiology and anatomy but also about age-specific peripheral physiology (i.e., sympathetic and parasympathetic systems), sensory development, and body and motor maturation. Precursor phenotypes should be documented and observable (e.g.,

contrast acuity developmental curves), not speculative (e.g., shared attention module).

- Propose a process by which neural plasticity/learning mechanisms and biological maturation processes cause the more mature phenotype to emerge from the minimal set. All of the elements in this explanatory model should be grounded in biological evidence, not purely hypothetical.
- Document pertinent events, objects, and human actions in the infant’s environment. This often requires exploratory ethnographic documentation, especially of patterns of events that tend to occur soon before, during, and after the behaviors of interest. From this and other evidence, hypothesize an ecological model of the information/experience patterns that are relevant to the emergent behavior.

The initial model will therefore specify the processes by which some minimal set of observable phenotypes, embodied in an organism that experiences a history of events-in-settings, will develop new phenotypes. The model can then be tested in the following ways:

- *Formal tests of the sufficiency of the precursor set, learning model, and ecological model.* Deák, Bartlett, and Jebara (2007) referred to the process of modeling both the agent’s encoding and learning processes and the information in the environment as a *dual modeling problem*. That is, if we construct an artificial agent with the proposed precursor set of phenotypes and an environment with naturalistic patterns of available experiences, and we simulate the proposed learning processes, we can observe whether the agent develops new responses that resemble the emerging behaviors of human infants. If it does, we can claim that the model was not falsified. If it does not develop the expected behaviors, then the theory is falsified—although we do not know whether the problem is in our model of the precursor set, in the critical ecological information structures, in the learning process, or in more than one of these elements. Note that simulation experiments are only proofs of plausibility, and the strength of any simulation outcomes rests on the number and range of natural phenomena—especially odd or noninevitable phenomena—that are replicated (Simmering, Spencer, Deák, & Triesch, 2010). For example, a model derives greater plausibility if it predicts how changing a particular parameter will evoke different patterns of disordered or disabled behavioral development (Richardson & Thomas, 2006; Triesch et al., 2006). Also, if

natural development shows discontinuities rather than constant, gradual improvement (which can be predicted by many models), then simulation tests become more informative: The discontinuities are critical tests that will not emerge inevitably from “any old” learning model. In general, then, the more phenomena (i.e., infant behaviors) simulated without “hand wiring” (i.e., overspecifying the agent), the more support (i.e., plausibility) is accorded the model.

- *Experimental longitudinal studies.* A powerful model should support predictions about how variability in precursor phenotypes and in ecological patterns will influence the emergence of new behaviors. For example, if a proposed precursor trait (e.g., speed of habituation) is truly important in the processing model, then individual differences in that trait should relate to later individual differences in the emergent behavior (e.g., its age of onset, efficiency, or benefit).
- *Naturalistic microbehavioral ethnographies and quantitative measures of infants’ environments.* This type of study can specify the event and information structures that permit and promote new behavioral phenotypes. Such studies are seldom available, however, and require laborious efforts. Fortunately, technological advances are allowing researchers to more easily collect, code, and analyze naturalistic data sets that are larger, richer, and more objective than ever before (see Spink et al., 2010). Without these studies, though, any theory or model is predicated on guesses or assumptions about what information might be available to infants and what information they naturally notice.

For related discussions, see Cangelosi et al. (2010), Deák et al. (2007), Goldstein et al. (2010), Grossberg and Vladusich (2010), and Roy et al. (2006). Examples of empirical work that tests biologically and ecologically viable theories include Messinger, Ruvolo, Ekas, and Fogel (2010), Yu and Ballard (2007), and Cameron et al. (2005). In the remaining pages of this chapter, we describe the PLeASES theory of infant attention-sharing and our efforts to test the theory.

#### PLeASES: A THEORY OF THE EMERGENCE OF ATTENTION-SHARING SKILLS

The PLeASES theory rests on the idea that infants’ attention-shifting decisions can become influenced by specific adult actions if those actions serve as predictive cues to the locations of relatively interesting things. It

assumes that infants have no prior expectations that other people’s eye, head, or arm actions are correlated with structures in the environment, or with their own behaviors, or with any internal state (e.g., intention) of the actor. PLeASES is therefore a nonnativist, instrumentalist, mechanistic theory (Deák, Fasel, & Movellan, 2001; Deák & Triesch, 2006; Fasel et al., 2002; Krasno, Deák, Triesch, & Jasso, 2007; Teuscher & Triesch, 2007; Triesch et al., 2006; Triesch, Jasso, & Deák, 2007). It is also biologically grounded: All processes of infant cue perception, attention, action selection, reward calculation, and so forth are to be specified, at least roughly, by neurobiological evidence. Finally, it is ecologically grounded because infants cannot learn skills like gaze following unless the adult cue actions systematically correlate with locations of stimuli that infants find relatively rewarding and unless the timing, form, and frequency of those cues make them detectable and learnable to infants.

The main postulate of PLeASES is that infants learn to follow adults’ gaze, pointing, or other actions because those cues can predict the locations of relatively interesting sights (C. Moore, 1996) in infants’ everyday environments. A second postulate is that infants’ interest is modulated by habituation and arousal. An ancillary assumption is that infants’ relative levels of interest in various stimuli are correlated with adults’ relative interest in the same stimuli. Thus, whatever grabs adults’ attention has a fair shot at getting infants’ attention. A third postulate is that the timing constraints on infants’ looking decisions (i.e., fixating and shifting) are close enough to adults’ looking and acting timing parameters that infants can follow parents’ action cues fast enough to yield some episodes of shared attention.

The PLeASES theory emerged from discussions among researchers across several disciplines, taking into consideration modern learning theory and neuroscience; research on infants’ perception, action, and physiology; and ethnographies of infant–parent social patterns in natural settings. As noted earlier, PLeASES is an acronym for *Perceptual routines, Learning mechanisms, Affective traits, and Social Ecological Structures*. Keeping with the strategy outlined previously, we attempted to define the minimum set of infant traits that are functioning before gaze or point following emerged and that would seem to be necessary for attention-sharing skills to emerge in their documented sequence. Traits were deemed necessary on the basis of a task analysis of the first attention-sharing skills to emerge. The theory eschews hypothetical special-purpose mechanisms (e.g., the shared attention mechanism; Baron-Cohen, 1995) under the philosophy that one should first prove that established, general mechanisms cannot explain some specific effect (e.g., gaze following) before postulating narrowly specialized mechanisms. This strategy indicated a starting age of 2 to 3 months. By that age, infants have all of the general precursor traits and ecological structures proposed (from the

task analysis) to be sufficient and needed to yield later, specific attention-following skills. The starting set includes the elements of the PLeASE acronym: perceptual routines, learning mechanisms, affective dispositions, and social ecology structures. (Note that this list is not comprehensive; many general phenotypes such as center-surround visual receptive fields, retinal heterogeneity of contrast and motion acuity, and audition attention traits are left implicit in the theory. However, we judged these phenotypes to be more distantly related to the phenotypes of interest.)

### Perceptual Routines

Several visual processes are hypothesized to be critical for attention sharing. Speed of attention shifting to a new target improves around 2 to 3 months (Butcher, Kalverboer, & Geuze, 2000; Johnson, Posner, & Rothbart, 1994) as recurrent connections from frontal eye fields (FEF) to superior parietal networks mature. This allows for top-down saccade planning that will be constrained by new (learned) factors, such as multidimensional salience maps (Itti & Koch, 2001) and experience-influenced autonomic neuromodulation (e.g., Aston-Jones & Cohen, 2005). These emergent changes in attention-shifting dynamics can have effects on infants' social attention (e.g., de Barbaro, Chiba, & Deák, 2011; Field, 1981). They also can be simulated in simple computational implementations (e.g., Nagai & Rohlfsing, 2009; Triesch et al., 2006).

The model also presumes that spatial mappings from the environment to retinal fields become mapped to sensorimotor loops (e.g., neck, torso, orbital muscles) within the first 4 to 6 months. These developments involve maturation of area V5/MT+ (higher level visual cortical regions), caudal FEF, and cerebellar networks (Rosander, 2007). Critically, they permit fast, smooth shifts of attention by coordination of multiple motor systems. Our simulations show that if infants' shifting is too slow, infants lose opportunities to use adults' gaze cues. Thus, we identified this aspect of visual maturation as a precursor of gaze following.

Another critical perceptual skill is discriminating adults' head poses (i.e., angles), which older infants use to estimate gaze direction (C. Moore, Angelopoulos, & Bennett, 1997). By 1 month of age, infants can discriminate frontal from profile head poses (Sai & Bushnell, 1988). Discrimination of head poses increases in acuity from 6 to 12 months (Butterworth & Jarrett, 1991), but little is known about this; for example, there are no data on acuity growth curves or on head features that infants use to discriminate head angles. Yet age limitations in infants' gaze following, especially to targets behind them, seem to be partly due to limited sensitivity to changes in head pose, durability of head pose representations, or both (Deák et al., 2000). Also,

the development of head pose sensitivity could plausibly explain premature capture errors in 6- to 9-month-olds (Butterworth & Jarrett, 1991). Younger infants sometimes follow gaze by turning to the correct hemifield but stop at the incorrect target (i.e., whichever is closer to midline), perhaps due to an imprecise encoding of the adult's head angle. Notably, in simulations of the PLeASES model (Triesch et al., 2006), infant-agents gained more rewards if they more accurately encoded caregivers' head pose. Thus, although young infants discriminated grossly dissimilar head poses, acuity improved with age. This could be a contributor to gaze following, a consequence of successful gaze following, or both.

The model also presumes that by 2 to 4 months of age, infants can, in optimal cases, discriminate different rotational positions of the eyes in the orbits—that is, eye direction. This presumption has empirical support (Farroni, Johnson, & Csibra, 2004; Symons, Hains, & Muir, 1998). However, there is no good evidence that infants under 12 to 18 months use eye direction, as opposed to head angle, to follow gaze (e.g., Butler, Caron, & Brooks, 2000; Doherty, 2006; C. Moore et al., 1997). Thus, young infants can detect eye direction but require many months to learn to associate it with adults' direction of attention, consistent with PLeASES: Infants would have to be able to perceive eye direction, but they require extensive input to learn that eye direction, which is subtler than but highly correlated with head angle, is uniquely predictive of adults' locus of attention. Thus, perceptual sensitivity alone is inadequate: We must also consider learning processes.

### Learning Mechanisms

The critical mechanisms of change proposed in PLeASES are temporal difference reinforcement learning (TD-RL) and habituation. Reinforcement learning is an established approach to machine learning (Sutton & Barto, 1998) with attractive features for modeling infant learning.<sup>1</sup> An insight of TD-RL is that adaptive behavior can be represented as learned policies of stochastic action selection based on a matrix of previously experienced situations and the outcomes (immediate and longer term) of actions previously taken in those situations. Outcomes are valued in terms of hedonic, material, or uncertainty-reducing outcomes, and immediate rewards are valued higher than delayed rewards. This model can be used to formalize a wide range of

<sup>1</sup>TD-RL functions can be captured by other learning approaches (e.g., partially observable Markov models, optimal control theory; Singh, Jaakkola, & Jordan, 1994; Wolpert, Ghahramani, & Flanagan, 2001). However, a TD-RL algorithm can capture some important and realistic ecological and cognitive constraints (e.g., partial feedback). The learning model was chosen on the basis of biological and cognitive plausibility: Infants are not optimal systems, and our goal is to replicate their errors and difficulties.



infant behaviors. For example, 2-month-olds can learn, after viewing alternating lights for a short time, to look at where the next light will be (Haith, Hazan, & Goodman, 1988). In this situation, if a light is more interesting than nothing and infants make different looking actions in the different event states (i.e., left on, right on, both off), then a TD-RL process can explain how infants' looking decision policy is shaped to yield such early and fast contingency learning.

There has been great progress in detailing some neural mechanisms of TD-RL (Schultz et al., 1997). Although we cannot experimentally test those mechanisms in human infants, we can test behavioral predictions in human or artificial infants (de Barbaro et al., 2011; Schlesinger & Parisi, 2001). For example, gaze following requires infants to notice a state of the environment (e.g., parent turned 90° to the left) and to choose either to keep looking or to look somewhere else, and if so, where. If, when the infant shifts gaze to the same target as the adult (after encoding his or her head or eye direction), and if that shared target is more interesting than other targets, the infant might strengthen the expected reward for selecting a similar attention shift in similar situations. That is not inevitable, as we shall explain below. First, however, we note several features of TD-RL models that are suited to problems of social prediction. One is that the models incorporate exploration (i.e., not always repeating the most-rewarded action) and stochastic action selection; this can explain the high variability of infants' responses to social cues. Another is that action policies can shift dynamically in response to changing environments (e.g., adapting to a new caregiver who is less demonstrative). Also, the models represent probabilistic memories for action outcomes, especially the most recent ones. This captures effects of personal history with a caregiver, as well as recency effects.

TD-RL models are complex enough that a full evaluation requires several distinct tests. First, it must be established how the agent classifies and differentiates states of the environment (e.g., head poses, regions of the environment), which over time can be associated with different actions. This requires psychophysical tests. Second, it is necessary to specify a priori reward values for different outcomes. This requires behavioral data such as looking time tests and facial expression coding. Third, it is necessary to show that previous actions affect future actions. These are significant challenges, but they are in fact challenges to all theories (e.g., How does a laboratory task generalize to everyday situations? Does praise carry the same value for every child in a sample?). The advantage of TD-RL models is that the assumptions are explicit, and researchers must be explicit about how they set the corresponding parameters or algorithms.

Habituation is another necessary learning mechanism. When an infant looks at a caregiver's face or a toy, habituation begins, and over time the

probability of a gaze shift gradually increases. This is necessary for infants to look away from parents, to seek out new and interesting sights, and to alternate gaze between caregivers and objects—a behavior taken as evidence that infants represent other people's attention (Tomasello, 1995). Habituation is often used as a methodological tool for infant studies but is mostly overlooked as a key learning mechanism in itself (Sirois & Mareschal, 2002). Yet, even neonates habituate to complex visual patterns (Slater, Earle, Morison, & Rose, 1985) and faces (Colombo, Mitchell, O'Brien, & Horowitz, 1987). Also, individual differences in habituation (e.g., rate) correlate with qualities of infant–parent interactions (Saxon, Frick, & Colombo, 1997; Tamis-LeMonda & Bornstein, 1989). It is possible that infants who habituate too fast or too slow relative to a caregiver's schedule of attention shifting miss opportunities to follow his or her cues to shift to interesting sights. In simulations, we parameterized habituation rate and found that very fast or very slow habituators were slower to learn gaze following (Triesch et al., 2006). This prediction is currently being tested in a longitudinal study (Ellis, Robledo, & Deák, 2012).

### Affective Traits

Reinforcement learning depends on the prior value of different outcomes. In social situations, shaping of action policies depends partly on what events or stimuli are interesting, fun, and so forth or are uncomfortable or stressful. Infants enjoy faces and voices, particularly those of caregivers (DeCasper & Fifer, 1980; Field, Cohen, Garcia, & Greenberg, 1984), and they enjoy participating in reciprocal, synchronized social interactions (Bigelow & Birch, 1999; G. A. Moore & Calkins, 2004). For example, 8- to 10-month-olds smile in anticipation of adults' next reaction (Venezia, Messinger, Thorp, & Mundy, 2004). Yet infants also show interest in complex objects such as toys (e.g., Rochat, 1989), and this can introduce conflict between the reward value of people and of objects (Bakeman & Adamson, 1984). The resulting dynamics can be modeled in simulations of the PLeASES model. For example, changing the relative values of faces and toys will, at extremes, yield patterns of disordered attention sharing: face avoidance that resembles autistic behaviors or hypersociability that resembles Williams syndrome (see Triesch et al., 2006). However, it is unknown how less extreme individual differences in these preferences affect the acquisition of attention sharing. It is also unknown what infants enjoy looking at in natural social settings, which must be known to test whether the PLeASES model can explain the acquisition of gaze and point following behaviors.

## Social Ecology Structures

Caregivers produce nonrandom social behavior patterns when interacting with infants (e.g., Cohn, Matias, Tronick, Connell, & Lyons-Ruth, 1986; Field, Healy, Goldstein, & Guthertz, 1990; Watson & Ramey, 1972), and infants learn these patterns (Kaye, 1982). For example, Nagai and Rohlfling (2009) and Zukow-Goldring and Arbib (2007) described how adults alter object manipulations to play with infants. Yu and Smith and their colleagues have described how parents' actions when sharing objects with toddlers, together with the toddlers' own activity and perception, jointly support toddlers' exploration and learning about the objects and even about object labels (Richert, Yu, & Favata, 2010; Smith, Yu, & Pereira, 2011; Yu & Smith, 2011). Parents also modify speech and gestures when interacting with infants (Brand, Baldwin, & Ashburn, 2002; Fernald & Kuhl, 1987). However, individual parents modify their behaviors with infants to varying degrees: Parents suffering from depression, for example, tend to be less reactive, and their infants learn different response policies than infants of nondepressed parents (Field et al., 1990; G. A. Moore & Calkins, 2004). Such findings must be accounted for by any theory. In initial simulations of the PLeASES model, Teuscher and Triesch (2007) varied the actions of virtual caregivers, making them, for example, very predictable, or neglectful, or chaotic and unpredictable. Infant-agents learned gaze following more or less quickly as a function of these styles, providing more evidence for the plausibility of the PLeASES theory.

However, little is known about how healthy, typical caregivers' social behavior patterns and individual or cultural differences affect infants' attention-sharing skills. To test whether PLeASES can explain the effects of these social structures, we must document what actual parents do while interacting with real infants and what infants do in response. Without this information, we cannot say what input infants have the opportunity to learn. Such information requires large videoethnographic data sets from which a wide range of caregiver actions are coded at high temporal resolution. Such data are costly to collect, however, so we are typically ignorant of infants' social information structures. Thus, our hypotheses about social input are often based on scant information from unnatural contexts, plus intuition and folk hypotheses. Yet these data sets are needed: They sometimes yield results (e.g., Smith et al., 2011) that are undetectable by other methods and at odds with conventional beliefs or inconsistent with results from the socially bizarre settings of laboratory experiments.

## Putting Together the Pieces

PLeASES predicts the following process: Infants in their first weeks develop a preference (i.e., reward) for the parent's face—frontal poses above

profiles—and for colorful objects. By 3 months, they smoothly shift gaze and habituate to static patterns on the same time scale as adults' fixation times. Parents often look at their infant when the infant is attentive, but they also look at other targets. Infants enjoy (i.e., are rewarded by) parents' direct gaze, but they often look away. In our model, this is because (a) other targets are also interesting; (b) habituation gradually reduces the value of the parent's face; and (c) in every time step, there is a possibility that the infant will explore—that is, look around to see if there happens to be something interesting. For these reasons, the cumulative likelihood of looking away from the parent gradually increases whenever the infant does look at the parent. Also, when the parent looks away, the infant sees a profile face, and instantaneous reward is reduced. Thus, the dynamics of TD-RL ensure that infants will eventually turn away. When they do, their next fixation target will not be random relative to the parent's gaze direction. Given the structure of the environment and similarities of the infant's and parent's visual systems, there is a relatively high probability that both will fixate on a nearby salient (e.g., colorful, moving, high-contrast) target. This requires the *Affective* and *Social Ecology* aspects of PLeASES. Sometimes, then, infant and parent will end up looking at the same target after the infant encoded the parent's head direction. Eventually, the infant will learn to expect higher accumulated reward by relating the parent's head poses (and later eye directions) to specific action commands (i.e., head turns) toward corresponding regions of the environment. This sequence involves *Perceptual* elements of PLeASES as well as TD-RL processes.

Testing the PLeASES theory required filling in unknown parameters. It was unknown what infants like to look at (i.e., find rewarding) in natural environments and whether these objects overlap with the things adults look at. It was also unknown how regularly infants see adults looking toward these objects or whether infants look toward these objects soon after they see the adult's head direction. One goal was to start to document these parameters in order to sharpen the predictions of PLeASES. Another goal was to sharpen and evaluate the model by formalizing and testing it in computational simulations using the above parameters to improve realism. A third goal has been to test behavioral predictions of PLeASES in a longitudinal study of infants. A fourth goal, more recently, has been to approach the mechanisms predicted by the PLeASES model using physiological studies of infants engaged in social interactions.

## Summary

PLeASES explains the emergence of attention-sharing skills in terms of dynamically interacting biological, sensorimotor, and ecological factors.



It has undergone initial testing through behavioral and physiological experiments, dense ethnographic studies, and computational simulations. Some preliminary findings are summarized next.

## TESTING THE PLeASES MODEL: INITIAL RESULTS

### Learning: What Do Infants Value?

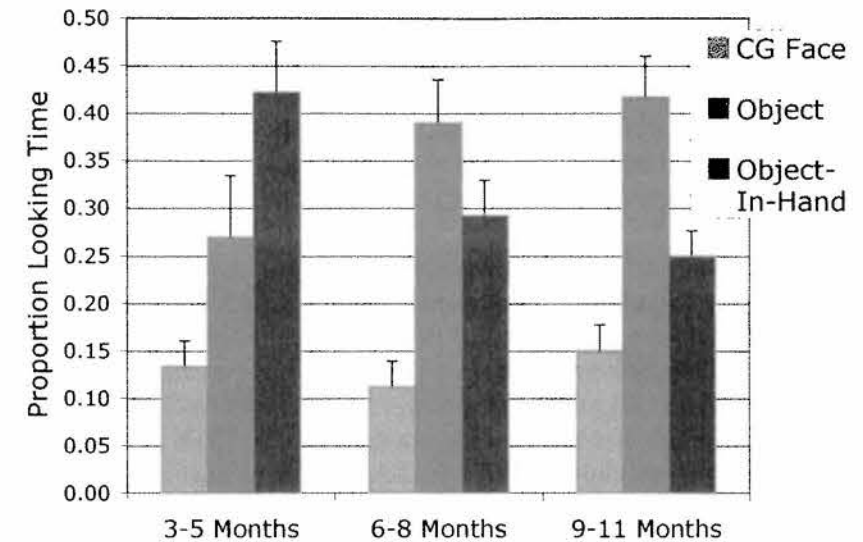
One goal is to replace assumptions about what infants find rewarding with empirical evidence. Without knowing what infants enjoy in social events, we cannot falsify TD-RL models. Our first test was a cross-sectional microethnographic study of 32 infants ages 3 to 11 months playing at home with caregivers (Deák, Krasno, Triesch, Lewis, & Sepeda, 2012). Initially, we expected infants to mostly fixate on caregivers' faces (it was commonly assumed that infants' favorite sight is their mother's face!). Then, when parents looked away, infants would eventually become bored and look elsewhere—fairly often, this would match where the parent was looking and would bring another fairly interesting target (e.g., toy) into view. Such sequences would gradually yield gaze following.

This neat story was not completely correct, but the real story is intriguing and provocative. First, although parents' faces are somewhat interesting to infants, they are not nearly as interesting as parents' hands manipulating objects. When parents were holding and moving things, infants of every age spent over 60% of total looking time, on average, watching adults' object handling. Considering the range of things that infants might look at when playing at home, this is a remarkably strong preference.

Figure 9.3 shows the proportion of total time infants spent looking at three sights: the parent's face, objects held by parents, or nonheld (static) objects. This graph includes the times when parents held nothing, so it underrepresents the robustness of infants' interest in watching object manipulation. These data belie the common assumption that infants most enjoy watching a parent's face (see also Smith et al., 2011). What, then, about the parents? Quite unlike infants, they spent most of the time looking at the infant's face. These facts, taken together, raise a puzzling question: If infants mostly watch objects, and parents mostly watch infants, how do infants ever learn gaze following?

### What Is the Learning Signal?

The answer lies in the margins—that is, in moments when infants do look at the parent and the parent happens to be looking away from the infant.



*Figure 9.3.* Infants' preferences for the caregiver's (CG's) face, objects held by the CG, and objects not held by the CG (e.g., sitting on highchair tray). Data are taken from a study of 35 infants ages 3 to 11 months, videotaped with their parents during a casual object play episode. Videos were coded frame by frame for gaze directions, manual actions, object kinds and locations, and speech and nonspeech sounds.

At those times, parents tend to look at their hands (Land, Mennie, & Rusted, 1999), and infants have an opportunity to see a parent looking toward their favorite sight (i.e., manipulating an object). If they do then turn from the parent's face to that sight, they receive a reward signal that would be temporally linked to a state of the environment that includes the adult's head pose. However, infants might instead see the parent's head (pointed to their hands), then look in another direction where they could see another, moderately rewarding stimulus, such as another toy (Krasno et al., 2007). This would provide misleading input for gaze following, because the infant would receive a reward signal for associating the parent's head pose with an unrelated location. Yet the scenario is plausible, especially in cluttered environments like homes. Critically, if such sequences occur as often as valid training sequences (i.e., seeing the parent's face, then high-value targets where the parent is also looking), then the learning component of the PLeASES would be disconfirmed. Thus, it was critical to determine where infants look just after they see the parent looking somewhere.

By virtue of having a large corpus of data coded frame by frame for gaze direction and gaze targets of mothers and infants, as well as manual actions of parents, we could verify that infants experience more of the former, teaching events than the latter, misleading events. This is true even with 3- to



5-month-olds, who cannot yet follow gaze, which means that infants receive a necessary training signal for RL in the form of structured events that are contingent on their actions. These results are captured in Figure 9.4: The left bar shows the mean rate of occurrences when infants turned from the mother's face to fixate in the same location where she was looking and manipulating an object. The right bar shows the mean rate of misleading events, when infants turned to a different location than the parent but nonetheless saw a toy (i.e., visual reward) there. This shows that infants were rewarded more often by looking in the same direction as an adult than by looking in another direction. This proves a necessary condition for a reinforcement learning process to yield gaze following.

These data also illustrate how social ecological data can establish the operating conditions for the development of a skill. Here, showing that infants like to look at adults' object handling not only disproved a common assumption (i.e., that infants most enjoy looking at faces) but also established what sort of events would serve as rewarding sights. Further, showing that

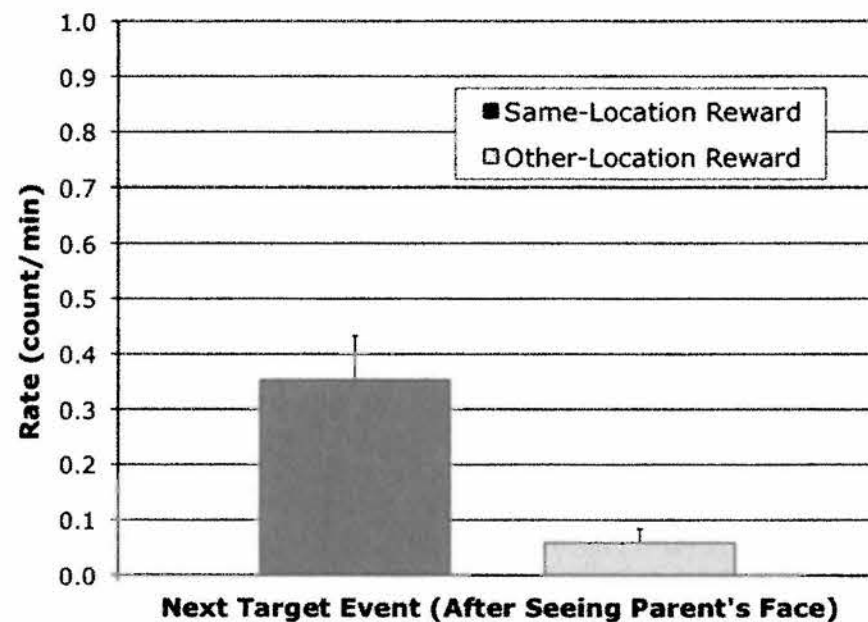


Figure 9.4. In the cross-sectional study, after (uncommon instances of) looking at the parent's face, infants were more likely to turn in the same direction as the parent's gaze and see a preferred (i.e., rewarded) sight than to turn in a different direction and see a rewarding sight in that other location. Typically the same-location "reward" was the parent manipulating an object (while looking at it), and the other-location reward was a static object. Thus, infants obtained a larger reward for turning in the same direction as the parent's gaze more often than they obtained a smaller reward for turning in a different direction.

infants are more likely to look toward their favorite sight after seeing the parent turned in that direction illustrates, for the first time in the infancy literature, that in the margins of everyday events, the social environment provides a statistically reliable, if weak, teaching signal.

As converging evidence that gaze following is an incidentally learned policy that has a modest profile, we queried the database for "pure" gaze following events—that is, when the infant saw the parent looking toward the region of a stationary toy and then turned to that same region. These events were quite rare, even among the oldest infants. This finding extends Deák et al.'s (2008) experimental finding that "pure" gaze following rarely occurs in face-to-face interactions within cluttered or distracting environments. Of course, older infants can follow gaze in stark experimental settings (Butterworth & Jarrett, 1991; Deák et al., 2000). This paradox has a possible solution in the previous results: Even if infants gradually learn gaze following via reinforcement learning, gaze following might almost never occur in situ because it is not needed. Parents do not usually sit on their hands, so to speak: They do things and often look at what they are doing. Infants might passively learn associations between parents' head poses and locations but respond to stronger cues and preferences in everyday settings. However, in the bizarre, stark laboratory setting, with none of the cues or rewards of an adult's manual actions or interesting toys, the most salient remaining source of information is the parent's head. In this low-stakes, low-conflict setting, infants can use their implicitly learned associations between head poses and location to select actions and look toward anything of marginal interest. In support of this explanation, Deák et al. (2008) found that infants in a stripped-down testing setting were more likely to follow gaze to interesting, distinctive targets than to boring, repetitive targets. One explanation is that when the few available targets were marginally more rewarding, the adult's gaze cue was marginally more valid, and infants learned to expect more reward for gaze following.

#### How Does the Social Environment Structure Attention Sharing?

A final analysis of the naturalistic data set stems from the question, If gaze following is nearly absent in naturalistic infant–parent interactions, how do attention-sharing episodes emerge? To determine this, we did a transitional state space analysis of dyadic states. These states include three characterized by coordinated attention: mutual gaze (State 1), shared attention (infant following parent; State 2), shared attention (parent following infant; State 3); they also include three states of uncoordinated attention: parent looking at infant but infant looking away (State 4), infant looking at parent but parent looking away (State 5), and each looking at different things (State 6).

Figure 9.5 schematizes the one-back transitions among these states. The frequency of each state is proportional to circle size. Line thickness is proportional to the number of transitions between states. Deviations from expected frequencies (based on marginal frequencies) are indicated by arrowheads (i.e., more than expected) or inhibition markers (i.e., fewer than expected). The results were surprising: We had expected that attention-sharing states would follow a canonical sequence: After a period of mutual gaze (State 1), one partner would turn away (State 4 or 5), and eventually the other part-

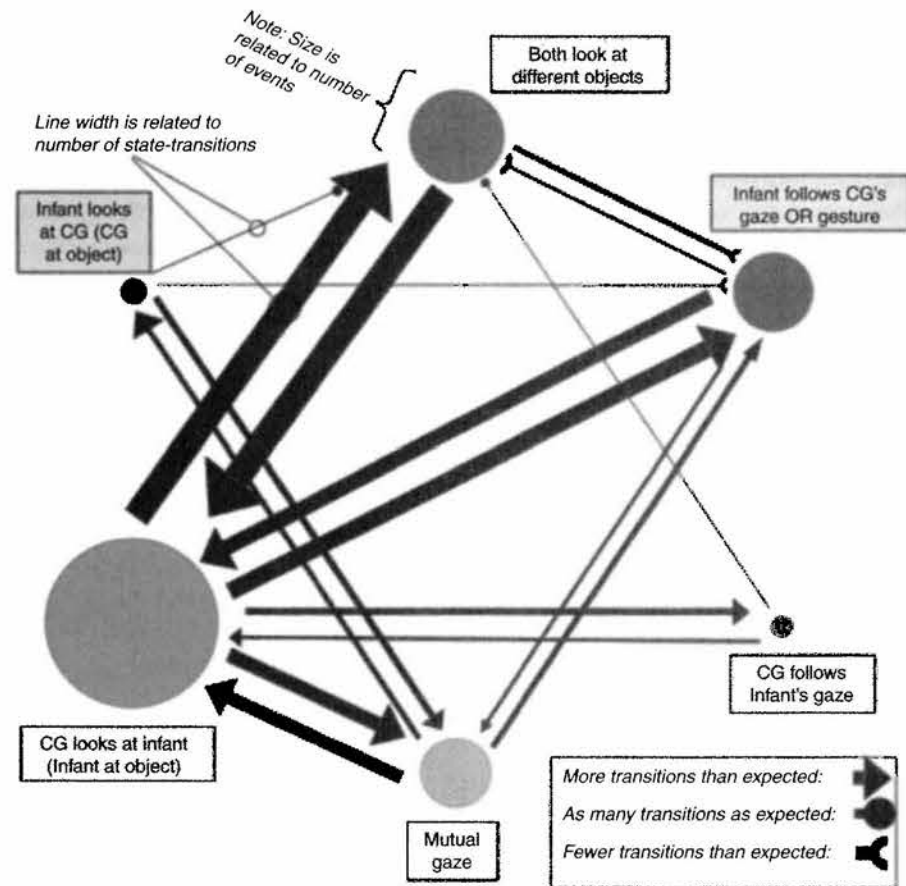


Figure 9.5. State-space transitions from cross-sectional observational study of 35 infants. Arrow thickness is proportional to frequency of successive state transitions. Note that states in which the caregiver (CG) is looking at the infant and the infant is looking at something else is a common "pivot" state leading either to shared attention (usually by infant looking at something the CG is manipulating or gesturing toward) or to both looking at different objects. Arrowheads indicate more transitions than expected given marginal expected frequencies; inhibition symbols indicate fewer transitions than expected.

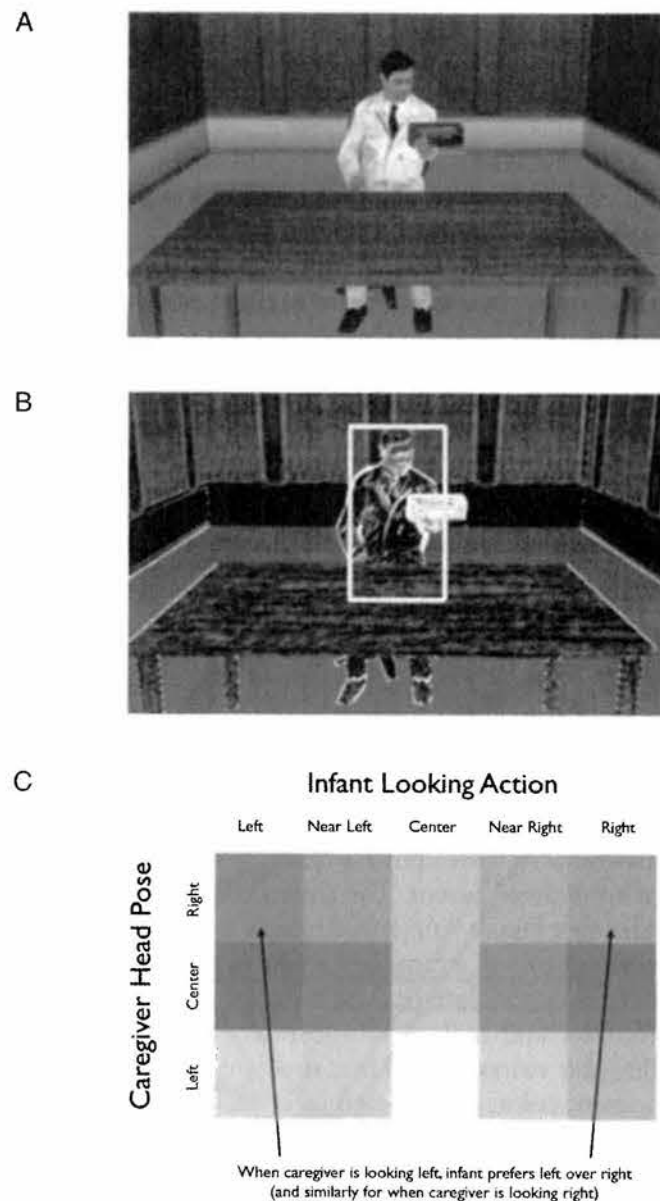
ner would turn to explore the same location, thereby establishing shared attention (State 2 or 3). However, as can be seen in Figure 9.5, mutual gaze (State 1) was uncommon and seldom proceeded in the expected sequence; in fact, only 0.7% of all three-step sequences followed that template. Rather, usually the infant looked away, and the parent kept watching the infant.

More commonly, shared attention was preceded by the parent's manual bid for the infant's attention: The parent kept watching the infant, primarily, and picked up a toy to play with. (It was in these moments when the parent sometimes looked to the object and the infant was set up to receive face-pose input, as explained previously.) Somewhat less often, infant and parent were looking at different things, then the infant shifted directly to the parent's target (usually because the parent was manipulating it). No one, including us, had predicted that these are the event sequences that set the stage for infant-parent attention sharing. However, two other data sets have shown some consistent phenomena (Amano, Kezuka, & Yamamoto, 2004; Smith et al., 2011).

### Can the PLeASES Theory Explain These Results?

In our early computational experiments, we did not know about infants' propensity to watch parents' object handling, and the simulations were not designed to test whether such patterns would emerge from a theoretical test of PLeASES. More recently, however, Lewis, Deák, Jasso, and Triesch (2010) used a 3D environment with a physics model in which a virtual infant-agent learned from a simulated parent. The parent was embodied as an anthropomorphic avatar (see Figure 9.6A), and targets were digitized images of multiple well-rendered toys in a furnished room, so the input was more visually realistic than in prior simulations. The infant-agent had TD-RL/habituation learning processes and biologically inspired visual processes (from the OpenCV computer vision library) that simulated a visual field with salience maps for contrast, color, and motion (see Itti & Koch, 2001). The visual field was converted to a multidimensional probability map, which tended to feature regions of salience around the caregiver and objects (Figure 9.6B). However, the presence of multiple toys and furnishings added clutter, or competing regions of salience, to simulate the effects of clutter on infants' attention following (Deák et al., 2008). The visual routines allowed rough discrimination of the parent-avatar's head poses, but specific knowledge of gaze direction was not given. Head poses might define different representational states in the infant-agent, but the infant-agent might not learn to associate the poses with locations in space.

A great advantage of this testing environment is that the parent-avatar could behave more realistically. Lewis et al. (2010) had the parent-avatar



**Figure 9.6.** Scenes and results from the MESA (Modeling the Emergence of Shared Attention) Project stage simulation environment. The caregiver replicates sequences of actions by “real” caregivers in naturalistic play with their infants. An infant learning agent carries out PLeASES (*P*erceptual routines, *L*earning mechanisms, *A*ffective traits, *S*ocial Ecology Structures) temporal difference reinforcement learning (TD-RL) and habituation learning on the dynamic scene. **A:** Full-vision snapshot of the infant’s visual field, with caregiver holding a high-contrast object. **B:** Weighted multidimensional saliency map calculated by the infant from that frame. **C:** Results from a pilot simulation (see text). Note that the original animations were in color, and color contrast was included in the OpenCV generated saliency maps.

produce series of actions from our frame-by-frame records of real parents’ actions during the naturalistic home play sessions (Krasno et al., 2007). Although the parent-avatar did not respond to the infant-agent, it produced strings of simplified actions—looking at things and picking up, moving, and putting down toys—with the same timing, order, and locations as real parents. Of course, real parents produce many more, and much richer, actions, and the simulation leaves out many potentially important details. Nonetheless, it was a first step toward an ecologically nonarbitrary test of theory-driven predictions about infants’ social learning and behaviors.

Initial results were promising: Within a reasonable number of time steps, the infant-agent learned to “predict” greater future reward for looking left when the parent-avatar looked left and for looking right when the parent-avatar looked right. This is shown in Figure 9.6C, in which lighter squares indicate a stronger reward prediction. It also generated a high expected value of locations at the center of the visual field, where the parent-avatar was located, as well as of objects held by the parent-avatar. Thus, the infant-agent learned first that its parent was interesting to look at and later learned that the parent’s head angle was predictive of the location (at least, left or right) of other interesting sights.

This result supports, broadly, the plausibility of the PLeASES model. An agent with no prior knowledge about head poses acquired rudimentary differentiation of, and responsiveness to, head and eye angles given a fairly sparse and weak set of very general perceptual, learning, and affective traits within an adult-structured, cluttered environment.

#### IMPLICATIONS OF A NEUROCONSTRUCTIVIST ACCOUNT

Microbehavioral ethnographic studies might give us insight into the social happenings of the real world, and computational simulations might establish the plausibility of a process model. However, PLeASES must also be biologically plausible. The facts of developmental neurobiology and systems neuroscience, as far as they are known, must constrain all predictions concerning infants’ processes of learning, perception, action, and affect in social settings.

What sorts of neural computations and representations might be required for gaze following? We know that humans represent both head angles and eye directions (Hooker et al., 2003). We know that locations in space become associated with one’s own actions, including head turns (e.g., Brotchie, Anderson, Snyder, & Goodman, 1995). We also can infer that vectors of perceived gaze direction in space are used to generate corresponding gaze shifting actions that will create an intersection of gaze. However, this mapping cannot



be represented simply as a “look-up table” (i.e., simple matrix) that relates each (observed) head angle to some “setting” for one’s own neck and eye muscles. The intersection of gaze vectors depends on the distances, locations, and relative headings of the infant and the adult. Gaze following thus requires the integration of information about the other person’s head (and eye) angles, the infant’s own visual and proprioceptive information, the infant’s representation of his or her own and others’ relative locations in an allocentric space, and possibly the layout and contents of the environment. Any theory must provide a plausible and detailed neurological account of how these spatial and perceptual–motor representations and routines can be learned so that they are reliably generated during social interactions. What, then, is known about the neural bases of these neural representations?

In most studies of adults’ hemodynamic responses to head and eye direction, activation is observed in several regions: posterior superior temporal sulcus (pSTS; e.g., Hooker et al., 2003), bilateral fusiform gyrus (FFG; George, Driver, & Dolan, 2001), and—especially in encoding direct versus averted gaze—amygdala (e.g., George et al., 2001). The converging evidence on activation of these areas might indicate a cortical and subcortical system that develops fairly efficient encoding of gaze and head direction cues (and perhaps other directional social cues like pointing; Macaluso, Driver, & Frith, 2003), even within a broader, multipurpose face processing network.

There is also evidence that the network that processes gaze information interacts with reward-computing networks: Kampe, Frith, Dolan, and Frith (2001) found that bilateral striatal responses to direct versus averted gaze were modulated by attractiveness of the stimulus face (see also Calder et al., 2002). Thus, perceived gaze direction impacts the reward values of outcomes (Schultz et al., 1997), which is consistent with PLeASES.

It is also noteworthy that encoded gaze direction activates not just the pSTS and FFG but also parts of intraparietal sulcus (IPS; e.g., Hoffman & Haxby, 2000). This finding is deemphasized, possibly due to a methodological artifact: Functional magnetic resonance imaging (fMRI) studies have almost always used disembodied, 2D face images that are static, canonical, repetitive, and cue invalid (e.g., there is no gaze target). Such stimuli and contexts are least likely to activate spatial maps. Yet even in these unnatural fMRI environments, Pelphrey, Singerman, Allison, and McCarthy (2003) found that IPS activation varied with whether or not a floating visual target matched a disembodied head’s eye direction. It was also modulated by the latency of the perceived gaze shift: Perhaps, as subjects’ attention to the target declined—that is, as they habituated—activation of that region of spatial maps also declined. This is what PLeASES would predict. Furthermore, and also consistent with the theory, one function of the FEF–IPS loop is to relate spatial cues to spatial maps for purposes of action planning (e.g., Andersen & Cui, 2009).

Interestingly, the parietal-prefrontal network that does saccade planning using, among other information, directional social cues also is subject to reinforcement learning. Campos, Breznen, Bernheim, and Andersen (2005) found that regions within the supplemental motor area compute expected reward related to gaze shifts. In light of the evidence reviewed above, this suggests that as adults, we have learned to relate social cues (e.g., direct vs. averted gaze) to gaze-shifting actions and to expect temporally discounted reward outcomes for those actions. This is exactly the prediction of PLeASES. Triesch et al. (2007) further explored the prediction in another simulation of gaze following acquisition and yielded a striking result: Through training, motor-planning units came to “mirror” location-specific units that corresponded to the eye and head direction input. That is, as the infant-agent learned to map caregiver head and eye angles to locations in allocentric space, location-sensitive processing units became coactivated by motor commands that shifted attention toward those same locations.

Such coactivated units fit the definition of mirror neurons (Rizzolatti & Craighero, 2004). Thus, the model predicts the existence of cells tuned for gaze following with mirror neuron properties (Triesch et al., 2007). Although at that time there was no evidence relevant to this prediction, Shepherd, Klein, Deaner, and Platt (2009) subsequently found cells that respond to both perceived gaze direction and saccade planning cells in macaque lateral intraparietal cortex. These are putative mirror cells for gaze direction. Although we cannot generalize from adult captive macaques to human infants, the converging evidence suggests a biologically plausible mechanism for the emergence of gaze following. The simulation result has a broader implication: The recent abundance of research and theory on the mirror neuron system has barely addressed how mirroring properties develop. Triesch et al.’s (2007) simulation, and the PLeASES theoretical framework, suggest one account. An expansive model by Grossberg and Vladusich (2010) offers another account. Both models share the goal of going beyond describing how mirror systems work to explain how they come to be.

#### FURTHER TESTS OF BEHAVIOR, PHYSIOLOGY, AND ECOLOGY OF ATTENTION SHARING

In a neuroconstructivist, embodied model like PLeASES, some early parameters will affect the trajectory of later-emerging behavioral phenotypes. For example, in any system that uses reinforcement learning, the prior values of various outcomes can affect learning and action selection. For example, preferred (i.e., higher valued) stimuli are more likely to be foci of attention. If those stimuli are related to some selective action, the preference can affect

skill learning. For example, in simulations, we varied the prior reward values of various stimuli for different infant-agents: The stimuli included the caregiver's face and objects. At the extremes, unbalanced values led to disordered joint attention. That is, autistic-looking gaze patterns were obtained by making the caregiver's face unrewarding; conversely, making faces too rewarding caused hypersociable gaze patterns reminiscent of Williams syndrome (Triesch et al., 2006, 2007). These results are evocative, but they point to just one possible phenotype among a range of phenotypic variations within each of these diverse disorders. A more theoretically powerful question is whether individual infants' preferences (i.e., comparative reward values) for different stimuli could affect their acquisition of joint attention skills. However, we found surprisingly little evidence on the stability and variability of infants' preferences for, for example, faces and toys. Do these preferences vary across individual infants? Do these preferences modulate the emergence of social routines such as gaze following? Perhaps, for example, infants who are less attracted to faces require more time to learn to map observed gaze directions onto their own saccades. We had hints that infants might show large individual differences in preferences for faces and objects: Infants in our ethnographic study (Krasno et al., 2007) spent an average of 12.8% of their time looking at their mother's face; however, across infants this ranged from 3% to 36% of time. Might these preferences predict how quickly an infant learns to discriminate gaze directions?

To address this, Robledo, Deák, and Kolling (2010) examined infants' sustained interest in photographs of faces and in colorful toys. This was tested every month and related to later gaze following skill. Interest in novel faces (i.e., total looking time) was moderately stable from 6 to 9 months (mean intermonth association  $r_{mean} = .39$ ). By contrast, attentiveness to toys was not stable. Also, infants' sustained interest in faces was weakly related to later gaze following: Recovery of interest in a novel face at 6, 8, and 9 months was correlated with gaze following from a controlled laboratory task at 9 months of age ( $r_{partials} .30-.46$ ), even with looking time to novel toys partialled out (to control for attentiveness or processing speed). Thus, interest in faces seems to be a modest predictor of gaze following in typically developing infants.

## AFFECTIVE DISPOSITIONS

The PLeASES model implies that the tendency to shift gaze at any moment is influenced by visual salience, the results of past action outcomes, and related predictive cues within structured environments. However, in reinforcement learning models, a parameter called *temperature* also matters. This refers to an agent's disposition to exploit actions that previously yielded

high rewards in similar situations versus exploring actions that yielded lower (or no) rewards in the past. Typically, machine-learning simulations have implemented temperature as a static variable. However, biological systems implement temperature as a dynamic variable. In particular, the locus coeruleus (LC), responding to signals from outcome- and expectancy-encoding networks, modulates norepinephrine (NE) expression. In the central nervous system, NE has widespread effects on cortical, cerebellar, and hippocampal targets (see Aston-Jones & Cohen, 2005). Some of these are effects on attention and arousal. LC output can shift between phasic and tonic states. During phasic LC output states, an animal's attention and learning are focused on a specific task or goal, and arousal (i.e., preparedness for action) is moderate. This can be construed as a bias toward *exploitation*. In tonic output states, attention and arousal are more distributed and less focused on a specific task or goal; this is overtly manifested in vigilant behavior and can be construed as a bias toward *exploration*. Notably, tonic (vigilance-biasing) LC states are negatively related to social affiliation, even in infants (Fortunato, Dribin, Granger, & Buss, 2008). Thus, we expect infants to shift gaze frequently and broadly instead of focusing on an adult social partner. In TD-RL models, this increase in "gain" to external stimuli can be represented as an increase in temperature, or a reduction in top-down guidance of attention. The model would therefore predict less cue following when infants are in high LC/NE-tonic states.

De Barbaro et al. (2011) reported evidence from human infants that fits this prediction: Infants in the MESA longitudinal study were, at 6 to 7 months of age, coded frame by frame for four visual behaviors, all related to NE-modulated vigilance in nonhuman mammals and in adult humans. The four behaviors were tightly correlated, indicating that individual infants showed coherent patterns of vigilant behavior. Notably, more-vigilant infants were less attentive to the adult experimenter, who periodically used pointing cues to indicate a distal target. More-vigilant infants did shift attention to the targets but were more compelled by intrinsic properties of the targets than by social cues.

Currently Zavala, de Barbaro, Chiba, and Deák (2010) are examining relations between these vigilance behaviors, social responsiveness, and concentrations of  $\alpha$ -amylase, a digestive molecule that is highly correlated with peripheral and central NE levels (Chatterton, Vogelsong, Lu, Ellman, & Hudgens, 1996). Saliva samples were collected from infants in the MESA longitudinal study at 6, 7, and 12 months of age. Levels of  $\alpha$ -amylase were assayed from these samples. Preliminary results show that  $\alpha$ -amylase levels were stable within sessions and somewhat stable across sessions. Moreover, the levels correlated with some, though not all, vigilance-related behaviors. We are currently analyzing the data to test for relations between joint attention



behaviors and  $\alpha$ -amylase levels. If so, it would suggest a link between LC-modulated attentiveness and arousal (temperature), and consequent changes in social actions including attention sharing.

## CONCLUSION

By understanding the development of infant social attention, we might gain insight into the processes of social development more generally. We might also gain insight into the origins of individual differences including, at the extreme, developmental disabilities (Karmiloff-Smith, 1998). We might even generate ideas for new interventions for the social-behavioral symptoms of disabilities. Only limited progress can be made from within a single discipline. Real progress will depend on integrating disciplines including psychology, neuroscience, artificial intelligence, anthropology and sociology, and others (e.g., linguistics). This is true whether or not the PLeASES theory continues to garner confirmatory data. To be sure, the current version of PLeASES lacks the means to explain later attention sharing outcomes, such as learning verbs of perception (e.g., *see*) or inferring another's visual perspective. These are elements we hope to develop in future work (e.g., Jao, Robledo, & Deák, 2010).

Our efforts so far have used computational simulations for proofs of the sufficiency of a theory and for greater specification of the theory. We stress that simulations are useful only insofar as they are biologically valid and insofar as the input structure is true to a "real" learning environment. Our efforts also make extensive use of microbehavioral and naturalistic behavioral evidence. Without that, we cannot know whether the behavioral output of the system is similar to real infants. Theory building and theory testing must be grounded in the structures and exigencies of real behavior in real environments. Qualitative, rough-coded ethnographic records are good sources of ideas but are inadequate for process models. Social actions occur within temporal intervals as short as 10 ms, so naturalistic social interactions should be coded at sampling rates that approach that granularity.

The goal of the PLeASES theory is to use the most relevant biological, cognitive, ecological, and microbehavioral information to explain how infants acquire new attention-sharing skills. It starts with a set of precursor phenotypes that we proposed as the minimum necessary for attention sharing behaviors to emerge, given a regimen of parent-provided social ecological structures. The precursor phenotypes are clearly demonstrable in young infants; no further special-purpose modules are proposed.

In tests of the PLeASES theory so far, we have discovered unknown phenomena. These include, for example, the fact that infants prefer to watch

adults manipulate objects and the role of this preference in indirect learning of gaze following responses. We also found that individual infants' sustained attentiveness to faces is a modest predictor of gaze following skills. Both of these findings are supported by simulations that implement the PLeASES model in an infant-agent. We have also extended the reinforcement learning aspect of the model to examine the temperature parameter, operationalized as the behavioral consequences of LC modulation of NE levels. Our results underscore the importance of this mechanism and its relevance to infants' attentiveness to social cues. Finally, our simulations have made novel predictions, such as the emergence of gaze following mirror neurons. Ongoing research is testing other predictions of the PLeASES model and exploring new questions that will flesh out the details of PLeASES. We believe that our approach is a model of the application of interdisciplinary concepts and methods to generate and test 21st-century theories of behavioral and cognitive development.

## REFERENCES

- Amano, S., Kezuka, E., & Yamamoto, A. (2004). Infant shifting attention from an adult's face to an adult's hand: A precursor of joint attention. *Infant Behavior & Development, 27*, 64–80. doi:10.1016/j.infbeh.2003.06.005
- Andersen, R. A., & Cui, H. (2009). Intention, action planning, and decision making in parietal-frontal circuits. *Neuron, 63*, 568–583. doi:10.1016/j.neuron.2009.08.028
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus–norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403–450. doi:10.1146/annurev.neuro.28.061604.135709
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother–infant and peer–infant interaction. *Child Development, 55*, 1278–1289. doi:10.2307/1129997
- Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418. doi:10.1017/S0305000900008345
- Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 41–59). Hillsdale, NJ: Erlbaum.
- Bigelow, A. E., & Birch, S. A. (1999). The effects of contingency in previous interactions on infants' preference for social partners. *Infant Behavior & Development, 22*, 367–382. doi:10.1016/S0163-6383(99)00016-8
- Brand, R., Baldwin, D., & Ashburn, L. (2002). Evidence for "motionese": Modifications in mothers' infant-directed action. *Developmental Science, 5*, 72–83. doi:10.1111/1467-7687.00211



- Brotchie, P. R., Anderson, R. A., Snyder, L. H., & Goodman, S. J. (1995, May 18). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, *375*, 232–235. doi:10.1038/375232a0
- 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 & Development*, *23*, 3–21. doi:10.1016/S0163-6383(00)00031-X
- Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant understanding of the referential nature of looking. *Journal of Cognition and Development*, *1*, 359–377. doi:10.1207/S15327647JCD0104\_01
- Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology*, *18*, 25–50. doi:10.1348/026151000165553
- 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., & Jarrett, N. L. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*, 55–72. doi:10.1111/j.2044-835X.1991.tb00862.x
- Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia*, *40*, 1129–1138. doi:10.1016/S0028-3932(02)00008-8
- Cameron, N. M., Champagne, F. A., Parent, C., Fish, E. W., Ozaki-Kuroda, K., & Meaney, M. J. (2005). The programming of individual differences in defensive responses and reproductive strategies in the rat through variations in maternal care. *Neuroscience and Biobehavioral Reviews*, *29*, 843–865. doi:10.1016/j.neubiorev.2005.03.022
- Campos, M., Breznen, B., Bernheim, K., & Andersen, R. A. (2005). The supplementary motor area encodes reward expectancy in eye movement tasks. *Journal of Neurophysiology*, *94*, 1325–1335. doi:10.1152/jn.00022.2005
- Cangelosi, A., Metta, G., Sagerer, G., Nehaniv, C., Fischer, K., Tani, J., . . . Zeschel, A. (2010). Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development*, *2*, 167–195. doi:10.1109/TAMD.2010.2053034
- Chatterton, R. T., Jr., Vogelsong, K. M., Lu, Y. C., Ellman, A. B., & Hudgens, G. A. (1996). Salivary alpha-amylase as a measure of endogenous adrenergic activity. *Clinical Physiology*, *16*, 433–448. doi:10.1111/j.1475-097X.1996.tb00731.x
- Cohn, J. F., Matias, R., Tronick, E. Z., Connell, D., & Lyons-Ruth, K. (1986). Face-to-face interactions of depressed mothers and their infants. *New Directions for Child and Adolescent Development*, *1986*, 31–45. doi:10.1002/cd.23219863405
- Colombo, J., Mitchell, D. W., O'Brien, M., & Horowitz, F. D. (1987). The stability of visual habituation during the first year of life. *Child Development*, *58*, 474–487. doi:10.2307/1130524
- Deák, G. O., Bartlett, M. S., & Jebara, T. (2007). How social agents develop: New trends in integrative theory-building. *Neurocomputing*, *70*, 2139–2147.
- Deák, G. O., Fasel, I., & Movellan, J. R. (2001). The emergence of shared attention: Using robots to test developmental theories. In C. Balkenius, J. Zlatev, H. Kozima, K. Dautenhahn, & C. Breazeal (Eds.), *Proceedings of the 1st International Workshop on Epigenetic Robotics: Modeling cognitive development in robotic systems* (Lund University Cognitive Studies Vol. 85, pp. 95–104). Lund, Sweden: Lund University Cognitive Science.
- Deák, G. O., Flom, R., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, *36*, 511–523. doi:10.1037/0012-1649.36.4.511
- Deák, G. O., Krasno, A., Triesch, J., Lewis, J., & Sepeda, L. (2012). *Watch the hands: Human infants can learn gaze-following by watching their parents handle objects*. Manuscript submitted for publication.
- Deák, G. O., & Triesch, J. (2006). The emergence of attention-sharing skills in human infants. In K. Fujita & S. Itakura (Eds.), *Diversity of cognition* (pp. 331–363). Kyoto, Japan: University of Kyoto Press.
- Deák, G., Wakabayashi, Y., Jasso, H., & Triesch, J. (2004, October). *Attention-sharing in human infants from 3 to 11 months of age in naturalistic conditions*. Paper presented at the 3rd International Conference on Development and Learning, La Jolla, CA.
- Deák, G. O., Walden, T. A., Yale, M., & Lewis, A. (2008). Driven from distraction: How infants respond to parents' attempts to elicit and re-direct their attention. *Infant Behavior & Development*, *31*, 34–50. doi:10.1016/j.infbeh.2007.06.004
- de Barbaro, K., Chiba, A., & Deák, G. O. (2011). Micro-analysis of infant looking in a naturalistic social setting: Insights from biologically based models of attention. *Developmental Science*, *14*, 1150–1160. doi:10.1111/j.1467-7687.2011.01066.x
- DeCasper, A. J., & Fifer, W. P. (1980, June 6). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*, 1174–1176. doi:10.1126/science.7375928
- Desrochers, S., Morissette, P., & Ricard, M. (1995). Two perspectives on pointing in infancy. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 85–101). Hillsdale, NJ: Erlbaum.
- Doherty, M. J. (2006). The development of mentalistic gaze understanding. *Infant and Child Development*, *15*, 179–186. doi:10.1002/icd.434
- Ellis, E., Robledo, M., & Deák, G. O. (2012). *Visual prediction in infancy: Does it predict later vocabulary?* Manuscript submitted for publication.
- 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. doi:10.1080/13506280050144399
- Farroni, T., Johnson, M. H., & Csibra, G. (2004). Mechanisms of eye gaze perception during infancy. *Journal of Cognitive Neuroscience*, *16*, 1320–1326. doi:10.1162/0898929042304787

- Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. (2004). Gaze following in newborns. *Infancy*, 5, 39–60. doi:10.1207/s15327078in0501\_2
- Fasel, I., Deák, G. O., Triesch, J., & Movellan, J. (2002). Combining embodied models and empirical research for understanding the development of shared attention. *Proceedings of the International Conference on Development and Learning*, 2, 21–27.
- Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*, 10, 279–293. doi:10.1016/0163-6383(87)90017-8
- Field, T., Healy, B., Goldstein, S., & Guthertz, M. (1990). Behavior-state matching and synchrony in infant–mother interactions of nondepressed versus depressed dyads. *Developmental Psychology*, 26, 7–14. doi:10.1037/0012-1649.26.1.7
- Field, T. M. (1981). Infant gaze aversion and heart rate during face-to-face interaction. *Infant Behavior & Development*, 4, 307–315. doi:10.1016/S0163-6383(81)80032-X
- Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother–stranger face discrimination by the newborn. *Infant Behavior & Development*, 7, 19–25. doi:10.1016/S0163-6383(84)80019-3
- Flom, R., Deák, G. O., Phill, C., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior & Development*, 27, 181–194. doi:10.1016/j.infbeh.2003.09.007
- Fortunato, C. K., Dribin, A. E., Granger, D. A., & Buss, K. A. (2008). Salivary alpha-amylase and cortisol in toddlers: Differential relations to affective behavior. *Developmental Psychobiology*, 50, 807–818. doi:10.1002/dev.20326
- George, N., Driver, J., & Dolan, R. J. (2001). Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing. *NeuroImage*, 13, 1102–1112. doi:10.1006/nimg.2001.0769
- Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, 14, 249–258. doi:10.1016/j.tics.2010.02.004
- Grossberg, S., & Vladusich, T. (2010). How do children learn to follow gaze, share joint attention, imitate their teachers, and use tools during social interactions? *Neural Networks*, 23, 940–965. doi:10.1016/j.neunet.2010.07.011
- 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. doi:10.2307/1130325
- Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3, 80–84. doi:10.1038/71152
- Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, 9, 131–134. doi:10.1111/1467-9280.00024
- Hooker, C. I., Paller, K. A., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2003). Brain networks for analyzing eye gaze. *Cognitive Brain Research*, 17, 406–418. doi:10.1016/S0926-6410(03)00143-5
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2, 194–203. doi:10.1038/35058500
- Jao, R. J., Robledo, M., & Deák, G. O. (2010). The emergence of referential gaze and perspective-taking in infants. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 284–289). Austin, TX: Cognitive Science Society.
- Johnson, M. K., Posner, M. I., & Rothbart, M. K. (1994). Facilitation of saccades toward a covertly attended location in early infancy. *Psychological Science*, 5, 90–93. doi:10.1111/j.1467-9280.1994.tb00636.x
- Kampe, K. K. W., Frith, C. D., Dolan, R. J., & Frith, U. (2001, October 11). Reward value of attractiveness and gaze. *Nature*, 413, 589–590. doi:10.1038/35098149
- Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, 2, 389–398. doi:10.1016/S1364-6613(98)01230-3
- Kaye, K. (1982). *The mental and social life of babies*. Chicago, IL: University of Chicago Press.
- Krasno, A., Deák, G., Triesch, J., & Jasso, H. (2007, April). *Watch the hands: Do infants learn gaze-following from parents' object manipulation?* Paper presented at the meeting of the Society for Research in Child Development, Boston, MA.
- Land, M., 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. doi:10.1068/p2935
- Lewis, J., Deák, G. O., Jasso, H., & Triesch, J. (2010). Building a model of infant social interaction. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 278–283). Austin, TX: Cognitive Science Society.
- Macaluso, E., Driver, J., & Frith, C. D. (2003). Multimodal spatial representations engaged in human parietal cortex during both saccadic and manual spatial orienting. *Current Biology*, 13, 990–999. doi:10.1016/S0960-9822(03)00377-4
- Messinger, D. M., Ruvolo, P., Ekas, N. V., & Fogel, A. (2010). Applying machine learning to infant interaction: The development is in the details. *Neural Networks*, 23, 1004–1016. doi:10.1016/j.neunet.2010.08.008
- Moore, C. (1996). Theories of mind in infancy. *British Journal of Developmental Psychology*, 14, 19–40. doi:10.1111/j.2044-835X.1996.tb00691.x
- 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. doi:10.1016/S0163-6383(97)90063-1
- Moore, G. A., & Calkins, S. (2004). Infants' vagal regulation in the still-face paradigm is related to dyadic coordination of mother–infant interaction. *Developmental Psychology*, 40, 1068–1080. doi:10.1037/0012-1649.40.6.1068

- Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development, 21*, 373–377. doi:10.1016/S0163-6383(98)90014-5
- Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*, 115–128. doi:10.1007/BF02206861
- Nagai, Y., & Rohlfing, K. J. (2009). Computational analysis of motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development, 1*, 44–54. doi:10.1109/TAMD.2009.2021090
- O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development, 67*, 659–677. doi:10.2307/1131839
- Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: The influence of context. *Neuropsychologia, 41*, 156–170. doi:10.1016/S0028-3932(02)00146-X
- Richardson, F., & 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*, 151–155. doi:10.1111/j.1467-7687.2006.00473.x
- Richert, M., Yu, C., & Favata, A. (2010). Joint attention through the hands: Investigating the timing of object labeling in dyadic social interaction. In *Proceedings of the IEEE 10th International Conference in Development and Learning* (pp. 114–119). Washington, DC: IEEE.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192. doi:10.1146/annurev.neuro.27.070203.144230
- Robledo, M., Danly, J. D., Acuña, J., Ramundo, A. A., & Deák, G. O. (2009, April). A longitudinal study of the emergence of attention-sharing in different contexts. Paper presented at the meeting of the Society for Research in Child Development, Denver, CO.
- Robledo, M., Deák, G. O., & Kolling, T. (2010). Infants' visual processing of faces and objects: Age-related changes in interest, and stability of individual differences. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2482–2487). Austin, TX: Cognitive Science Society.
- Rochat, P. (1989). Object manipulation and exploration in 2- to 5-month-old infants. *Developmental Psychology, 25*, 871–884. doi:10.1037/0012-1649.25.6.871
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York, NY: Oxford University Press.
- Rosander, K. (2007). Visual tracking and its relationship to cortical development. *Progress in Brain Research, 164*, 105–123. doi:10.1016/S0079-6123(07)64006-0
- Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., . . . Gorniak, P. (2006). The Human Speechome project. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 2059–2064). Wheat Ridge, CO: Cognitive Science Society.
- Sai, F., & Bushnell, W. R. (1988). The perception of faces in different poses by 1-month-olds. *British Journal of Developmental Psychology, 6*, 35–41. doi:10.1111/j.2044-835X.1988.tb01078.x
- Saxon, T. F., Frick, J. E., & Colombo, J. (1997). A longitudinal study of maternal interactional styles and infant visual attention. *Merrill-Palmer Quarterly, 43*, 48–66.
- Schlesinger, M., & Parisi, D. (2001). The agent-based approach: A new direction for computational models of development. *Developmental Review, 21*, 121–146. doi:10.1006/drev.2000.0520
- Schultz, W., Dayan, P., & Montague, P. R. (1997, March 14). A neural substrate of prediction and reward. *Science, 275*, 1593–1599. doi:10.1126/science.275.5306.1593
- Shepherd, S. V., Klein, J. T., Deaner, R. O., & Platt, M. L. (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 9489–9494. doi:10.1073/pnas.0900419106
- Simmering, V. R., Spencer, J., Deák, G., & Triesch, J. (2010). To model or not to model? A dialogue on the role of computational modeling in developmental science. *Child Development Perspectives, 4*, 152–158. doi:10.1111/j.1750-8606.2010.00134.x
- Singh, S., Jaakkola, T., & Jordan, M. I. (1994). Learning without state-estimation in partially observable Markovian decision processes. In W. W. Cohen & H. Hirsch (Eds.), *Proceedings of the 11th International Conference on Machine Learning*. San Francisco, CA: Morgan Kaufman.
- Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences, 6*, 293–298. doi:10.1016/S1364-6613(02)01926-5
- Slater, A., Earle, D. C., Morison, V., & Rose, D. (1985). Pattern preferences at birth and their interaction with habituation-induced novelty preferences. *Journal of Experimental Child Psychology, 39*, 37–54. doi:10.1016/0022-0965(85)90028-1
- Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science, 14*, 9–17. doi:10.1111/j.1467-7687.2009.00947.x
- Spink, A., Grieco, F., Krips, O., Loijens, L., Noldus, L., & Zimmerman, P. (Eds.). (2010). *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research*. Eindhoven, the Netherlands: Association for Computing Machinery.
- Sroufe, L. A. (1996). *Emotional development: The organization of emotional life in the early years*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511527661
- Stern, D. N. (2000). *The interpersonal world of the infant* (rev. ed.). New York, NY: Basic Books.
- Strogatz, S. H. (2001). Exploring complex networks. *Nature, 410*, 268–276. doi:10.1038/35065725



- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.
- Symons, L. A., Hains, S. M., & Muir, D. W. (1998). Look at me: Five-month-old infants' sensitivity to very small deviations in eye-gaze during social interactions. *Infant Behavior & Development, 21*, 531–536. doi:10.1016/S0163-6383(98)90026-1
- Tamis-LeMonda, C. S., & Bornstein, M. H. (1989). Habituation and maternal encouragement of attention in infancy as predictors of toddler language, play, and representational competence. *Child Development, 60*, 738–751. doi:10.2307/1130739
- Teuscher, C., & Triesch, J. (2007). To each his own: The caregiver's role in a computational model of gaze following. *Neurocomputing, 70*, 2166–2180. doi:10.1016/j.neucom.2006.02.023
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Mahwah, NJ: Erlbaum.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Triesch, J., Jasso, H., & Deák, G. O. (2007). Emergence of mirror neurons in a model of gaze following. *Adaptive Behavior, 15*, 149–165. doi:10.1177/1059712307078654
- Triesch, J., Teuscher, C., Deák, G., & Carlson, E. (2006). Gaze-following: Why (not) learn it? *Developmental Science, 9*, 125–147. doi:10.1111/j.1467-7687.2006.00470.x
- Venezia, M., Messinger, D. S., Thorp, D., & Mundy, P. (2004). The development of anticipatory smiling. *Infancy, 6*, 397–406. doi:10.1207/s15327078in0603\_5
- Watson, J. S., & Ramey, C. T. (1972). Reactions to response-contingent stimulation in early infancy. *Merrill-Palmer Quarterly, 18*, 218–227
- Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences, 5*, 487–494. doi:10.1016/S1364-6613(00)01773-3
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70*, 2149–2165. doi:10.1016/j.neucom.2006.01.034
- Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science, 14*, 165–180.
- Zavala, C., de Barbaro, K., Chiba, A., & Deák, G. O. (2010, January). *Does stress influence infant attention patterns over time?* Paper presented at the Temporal Dynamics of Learning Center Meeting, San Diego, CA.
- Zukow-Goldring, P., & Arbib, M. A. (2007). Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention. *Neurocomputing, 70*, 2181–2193. doi:10.1016/j.neucom.2006.02.029