The Development of Gaze Following as a Bayesian Systems Identification Problem

(INC MPLAB TR 2002.01 January 29 2002)

Javier R. Movellan

Institute for Neural Computation
UC San Diego

John S. Watson

Department of Psychology
UC Berkeley

Abstract

We propose a view of gaze following in which infants act as Bayesian learners actively attempting to identify the operating characteristics of the systems with which they interact. We present results of an experiment in which 28 infants (average age 10 months) interacted for a 3 minute period with a non-humanoid robot. For half the infants the robot simulated contingency structure typically produced by human beings. In particular it provided causal information about the existence of a line of regard. For the other 14 infants, the robot behaved in a manner which was not contingent with the environment. We found that a few minutes of interaction with the contingent robot was sufficient to elicit statistically detectable gaze following. There were clear signs that some of these infants were actively attempting to identify whether or not the robot was responsive to them. We propose that the infant brain is equipped to learn and analyze the contingency structure of real-time social interactions. Contingency is a fundamental perceptual dimension used by infants to recognize the operational properties of humans and to generalize existing behaviors to new social partners.

1 Introduction

By the end of the first year infants exhibit a variety of behaviors that reflect a rather sophisticated understanding of the operating characteristics of other human beings. These behaviors, which include pointing, ritualized reaching, and following the adult's line of regard, revolutionize the way infants interact with their caregivers. Developmental psychologists have been particularly interested on the emergence of gaze following as a basic indicator of our capacity to share attention and learn about the world by means of others. Scaife and Bruner [9] were first to investigate the development of gaze following in laboratory conditions. Infants were seated in front of an adult that interacted with them. At predetermined times the experimenter turned his head 90 degrees left or right and stayed there for 7 seconds. A experimenter scored whether during the 7 second period the infant looked in the direction pointed by the adult's face without intermediate looks elsewhere. They found that 81 % of infants of 8 months of age or older followed the line of regard on at least one trial. In contrast only 36 % of the infants between 2-7 months of age did so. Moore [10] recently reviewed current experiments on the emergence of gaze following and summarized them as follows: Under simplified conditions some experiments show that 3 month olds look statistically more often in the general direction of a head turn when there are peripheral targets in the visual field. By 6 months this behavior appears reliably. By 9 months infants follow head turns even if there are no visible targets, but they are not particularly sensitive to eye direction. Sensitivity to eye direction, not just head turns, is detectable at about 18 months of age.

While the experimental evidence is quite stable, the theoretical interpretations vary dramatically. One explanation is that infants follow gaze because they want to see what other people are looking at. According to this interpretation infants know that people can see and they put themselves in the perspective of others. Those who favor this interpretation point that it describes well why people follow gaze and that the knowledge required to do so is unlikely to be learned. The emergence of gaze following at about 9-months is explained by the maturation of highly specific knowledge modules. Interestingly, autistic children, who begin following head turns much later than typically developing infants, also have problems putting themselves in the perspective of others. In 1985 Baron-Cohen, Leslie, and Frith [18] tested autistic children and Down-Syndrome children, on the now famous "Sally-Ann" version of the "Wimmer-Perner" false belief task: The child is shown two dolls, one called Sally, and one called Ann. Sally places a marble in a covered basket basket and goes out. While she is out, her friend Ann moves Sally's marble from the basket to her own box, then she goes out. Sally comes back in and the child is asked "Where will Sally look for her marble?". 16 out of 20 autistic children with mental ages above 4 years said that Sally will look in the box, where the marble really is, and not in the basket. 12/14 children with Down Syndrome of lower mental age said she will look in the basket. Normally developing children about 4 year old are also known to succeed on this task.

However, some feel that simpler explanations about the emergence of gaze following are possible which do not require explicit knowledge that other people can see [2, 5]. Early forms of head turning can be explained as reflexive shifts in visual orienting caused by the head movement and followed by a capture of attention by a peripheral target [10]. The more advanced forms of gaze following that appear about 9 months of age can be explained by the fact that head turning is a cue to the appearance of interesting events in the direction of the turn, and infants progressively learn to rely on that cue. Chimpanzees are actually believed to follow gaze using this strategy [6, 7]. In fact learning experiments in which head turning of an adult is paired with activation of a toy in the direction of the turn produce significant increase in the gaze following behavior of 8-9 month old infants [23]. While we as adults believe that other people see and sometimes we turn our heads with an explicit intent to see what other people are looking at, it is unclear to what extent this knowledge controls the real time constrains required in social interactions. When a quarterback suddenly turns his head, defensive players have to rapidly react to this cue and behave accordingly. There is not much time for thinking. It is likely that the mechanisms responsible for learning these fast, real-time reactions play a crucial role in the development of social interactions from very early on. Rather than high-level knowledge about other people's minds causing the emergence of gaze-following, it is possible that learning to follow gaze in real-time interactions provides a foundation from which the more explicit (and slow) knowledge about others develops. In favor of this view is the fact that brains of autistic individuals very reliably exhibit severe abnormalities in the cerebellum, an organ known to handle real-time interactions with the environment [4].

2 Contingency detection and social development

Throughout his research career the second author of this paper has championed a view of infant development in which contingency detection plays a crucial role. According to this view infants are particularly good at analyzing the real-time causal texture of the world, and perceive it in a manner not unlike the way we perceive morphological features of a face. Intuitively we can see this form of perception at work when we recognize an old friend by his facial gestures in response to us even though his face may be barely recognizable due to facial hair, age and weight changes. Modern computer animation offers good opportunities to see this system at work. In some computer animated movies the characters are actually driven by actual human beings whose facial gestures and body movements are tracked in real time. In some occasions one has the distinct impression of recognizing the actor behind a computer character even though the character has very little physical resemblance with the actor animating it. The view of contingency as a fundamental perceptual property originated from an early learning experiment conducted by the second author of this document [24] in which 2-month-old infants learned to kick their legs to activate a mobile above their cribs. After 4 days of exposure to this controllable mobile, infants exhibited social smiles, positive affect, and cooing when the mobile was present. These social behaviors did not appear in a control group for which the mobile moved in a non-contingent manner. Watson proposed that contingency was a perceptual property used by infants to identify other humans and that in fact it was more powerful than other morphological properties of human faces (like the presence of eyes).

There is some evidence to support the idea that the infants' capacity to analyze contingency is quite sophisticated and that it is used to identify other humans. By about 4 months of age infants prefer to interact with objects which are responsive but not perfectly controllable [16, 25] suggesting, at least qualitatively, a preference for levels of responsiveness typical of social interactions. Bigelow [3] found that 4-5 month infants produced more contingent vocalization and social responses towards strangers which best approximated the level of responsiveness found in mother-infant interaction. Strangers that were more responsive or less responsive than mom were less preferred.

3 Bayesian Systems Identification

In statistics "systems identification" refers to the problem of making inferences about the structure of a system by observing how it responds to inputs [12]. The goal is to form a model of the system that can be used for prediction and control. Such models are specified in the form of a parameterized set of conditional probability distributions. Bayesian approaches to system identification emphasize the use of prior knowledge, i.e. a prior probability distribution over the set of possible system parameters. Recently the first author of this paper proposed an active learning approach for systems identification. In this approach the learner probes the system with those inputs that are expected to provide maximum information value [14, 13]. The approach described well how people identified concepts that other people were thinking about. Figure 1 shows an interesting example. The goal of subjects was to identify which number concept a person was thinking about. On the trial displayed in this figure, they were told that the numbers 16, 23, 19, and 20 where random examples from that concept. Subjects could then chose an integer from 1 to 100 and ask whether that integer belonged to the concept they where trying to identify. The figure shows the Bayesian information value[1] in bits of each question (continuous line) and the proportion of times subjects asked each question (dots). Note how people tended to ask questions with high information value, as one would expect from active Bayesian learners.

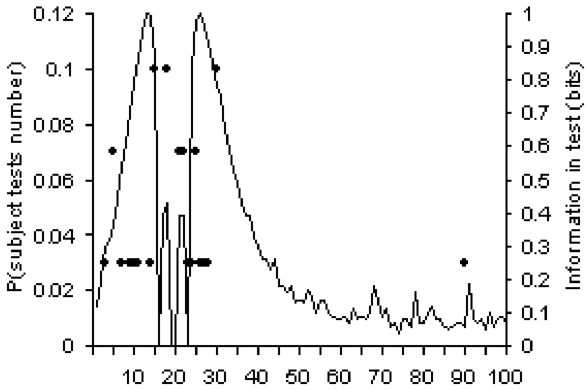

Figure 1: Example figure from Nelson, Tenenbaum and Movellan (2001) study on active Bayesian learning.

4 Bayesian functionalism

In this paper we promote a view of infants as active Bayesian learners whose goal is to identify the operating characteristics of the objects they interact with. We do so in the spirit of what the first author of this paper recently named "Bayesian functionalism" [11] which is closely related to the ideas in the rational movement in cognitive science [15]. While structural approaches emphasize the development of specific information processing models that can exhibit observed behaviors, in functional analysis the goal is to understand observed behaviors by showing that they are reasonable solutions to specific problems. The focus here is on specifying those problems and on providing methods to evaluate the goodness of the observed solutions. Bayesian theory is a useful mathematical framework for formalizing this approach, thus the name "Bayesian functionalism".

5 Robotics and Development

Robots present an ideal opportunity to study cognitive and social development in infants [21, 22, 2, 5]. It is possible to create robots that do not look particularly human and to program them to exhibit precisely controlled contingency structures. By observing how infants interact with these robots we may gain an understanding of the strategies they use to identify the operating characteristics of the objects with which they interact.

In this paper we present results from one of the first studies on infant-robot interaction we are aware of. We conducted the experiment in 1986 but due to historical reasons the work was only published as a short abstract [17]. Because of the recent interest on the use of robots to study development, we felt it is important to document the experiment in more detail, including our views regarding what the experiment tells us about the development of shared attention. The issue we addressed in this experiment was whether infants' sensitivity to the line of regard of others may be understood as a system's identification process which relies primarily on contingency information, not just the presence/absence of specific humanoid features.

6 Methods

Participants: 28 infants from a pool of volunteer families from the San Francisco Bay Area were randomly assigned to one of two groups. The experimental group consisted of 7 females and 7 males (mean age=10.5 months; sd=1.0). The control group consisted of (6 females, 8 males; mean age=10.6 months, sd=0.9).

Procedure: Infants were seated on their mother's laps, 1.5 meters in front of a robot head ($56 \times 21 \times 21$ cm). The mothers were wearing dark glasses and could not see the robot. Each of the sides of the robot's head was distinguished by an abstract pattern. In particular one of the sides (which played the role of the robot's face) was symmetric while the others were not (See Figure 2). Ninety degrees left and right of the robot there were two small boxes ($9 \times 9 \times 3$ cm) each of which had a small loud-speaker and a colored light. An IBM PC Jr. with 64KB of memory controlled the behavior of the robot and of the side boxes via a serial port interface. The computer could rotate the robot's head to "face" right or left, flash lights on its surface, or make sounds. It could also control the sounds and lights produced by the side boxes. The behavior of the system was programmed using a MSBasic interpreter. The first author of this document acted as a "sensor" informing the robot, via a joystick, that the infant had produced an interesting behavior (vocalizations or sudden movements of arms or legs). Other than this, the system was fully automatic.

Independent Variable: For the infants in the experimental group the robot was programmed to respond to the environment in a manner that simulated the contingency properties of human beings: The robot was only responsive

[1] Information value of questions was explicitly measured in number of bits with respect to a model of prior beliefs recently developed by Tenenbaum [1].

to visual events in front of one of its sides; these visual events included behaviors from the infant and flashes of lights produced by the box to the right side of the robot. In addition when objects in the environment (the infant and the box at the left side of the robot) produced interesting sounds, the robot's head rotated to "face" those objects. Thus this robot provided information that while it could respond to sounds from objects arbitrarily located in the room, it could only respond to visual events facing one of its sides. Each infant in the control group was matched to an infant in the experimental group and was presented the same temporal distribution of lights, sounds and turns of the central robot as was experienced by his/her matched participant. He/she also received the same number of stimuli from the side boxes but randomly distributed over the experimental session. Thus in the control group the robot behaved in a manner that was not predictable from the infant's behavior or from the side boxes. After 3.5 minutes of the robot interacting with the infant and with the box located at its left, all infants were tested for sensitivity to directional attention. There were 4 test trials each of which started with the robot producing intermittent light flashes and sounds until the infant fixated it. Immediately the robot turned its "face" to one side maintaining it there for 7 seconds. In 2 of the trials the robot's head rotated to face the box to its left, in the other 2 trials it rotated to face the box to its right. The order of the rotations was randomized across participants. On test trials the side boxes produced no lights or sounds.

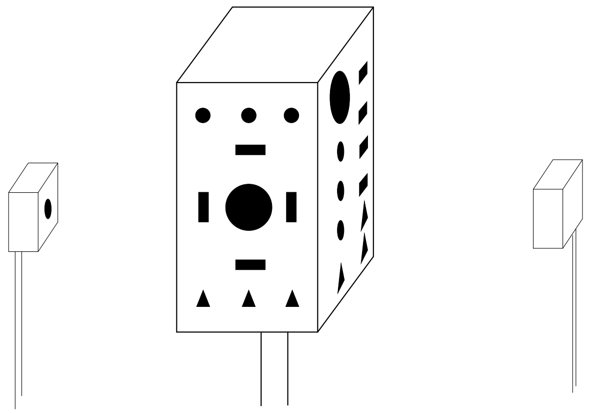

Figure 2: Schematic of the head of the center robot and side boxes.

Dependent Variables: The subjects' line of regard was coded as: looking towards the center robot, the right side robot, the left side robot, or looking away (e.g. up, down). Interactive behaviors (vocalizations and sudden arm or leg movements while looking at any of the three objects) were also recorded. Reliabilities across two different coders ranged from $r = 0.78$ to $r = 0.9$.

7 Results

Contingency detection: There was very clear evidence that infants behaved differently in response to the two contingency schedules. Infants in the experimental group exhibited about 5 times more vocalizations and sudden arm or leg movements during the 3.5 minute training period (Experimental Group =10.33 bpm, Control Group = 1.84 bpm[2], $p < .002$, one tail).

Emergence of shared attention: We found evidence that the infants were sensitive to the directional properties of the robot's behavior. First we measured the percentage of times the subjects looked to the left and right side boxes vs anywhere else. The same response measure during training was used as covariate. The experimental and control groups were significantly different (Looking rate: Experimental group=6.38 lpm[3] , Control group=3.23 lpm, $p < .005$, one tail, Percentage: Experimental group=61.2 %, Control group=38.2 %, $p < .03$, one tail).

In addition we assessed the proportion of testing trials in which the infants looked in the direction specified by the robot vs anywhere else. Infants in the experimental group looked proportionally more in the direction specified by the robot's rotation (Experimental group =32.29 %, Control group =18.35 %, $p < .04$, one tail[4]).

Finally, we also found evidence that infants followed the line of regard when the robot turned towards the box that had not been active during the training period, i.e., looking to the box at the robot's right side when the robot turned towards that box (Experimental group =1.8 lpm, Control group =0.6 lpm, $p < .03$, one tail, using the rate of looking left during the training phase as a covariate).

Qualitative observations: There were two particularly salient aspects of our experience running this experiment. First, a significant number of infants in the experimental group seemed to greatly enjoy the interaction with the robot, produced a large number of vocalizations, social smiles, and gave a distinct impression that they were treating the robot as if it were a conversational partner. One of the infants in particular laughed and "conversed" with the robot so loudly that a staff member entered the lab worried that something was wrong with this infant. Contrary to this, all the infants in the control group appeared

[2]bmp = behaviors per minute.

[3]lmp= looks per minute.

[4]After residual analysis [8] one control subject and the matched experimental subject were deleted (studentized residual= 3.97, Weisberg's t for residuals= 6.53, $p < .05$).

rather bored after a few moments of interaction with the robot. Second, some infants recognized whether the robot was responsive to them very quickly, in a matter of seconds. These infants showed clear qualitative signs that they were actively assessing the operational characteristics of the robot[5]. For example, infants would spontaneously vocalize. If such vocalization was followed by a robot's response, infants would intensively look at the robot and stop moving and vocalizing for a period of about 10 seconds, followed by another vocalization and an observation period. As learning progressed the observation periods became shorter and shorter. To provide a sense of the active nature of the infant's exploration we put an example video at http://mplab.ucsd.edu following links to demos and infant-robot interaction. At the time, the first author found it difficult to reconcile this behavior with the classic views on conditioning and the new associationist learning rules that were appearing in the connectionist literature. Only now, 16 years after the experiment was conducted, we feel we have a good formal framework to understand these behaviors from the point of view of active Bayesian systems identification.

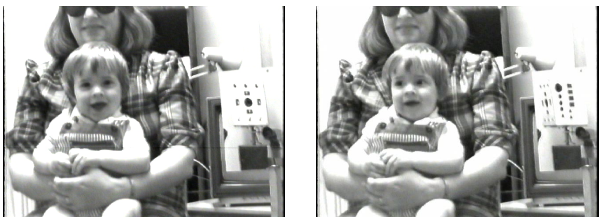

Figure 3: Example of infant following the robot's line of regard. The reflection of the robot through a mirror can be seen to the right of each image.

8 Discussion

Learning to interact efficiently with others is, at its roots, a systems-identification problem, for which human infants are typically well equipped. Our experiment suggests that by the end of the first year infants can use contingency information to ascertain in a matter of seconds whether a new object is responsive to them and to identify in a matter of minutes important general aspects of its operating characteristics (e.g., the fact that they have a line of regard).

These results cast some doubt on theories of gaze following which explain its emergence via specialized innate modules specifically designed to track morphological structures of the human face [20]. Infants appear perfectly capable of following the line of regard of systems that do not have eyes. We also found informal evidence that instead of slow associative learners, 10-month old infants behave in a manner reminiscent of the new Bayesian models of active learning that have been recently investigated by the first author of this article [13, 14].

Inspired by our experiment, Johnson, Slaughter and Carey [19] conducted a study in which eighty-three 12 month old infants faced an active object. After one minute of interaction with the infant, the object turned 45 degrees left or right. Their experimental design included the presence or absence of humanoid features and the presence of absence of contingency. One important difference between our study and theirs is that they did not modeled the robot turning to the side and finding something interesting happening there. Their results replicated our main finding that a few minutes of experience with a contingent object with no humanoid features could elicit gaze following. Their interpretation of the results was perhaps less friendly to learning approaches than we wish to be. In their view infants follow the line of regard because they are attributing intentions to the robot and contingency happens to be a marker for an early concept of "intentional being". Instead we rather think of infants as using the dimension of contingency to generalize appropriate behaviors about how to interact with other objects. For example based on interaction with human begins and other objects, unsupervised learning algorithms could create clusters of contingency structures. Prominent cluster would likely include some human beings, but the composition of these clusters may be impossible to describe with words. For example one such cluster may include a subset of human beings, animals and toys, while excluding other human beings. When a robot exhibits contingency structures characteristic of a cluster, behaviors used to interact with members of that cluster generalize to the robot.

At a qualitative level of analysis the infants in our experiment appeared to have a rather sophisticated prior knowledge about the space of possibilities for the operating characteristics of the robot. While the results do not inform us about how this knowledge was acquired, the rejection of learning as plausible explanation is not unlike the dismissal of mutation and natural selection as a plausible explanation for evolution. Our main message here is that contingency structure, not just morphological structure should play a critical role in whichever learning algorithms are proposed.

Thinking of early social development as a problem of real-time systems identification and control, brings about a change in priorities for future research. Critical is the gathering of databases and statistical analysis to characterize the causal properties of adult behavior in response to infants (including time constants, and contingency statistics).

[5] We are currently developing a Bayesian model of active sampling for this task to asses this point more formally.

Critical is the development of formal optimality models for systems identification and control of social objects. Critical is the study of the different trade-offs involved when endowing brains with different forms of prior knowledge. Ideally a functional approach may help us understand typical and atypical development as alternative solutions to these different trade-offs.

Acknowledgments

We thank the Fulbright program, which made it possible for the authors to interact for a period of five years. The first author dedicates his work in this paper to the memory of maestro Angel Riviere, a Spanish explorer of the human mind, who sparked the first author's interest on early social development. Some of the ideas presented here and the motivation to finally publish this work emerged from recent interactions between the first author of this document and Gedeon Deak, Jochen Triesch and Ian Fasel.

References

[1] Tenenbaum J. B. Rules and similarity in concept learning. In S. A. Solla, T. K. Leen, and K. R. Muller, editors, *Advances in Neural Information Processing Systems*, number 12, pages 59–65. MIT Press, Cambridge, Massachusetts, 2000.

[2] G. O. Deak, I. R. Fasel, and J. R. Movellan. The emergence of shared attention: using robots to test developmental theories. In Balkenius, editor, *Proceedings of the first international workshop on epigenetic robotics: modeling cognitive development*, pages 95–104. Lund University Cognitive Studies, Lund, Sweden, 2001.

[3] Bigelow A. E. Infant's sensitivity to imperfect contingency in social interaction. In P. Rochat, editor, *Early social cognition: understanding others in the first months of life*, pages 241–256. LEA, New York, 1999.

[4] Courchesne E., Saitoh O., Yeung-Courchesne R., Press G.A., Lincoln A.J., Haas R.H., and Schreibman L. Abnormality of cerebellar vermian lobules vi and vii in patients with infantile autism: Identification of hypoplastic and hyperplastic subgroups by mr imaging. *American Journal of Roentgenology*, 162:123–130, 1994.

[5] I. Fasel, G. O. Deak, J. Triesch, and J. R. Movellan. Combining embodied models and empirical research for understanding the development of shared attention. In *Proceedings of the second international conference on development and learning (ICDL02)*. MIT, Submitted.

[6] Povinelli D. J. and Eddy T. J. Chimpanzees: Joint visual attention. *Psychological Science*, 7:129–135, 1996.

[7] Povinelli D. J. and Eddy T. J. Specificity of gaze-following in young chimpanzees. *British Journal of De- velopmental Psychology*, 15:213–222, 1997.

[8] Stevens J. Outliers and influential data points in regression analysis. *Psychological Bulletin*, 95(2):233–344, 1984.

[9] Scaife M. and Bruner J. S. The capacity for joint visual attention in the infant. *Nature*, 253:265–266, 1975.

[10] C. Moore. Gaze following and the control of attention. In P. Rochat, editor, *Early social cognition: understanding others in the first months of life*, pages 241–256. LEA, New York, 1999.

[11] J. R. Movellan and J. Nelson. Probabilistic functionalism: A unifying paradigm for the cognitive sciences. *Behavioral and Brain Sciences*, In Press.

[12] O. Nelles. *Nonlinear System Identification : From Classical Approaches to Neural Networks and Fuzzy Models*. Springer, 2001.

[13] J. D. Nelson and J. R. Movellan. Active inference in concept induction. In T. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, number 13, pages 45–51. MIT Press, Cambridge, Massachusetts, In Press.

[14] J. D. Nelson, J. B. Tenenbaum, and J. R. Movellan. Active inference in concept learning. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 381–384. LEA, Edinburgh, Scotland, 2001.

[15] M. Oaksford and N. Chater. *Rational models of cognition*. Oxford University Press, Oxford, 1998.

[16] Bahrick L. R. and Watson J. S. Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy. *Developmental Psychology*, 21:963–973, 1985.

[17] Movellan J. R. and J. S. Watson. Perception of directional attention. In *Infant Behavior and Development: Abstracts of the 6th International Conference on Infant Studies*, NJ, 1987. Ablex.

[18] Baron-Cohen S., Leslie A.M., and Frith U. Does the autistic have a "theory of mind"? *Cognition*, 21:37–46, 1985.

[19] Johnson S., Slaughter V., and Carey S. Whose gaze will infants follow? the elicitation of gaze-following in 12-month-olds. *Developmental Science*, 1:233–238, 1998.

[20] Leekam S., S. Baron-Cohen, Perret D., Milders M., and Brown S. Eye direction detection: a dissociation between geometric and joint attention skills in autism. *British journal of developmental psychology*, 15:77–95, 1977.

[21] B. Scasselatti. Building behaviors developmentally:a new formalism. *Proceedings of the AAAI Spring Symposium on Integrating Robotics Research*, 1998.

[22] B. Scasselatti. How robotics and developmental psychology complement each other. *NSF/DARPA Workshop on Development and Learning*, 2000.

[23] Corkum V. and Moore C. The origins of joint visual attention in infants. *Developmental Psychology*, 34:28–38, 1998.

[24] J. S. Watson. Smiling, cooing and ”the game”. *Merrill-Palmer Quarterly*, 18:323–339, 1972.

[25] J. S. Watson. Contingency perception in earlyh social development. In Field T. M. and Fox N. A., editors, *Social perception in infants*, pages 157–176. Ables, New Jersey, 1985.