How to build robots that make friends and influence people

Cynthia Breazeal
cynthia@ai.mit.edu

Brian Scassellati
scaz@ai.mit.edu

MIT Artificial Intelligence Lab
545 Technology Square
Cambridge, MA 02139

Abstract

In order to interact socially with a human, a robot must *convey intentionality*, that is, the human must believe that the robot has beliefs, desires, and intentions. We have constructed a robot which exploits natural human social tendencies to convey intentionality through motor actions and facial expressions. We present results on the integration of perception, attention, motivation, behavior, and motor systems which allow the robot to engage in infant-like interactions with a human caregiver.

1 Introduction

Other researchers have suggested that in order to interact socially with humans, a software agent must be believable and life-like, must have behavioral consistency, and must have ways of expressing its internal states [2, 3]. A social robot must also be extremely robust to changes in environmental conditions, flexible in dealing with unexpected events, and quick enough to respond to situations in an appropriate manner [6].

If a robot is to interact socially with a human, the robot must *convey intentionality*, that is, the robot must make the human believe that it has beliefs, desires, and intentions [8]. To evoke these kinds of beliefs, the robot must display human-like social cues and exploit our natural human tendencies to respond socially to these cues.

Humans convey intent through their gaze direction, posture, gestures, vocal prosody, and facial displays. Human children gradually develop the skills necessary to recognize and respond to these critical social cues, which eventually form the basis of a theory of mind [1]. These skills allow the child to attribute beliefs, goals, and desires to other individuals and to use this knowledge to predict behavior, respond appropriately to social overtures, and engage in communicative acts. Using the development of human infants as a guideline, we have been building a robot that can interact socially with people.

From birth, an infant responds with various innate proto-social responses that allow him to convey subjective states to his caregiver. Acts that make internal processes overt include focusing attention on objects, orienting to external events, and handling or exploring objects with interest [14]. These responses can be divided into four categories. *Affective responses* allow the caregiver to attribute feelings to the infant. *Exploratory responses* allow the caregiver to attribute curiosity, interest, and desires to the infant, and can be used to direct the interaction to objects and events in the world. *Protective responses* keep the infant away from damaging stimuli and elicit concerned and caring responses from the caregiver. *Regulatory responses* maintain a suitable environment for the infant that is neither overwhelming nor under-stimulating.

These proto-social responses enable the adult to interpret the infant's actions as intentional. For example, Trevarthen found that during face-to-face interactions, mothers rarely talk about what needs to be done to tend to their infant's needs. Instead, nearly all the mothers' utterances concerned how the baby felt, what the baby said, and what the baby thought. The adult interprets the infant's behavior as communicative and meaningful to the situation at hand. Trevarthen concludes that whether or not these young infants are aware of their consequences of their behavior, that is, whether or not they have intent, their actions acquire meaning because they are interpreted by the caregiver in a reliable and consistent way.

The resulting *dynamics* of interaction between caregiver an infant is surprisingly natural and intuitive – very much like a dialog, but without the use of natural language (sometimes these interactions have been called proto-dialogs). Tronick, Als, and Adamson [15] identify five phases that characterize social exchanges between three-month-old infants and their caregivers:

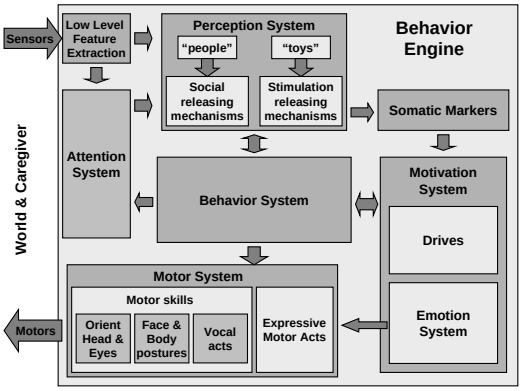

Figure 1: Overview of the software architecture. Perception, attention, internal drives, emotions, and motor skills are integrated to provide rich social interactions.

initiation, mutual-orientation, greeting, play-dialog, and *disengagement*. Each phase represents a collection of behaviors which mark the state of the communication. The exchanges are flexible and robust; a particular sequence of phases may appear multiple times within a given exchange, and only the initiation and mutual orientation phases must always be present.

The proto-social responses of human infants play a critical role in their social development. These responses enable the infant to convey intentionality to the caregiver, which encourages the caregiver to engage him as a social being and to establish natural and flexible dialog-like exchanges. For a robot, the ability to convey intentionality through infant-like proto-social responses could be very useful in establishing natural, intuitive, flexible, and robust social exchanges with a human. To explore this question, we have constructed a robot called Kismet that performs a variety of proto-social responses (covering all four categories) by means of several natural social cues (including gaze direction, posture, and facial displays). These considerations have influenced the design of our robot, from its physical appearance to its control architecture (see Figure 1). We present the design and evaluation of these systems in the remainder of this paper.

2 A Robot that Conveys Intentionality

Kismet is a stereo active vision system augmented with facial features that can show expressions analogous to happiness, sadness, surprise, boredom, anger, calm, displeasure, fear, and interest (see Figure 2).

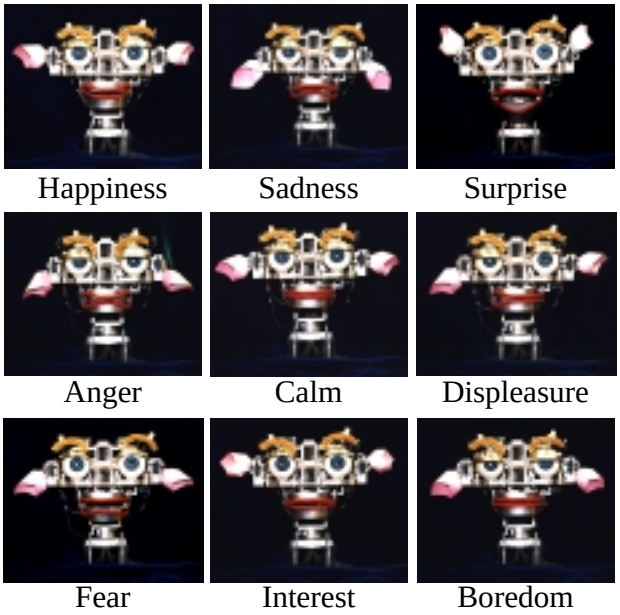

Figure 2: Kismet, a robot capable of conveying intentionality through facial expressions and behavior.

Kismet has fifteen degrees of freedom in facial features, including eyebrows, ears, eyelids, lips, and a mouth. The platform also has four degrees of freedom in the vision system; each eye has an independent vertical axis of rotation (pan), the eyes share a joint horizontal axis of rotation (tilt), and a one degree of freedom neck (pan). Each eyeball has an embedded color CCD camera with a 5.6 mm focal length. Kismet is attached to a parallel network of eight 50MHz digital signal processors (Texas Instruments TMS320C40) which handle image processing and two Motorola 68332-based microcontrollers which process the motivational system.

2.1 Perception and Attention Systems

Human infants show a preference for stimuli that exhibit certain low-level feature properties. For example, a four-month-old infant is more likely to look at a moving object than a static one, or a face-like object than one that has similar, but jumbled, features [10]. To mimic the preferences of human infants, Kismet's perceptual system combines three basic feature detectors: face finding, motion detection, and color saliency analysis. The face finding system recognizes frontal views of faces within approximately six feet of the robot under a variety of lighting conditions [12]. The motion detection module uses temporal differencing and region growing to obtain bounding boxes of mov-

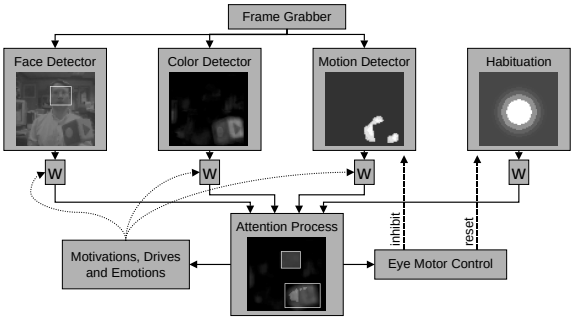

Figure 3: Kismet's attention and perception systems. Low-level feature detectors for face finding, motion detection, and color saliency analysis are combined with top-down motivational influences and habituation effects by the attentional system to direct eye and neck movements. In these images, Kismet has identified two salient objects: a face and a colorful toy block.

ing objects [5]. Color content is computed using an opponent-process model that identifies saturated areas of red, green, blue, and yellow [4]. All of these systems operate at speeds that are amenable to social interaction (20-30Hz).

Low-level perceptual inputs are combined with high-level influences from motivations and habituation effects by the attention system (see Figure 3). This system is based upon models of adult human visual search and attention [16], and has been reported previously [4]. The attention process constructs a linear combination of the input feature detectors and a time-decayed Gaussian field which represents habituation effects. High areas of activation in this composite generate a saccade to that location and compensatory neck movement. The weights of the feature detectors can be influenced by the motivational and emotional state of the robot to preferentially bias certain stimuli. For example, if the robot is searching for a playmate, the weight of the face detector can be increased to cause the robot to show a preference for attending to faces.

Perceptual stimuli that are selected by the attention process are classified into *social* stimuli (i.e. people, which move and have faces) which satisfy a drive to be social and *non-social* stimuli (i.e. toys, which move and are colorful) which satisfy a drive to be stimulated by other things in the environment. This distinction can be observed in infants through a preferential looking paradigm [14]. The percepts for a given classification are then combined into a set of *releasing mechanisms* which describe the minimal percepts necessary for a behavior to become active [11, 13].

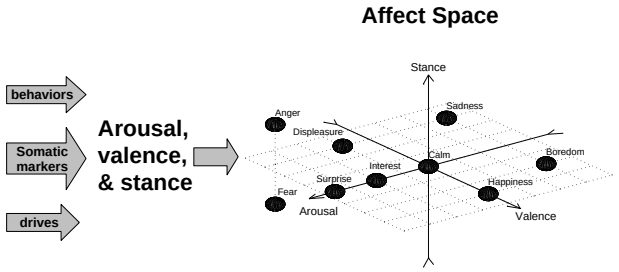

Figure 4: Kismet's affective state can be represented as a point along three dimensions: arousal, valence, and stance. This affect space is divided into `emotion` regions whose centers are shown here.

2.2 The Motivation System

The motivation system consists of `drives` and `emotions`. The robot's `drives` represent the basic "needs" of the robot: a need to interact with people (the `social` drive), a need to be stimulated by toys and other objects (the `stimulation` drive), and a need for rest (the `fatigue` drive). For each drive, there is a desired operation point, and an acceptable bounds of operation around that point (the homeostatic regime). Unattended, drives drift toward an under-stimulated regime. Excessive stimulation (too many stimuli or stimuli moving too quickly) push a drive toward an over-stimulated regime. When the intensity level of the drive leaves the homeostatic regime, the robot becomes motivated to act in ways that will restore the drives to the homeostatic regime.

The robot's `emotions` are a result of its affective state. The affective state of the robot is represented as a point along three dimensions: *arousal* (i.e. high, neutral, or low), *valence* (i.e. positive, neutral, or negative), and *stance* (i.e. open, neutral, or closed) [9]. The affective state is computed by summing contributions from the drives and behaviors. Percepts may also indirectly contribute to the affective state through the releasing mechanisms. Each releasing mechanism has an associated *somatic marker* processes, which assigns arousal, valence and stance tags to each releasing mechanism (a technique inspired by Damasio [7]).

To influence behavior and evoke an appropriate facial expression, the affect-space is divided into a set of `emotion` regions (see Figure 4). Each region is characteristic of a particular emotions in humans. For example, `happiness` is characterized by positive valence and neutral arousal. The region whose center is closest to the current affect state is considered to be active.

The motivational system influences the behavior selection process and the attentional selection process

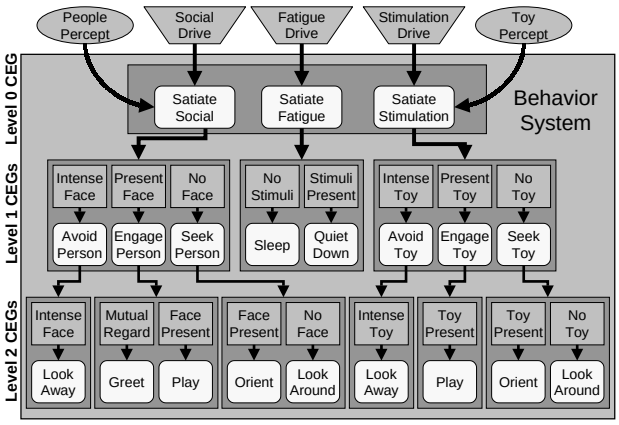

Figure 5: Kismet's behavior hierarchy consists of three levels of behaviors. Top level behaviors connect directly to drives, and bottom-level behaviors produce motor responses. Cross exclusion groups (CEG) conduct winner-take-all competitions to allow only one behavior in the group to be active at a given time.

based upon the current active `emotion`. The active `emotion` also provides activation to an affiliated expressive motor response for the facial features. The intensity of the facial expression is proportional to the distance from the current point in affect space to the center of the active `emotion` region. For example, when in the `sadness` region, the motivational system applies a positive bias to behaviors that seek out people while the robot displays an expression of sadness.

2.3 The Behavior System

We have previously presented the application of Kismet's motivation and behavior systems to regulating the intensity of social interaction via expressive displays [5]. We have extended this work with an elaborated behavior system so that Kismet exhibits key infant-like responses that most strongly encourage the human to attribute intentionality to it. The robot's internal state (emotions, drives, concurrently active behaviors, and the persistence of a behavior) combines with the perceived environment (as interpreted thorough the releasing mechanisms) to determine which behaviors become active. Once active, a behavior can influence both how the robot moves (by influencing motor acts) and the current facial expression (by influencing the arousal and valence aspects of the emotion system). Behaviors can also influence perception by biasing the robot to attend to stimuli relevant to the task at hand.

Behaviors are organized into a loosely layered, heterogeneous hierarchy as shown in Figure 5. At each level, behaviors are grouped into *cross exclusion groups* (CEGs) which represent competing strategies for satisfying the goal of the parent [3]. Within a CEG, a winner-take-all competition based on the current state of the emotions, drives, and percepts is held. The winning behavior may pass activation to its children (level 0 and 1 behaviors) or activate motor skill behaviors (level 2 behaviors). Winning behaviors may also influence the current affective state, biasing towards a positive valence when the behavior is being applied successfully and towards a negative valence when the behavior is unsuccessful.

Competition between behaviors at the top level (level 0) represents selection at the *global task* level. Level 0 behaviors receive activation based on the strength of their associated drive. Because the satiating stimuli for each drive are mutually exclusive and require different behaviors, all level 0 behaviors are members of a single CEG. This ensures that the robot can only act to restore one drive at a time.

Competition between behaviors within the active level 1 CEG represents *strategy* decisions. Each level 1 behavior has its own distinct winning conditions based on the current state of the percepts, drives, and emotions. For example, the `avoid person` behavior is the most relevant when the `social` drive is in the overwhelmed regime and a person is stimulating the robot too vigorously. Similarly, `seek person` is relevant when the `social` drive is in the under-stimulated regime and no face percept is present. The `engage person` behavior is relevant when the `social` drive is already in the homeostatic regime and the robot is receiving a good quality stimulus. To preferentially bias the robot's attention to behaviorally relevant stimuli, the active level 1 behavior can adjust the feature gains of the attention system.

Competition between level 2 behaviors represents *sub-task* divisions. For example, when the `seek person` behavior is active at level 1, if the robot can see a face then the `orient to face` behavior is activated. Otherwise, the `look around` behavior is active. Once the robot orients to a face, bringing it into mutual regard, the `engage person` behavior at level 1 becomes active. The `engage person` behavior activates its child CEG at level 2. The `greet` behavior becomes immediately active since the robot and human are in mutual regard. After the greeting is delivered, the internal persistence of the `greet` behavior decays and allows the `play` behavior to become active. Once the satiatory stimulus (in this case a face in mutual regard) has been obtained, the appropriate drive is

adjusted according to the quality of the stimulus.

2.4 The Motor System

The motor system receives input from both the emotion system and the behavior system. The emotion system evokes facial expressions corresponding to the currently active `emotion` (anger, boredom, displeasure, fear, happiness, interest, sadness, surprise, or calm). Level 2 behaviors evoke motor skills including `look around` which moves the eyes to obtain a new visual scene, `look away` which moves the eyes and neck to avoid a noxious stimulus, `greet` which wiggles the ears while fixating on a persons face, and `orient` which produces a neck movement with compensatory eye movement to place an object in mutual regard.

3 Mechanics of Social Exchange

The software architecture described above has allowed us to implement all four classes of social responses on Kismet. The robot displays *affective* responses by changing facial expressions in response to stimulus quality and internal state. A second class of affective response results when the robot expresses preference for one stimulus type. *Exploratory* responses include visual search for desired stimuli and maintenance of mutual regard. Kismet currently has a single *protective* response, which is to turn its head and look away from noxious or overwhelming stimuli. Finally, the robot has a variety of *regulatory* responses including: biasing the caregiver to provide the appropriate level of interaction through expressive feedback; the cyclic waxing and waning of affective, attentive, and behavioral states; habituation to unchanging stimuli; and generating behaviors in response to internal motivational requirements.

Figure 6 plots Kismet's responses while interacting with a toy. All four response types can be observed in this interaction. The robot begins the trial looking for a toy and displaying sadness (an affective response). The robot immediately begins to move its eyes searching for a colorful toy stimulus (an exploratory response) ($t < 10$). When the caregiver presents a toy ($t \approx 13$), the robot engages in a play behavior and the stimulation drive becomes satiated ($t \approx 20$). As the caregiver moves the toy back and forth ($20 < t < 35$), the robot moves its eyes and neck to maintain the toy within its field of view. When the stimulation becomes excessive ($t \approx 35$), the robot

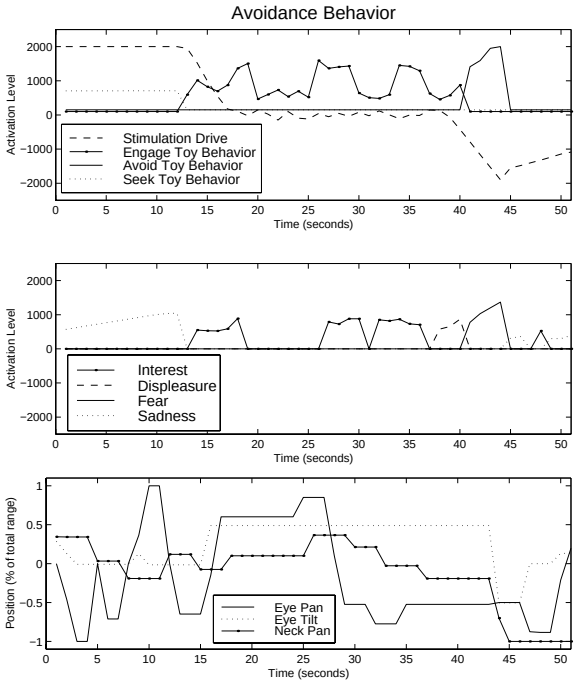

Figure 6: Kismet's response to excessive stimulation. Behaviors and drives (top), emotions (middle), and motor output (bottom) are plotted for a single trial of approximately 50 seconds. See text for description.

becomes first displeased and then fearful as the stimulation drive moves into the overwhelmed regime. After extreme over-stimulation, a protective avoidance response produces a large neck movement ($t = 44$) which removes the toy from the field of view. Once the stimulus has been removed, the stimulation drive begins to drift back to the homeostatic regime (one of the many regulatory responses in this example).

To evaluate the effectiveness of conveying intentionality (via the robot's proto-social responses) in establishing intuitive and flexible social exchanges with a person, we ran a variant of a social interaction experiment. Figure 7 plots Kismet's dynamic responses during face-to-face interaction with a caregiver in one trial. This architecture successfully produces interaction dynamics that are similar to the five phases of infant social interactions described in [15]. Kismet is initially looking for a person and displaying sadness (the initiation phase). The robot begins moving its eyes looking for a face stimulus ($t < 8$). When it finds the caregiver's face, it makes a large eye movement to enter into mutual regard ($t \approx 10$). Once the face is foveated, the robot displays a greeting behavior by wiggling its ears ($t \approx 11$), and begins a play-dialog phase of interaction with the caregiver ($t > 12$).

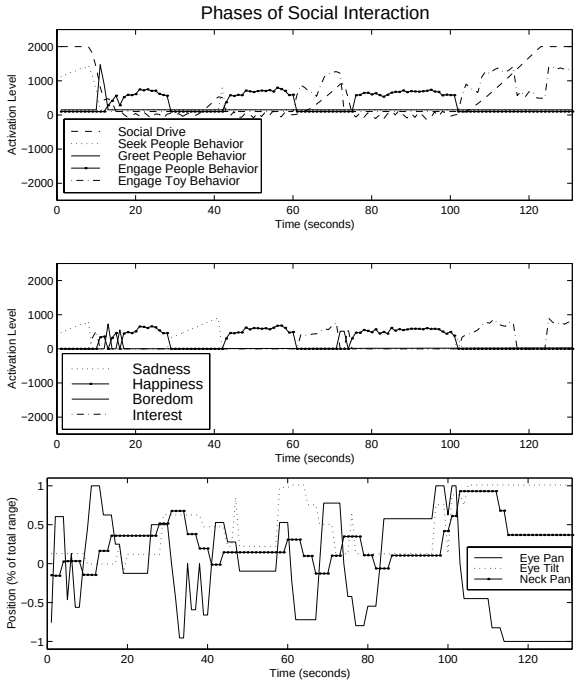

Figure 7: Cyclic responses during social interaction. Behaviors and drives (top), emotions (middle), and motor output (bottom) are plotted for a single trial of approximately 130 seconds. See text for description.

Kismet continues to engage the caregiver until the caregiver moves outside the field of view ($t \approx 28$). Kismet quickly becomes sad, and begins to search for a face, which it re-acquires when the caretaker returns ($t \approx 42$). Eventually, the robot habituates to the interaction with the caregiver and begins to attend to a toy that the caregiver has provided ($60 < t < 75$). While interacting with the toy, the robot displays interest and moves its eyes to follow the moving toy. Kismet soon habituates to this stimulus, and returns to its play-dialog with the caregiver ($75 < t < 100$). A final disengagement phase occurs ($t \approx 100$) as the robot's attention shifts back to the toy.

In conclusion, we have constructed an architecture for an expressive robot which enables four types of social responses (affective, exploratory, protective, and regulatory). The system dynamics are similar to the phases of infant-caregiver interaction [15]. These dynamic phases are not explicitly represented in the software architecture, but instead are emergent properties of the interaction of the control systems with the environment. By producing behaviors that convey intentionality, we exploit the caregiver's natural tendencies to treat the robot as a social agent, and thus to respond in characteristic ways to the robot's overtures. This reliance on the external world produces dynamic behavior that is both flexible and robust. Our future work will focus on measuring the quality of the interactions as perceived by the human caregiver and on enabling the robot to learn new behaviors and skills which facilitate interaction.

References

[1] S. Baron-Cohen. *Mindblindness*. MIT Press, 1995.

[2] J. Bates. The role of emotion in believable characters. *Communications of the ACM*, 1994.

[3] B. Blumberg. *Old Tricks, New Dogs: Ethology and Interactive Creatures*. PhD thesis, MIT, 1996.

[4] C. Breazeal and B. Scassellati. A context-dependent attention system for a social robot. In *1999 International Joint Conference on Artificial Intelligence*, 1999. Submitted.

[5] C. Breazeal and B. Scassellati. Infant-like social interactions between a robot and a human caretaker. *Adaptive Behavior*, 8(1), 2000. To appear.

[6] R. Brooks. Challenges for complete creature architectures. In *Proceedings of Simulation of Adaptive Behavior (SAB90)*, 1990.

[7] A. R. Damasio. *Descartes' Error*. G.P. Putnam's Sons, New York, 1994.

[8] K. Dautenhahn. Ants don't have friends – thoughts on socially intelligent agents. Technical report, AAAI Technical Report FS 97-02, 1997.

[9] P. Ekman and R. Davidson. *The Nature of Emotion: Fundamental Questions*. Oxford University Press, New York, 1994.

[10] J. F. Fagan. Infants' recognition of invariant features of faces. *Child Development*, 47:627–638, 1976.

[11] K. Lorenz. *Foundations of Ethology*. Springer-Verlag, New York, NY, 1973.

[12] B. Scassellati. Finding eyes and faces with a foveated vision system. In *Proceedings of the American Association of Artificial Intelligence (AAAI-98)*, 1998.

[13] N. Tinbergen. *The Study of Instinct*. Oxford University Press, New York, 1951.

[14] C. Trevarthen. Communication and cooperation in early infancy: a description of primary intersubjectivity. In M. Bullowa, editor, *Before Speech*, pages 321–348. Cambridge University Press, 1979.

[15] E. Tronick, H. Als, and L. Adamson. Structure of early face-to-face communicative interactions. In M. Bullowa, editor, *Before Speech*, pages 349–370. Cambridge University Press, 1979.

[16] J. M. Wolfe. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238, 1994.