THE ROYAL SOCIETY

Published online 20 January 2003

The enactive mind, or from actions to cognition: lessons from autism

Ami Klin*, **Warren Jones, Robert Schultz and Fred Volkmar**

Yale Child Study Center, Yale University School of Medicine, 230 South Frontage Road, New Haven, CT 06520, USA

Normative-IQ individuals with autism are capable of solving explicit social cognitive problems at a level that is not matched by their ability to meet the demands of everyday social situations. The magnitude of this discrepancy is now being documented through newer techniques such as eye tracking, which allows us to see and measure how individuals with autism search for meaning when presented with naturalistic social scenes. This paper offers an approach to social cognitive development intended to address the above discrepancy, which is considered a key element for any understanding of the pathophysiology of autism. This approach, called the enactive mind (EM), originates from the emerging work on 'embodied cognitive science', a neuroscience framework that views cognition as bodily experiences accrued as a result of an organism's adaptive actions upon salient aspects of the surrounding environment. The EM approach offers a developmental hypothesis of autism in which the process of acquisition of embodied social cognition is derailed early on, as a result of reduced salience of social stimuli and concomitant enactment of socially irrelevant aspects of the environment.

Keywords: autism; enactive mind; embodied cognition; theory of mind

1. SOCIAL FUNCTIONING IN EXPLICIT VERSUS NATURALISTIC SITUATIONS

One of the most intriguing puzzles posed by individuals with autism is the great discrepancy between what they can do on explicit tasks of social reasoning (when all of the elements of a problem are verbally given to them), and what they fail to do in more naturalistic situations (when they need to spontaneously apply their social reasoning abilities to meet the moment-by-moment demands of their daily social life) (Klin *et al.* 2000). While even the most intellectually gifted individuals display deficits in some complex social reasoning problems (Happé 1994; Baron-Cohen *et al.* 1997), some, particularly those without cognitive deficits, can solve such problems at relatively high levels (Bowler 1992; Dahlgren & Trillingsgaard 1996) without showing commensurate levels of social adaptation. This discrepancy is troublesome because, while it is possible to teach them better social reasoning skills, such new abilities may have little impact on their real-life social or communicative competence (Ozonoff & Miller 1995; Hadwin *et al.* 1997).

There has been little systematic research to investigate the magnitude of this discrepancy. Nevertheless, an indicator of its size can be derived from a sample of 40 older adolescents and adults with autism followed in our centre. Their full-scale IQs are within the normative range, whereas their mean age equivalent score on the interpersonal relationships sub-domain of the *Vineland Adaptive Behaviour Scales* (Sparrow *et al.* 1984) is 4 years. These individuals have many cognitive, linguistic, knowledge-based and potentially useful vocational assets, and yet this social adaptive score would suggest that if left to their own devices in a challenging social situation, their 'social survival' skills or 'street smarts' might be equivalent to those of young children. However, many of these individuals are capable of a degree of self-sufficiency that is much higher than 4 years. It is possible that they are able to achieve this level of independence despite significant social disabilities by choosing highly structured and regimented life routines that avoid novelty and the inherent unpredictability of typical social life. In other words, they may be able to constrain the inevitable complexity of social life by setting themselves a routine of rigid rules and habits, adhering very closely to this lifestyle in what is, typically, a very solitary life.

Some recent studies focusing on responses to naturalistic social situations suggest that the discrepancy between performance on structured as against naturalistic tasks may be even greater than hitherto thought possible. Consider the following two examples from eye-tracking studies of normative-IQ adolescents and adults with autism. In these experiments (Klin *et al.* 2002*a*,*b*), eye-tracking technology allows researchers to see and measure what a person is visually focusing on when viewing complex social situations. This paradigm allows for an appreciation of a person's spontaneous reactions to naturalistic demands inherent in seeking meaning in what is viewed. In real-life social situations, many crucial social cues occur very rapidly. Failure to notice them may lead to a general failure in assessing the meaning of entire situations, thus precluding adaptive reactions to them. Figure 1 shows a still image of two characters from a film: a young man on the left and a young woman on the right. Overlaid on the image are crosses that mark, in black, the focus of a normative-IQ adult with autism and, in white, the focus of a

*Author for correspondence (ami.klin@yale.edu).

One contribution of 14 to a Theme Issue 'Autism: mind and brain'.

DOI 10.1098/rstb.2002.1202

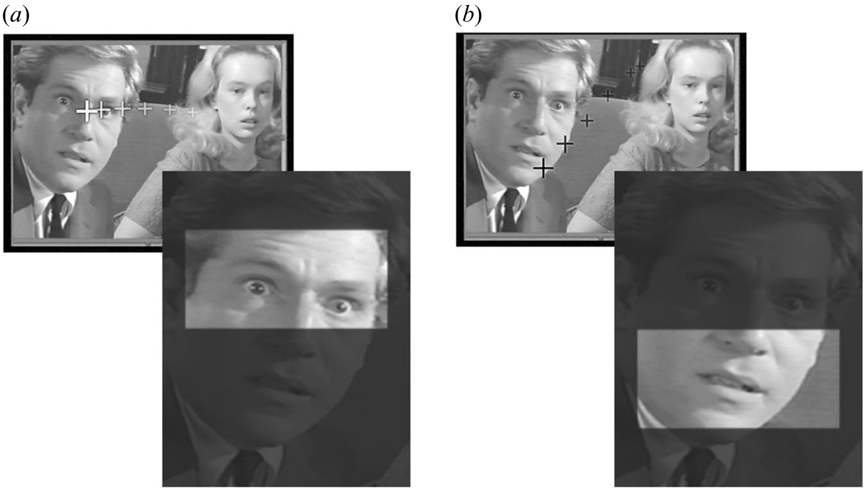

Figure 1. Focus on eyes versus mouth: cut to shocked young man. (*a*) Focus of typically developing viewer. (*b*) Focus of viewer with autism.

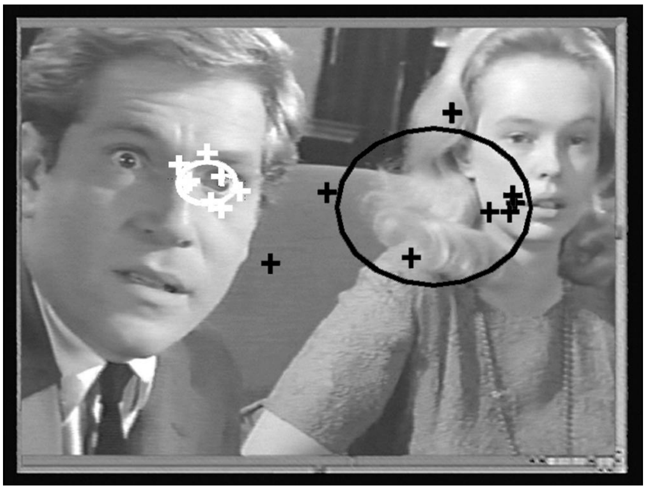

Figure 2. Group data ($n = 16$) illustrating focus on eyes versus mouth. Viewers with autism: black crosses; typically developing viewers: white crosses.

typical adult viewer matched for gender and IQ. The boldest crosshairs mark each viewer's point-of-regard at the moment of this still, while the gradated crosses reveal the path of each viewer's focus over the preceding five frames. The image in this figure is a still from a shot immediately following an abrupt camera cut. In the preceding shot, a character smashes a bottle in the right half of the frame (where both viewers were focused). The camera cuts to show the reaction of the young man and woman, and both viewers respond immediately. While the typical viewer responds directly to the look of surprise and horror in the young man's wide eyes, the viewer with autism is seen trying to gather information from the young man's mouth. The young man's mouth is slightly open but quite expressionless, and it provides few clues about what is happening in the scene.

This discrepancy in viewing patterns is also seen in group data. Figure 2 plots the focus of eight normative-IQ adults with autism (in black) and eight age-, IQ- and gender-matched typical controls (in white) (this is a subsample from the data in Klin *et al.* 2002*b*) for one frame of this video sequence. This sub-sample is used here to visually illustrate the findings obtained for the entire sample summarized below. While typical viewers converge on the eye region, some individuals with autism converge on the mouth regions, whereas others' focus is peripheral to the face. When the visual fixation patterns were summarized for the entire sample in this study ($n = 30$, 15 participants in each group), individuals with autism, relative to controls, focused twice as much time on the mouth region of faces and 2.5 times less on the eye region of faces when viewing dynamic social scenes. There was virtually no overlap in the distributions of visual fixation patterns across the two groups of participants. Figure 3 presents these data as per cent of overall viewing time focused on eyes and on mouths.

These results contrast markedly with another recent study of face scanning in autism (van der Geest *et al.* 2002), in which participants showed normative visual fixation patterns when viewing photographs of human faces relative to controls. The difference between the two studies was that while in the latter investigation participants were presented with static pictures of faces, in the former study participants were presented with dynamic (i.e., video) depictions of social interactions, coming perhaps closer to replicating a more naturalistic social situation (i.e., we almost never encounter static depictions of faces in our daily social interactions). In such more 'spontaneous situations', the deviation from normative face-scanning patterns in autism seems to be magnified. And the magnitude of this deviation is put in context if one appreciates the fact that preferential looking at the eyes rather than at the mouths of an approaching person has been shown in infants as young as three months of age (Haith *et al.* 1979).

A second example from the same eye-tracking studies (Klin *et al.* 2002*a*) focuses on a developmental skill that emerges and is fully operational by the time a child is approximately 12–14 months of age. It involves the joint-attention skill of following a pointing gesture to the target indicated by the direction of pointing (Mundy & Neal 2000). Pointing, like many other non-verbal social cues, can both modify and further specify what is said. For effective communication exchange, verbal and non-verbal

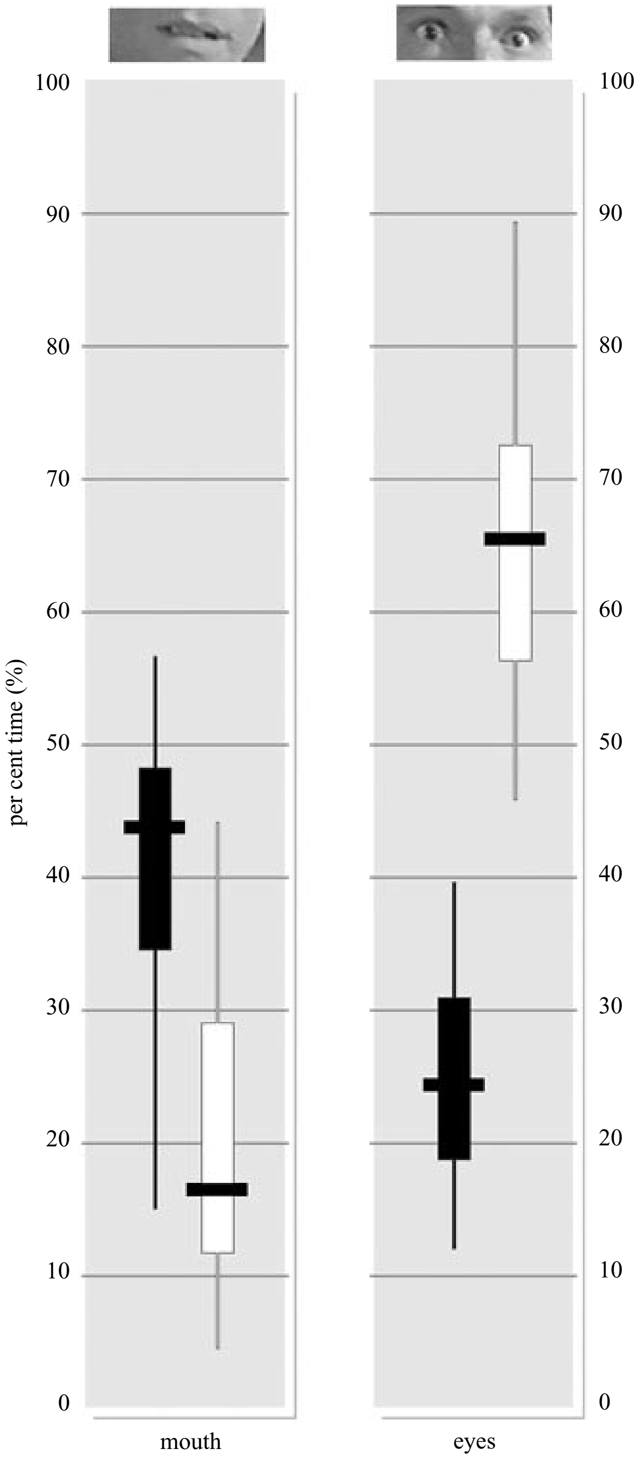

Figure 3. Box plot comparison of visual fixation time on mouth and eye regions for 15 viewers with autism and 15 typically developing viewers (controls). The upper and lower boundaries of the standard box plots are the 25th and 75th percentiles. The horizontal line across the box marks the median of the distribution and the vertical lines below and above the box extend to the minimum and maximum, respectively. Viewers with autism: areas shaded in black; typically developing viewers: white areas.

cues need to be quickly integrated. Figure 4 shows a scene from a film in which the young man enquires about a painting hanging on a distant wall. In doing so, he first points to a specific painting on the wall and then asks the older man (who lives in the house) 'Who did the painting?' While the verbal request is more general (as there are several pictures on the wall), the act of pointing has already specified the painting in which the young man is interested. The figure shows the visual scanning paths of the adult viewer with autism (in black) and the typical viewer (in white). As can be seen in figure 4*a*, and more clearly in the schematic renditions in figure 4*b*,*c*, the viewer with autism does not follow the pointing gesture but instead waits until he hears the question and then appears to move from picture to picture without knowing which one the conversation is about. The typical viewer (white track) follows the young man's pointing immediately, ending up, very deliberately, on the correct (large) picture. Hearing the question, he then looks to the older man for a reply and back to the young man for his reaction. The visual path he follows clearly illustrates his ability to use the non-verbal gesture to immediately inspect the painting referenced by the young man. By contrast, the viewer with autism uses primarily the verbal cue, neglecting the non-verbal gesture, and in doing so, resorts to a much more inefficient pursuit of the referenced painting. When the viewer with autism was later questioned, in an explicit fashion, about whether he knew what the pointing gesture meant, he had no difficulty defining the meaning of the gesture. And yet, he failed to apply this knowledge spontaneously when viewing the scene from the film.

That normative-IQ adolescents and adults with autism fail to display normative reactions exhibited by typical young children does not mean, of course, that their ability to function in the world is at this very early stage of development. Rather, it raises the possibility that these individuals learn about the social world in a different manner. What form this developmental path takes is of both clinical and research importance. Collectively, the various examples presented here suggest a need to explain the discrepancy between performance on structured and explicit, as against naturalistic and spontaneous, tasks, and in so doing, to explore what might be a unique social developmental path evidenced in autism. This paper contends that theories of the social dysfunction in autism need to address both of these phenomena. Traditionally, theories of social cognitive development have relied on a framework delineated by computational models of the mind and of the brain (Gardner 1985), which focus on abstracting problem-solving capacities necessary to function in the social environment. The methodologies used typically employ explicit and often verbally mediated tasks to probe whether or not a person has these capacities. In real life, however, social situations rarely present themselves in this fashion. Rather, the individual needs to go about defining a social task as such by paying attention to, and identifying, the relevant aspects of a social situation prior to having an opportunity to use their available social cognitive problem-solving skills. Thus, in order to study more naturalistic social adaptation, there may be some justification in using an alternative theoretical framework that centres around a different set of social cognitive phenomena, for example people's predispositions to orient to salient social stimuli, to naturally seek to impose social meaning on what they see and hear, to differentiate what is relevant from what is not, and to be intrinsically motivated to solve a social problem once such a problem is

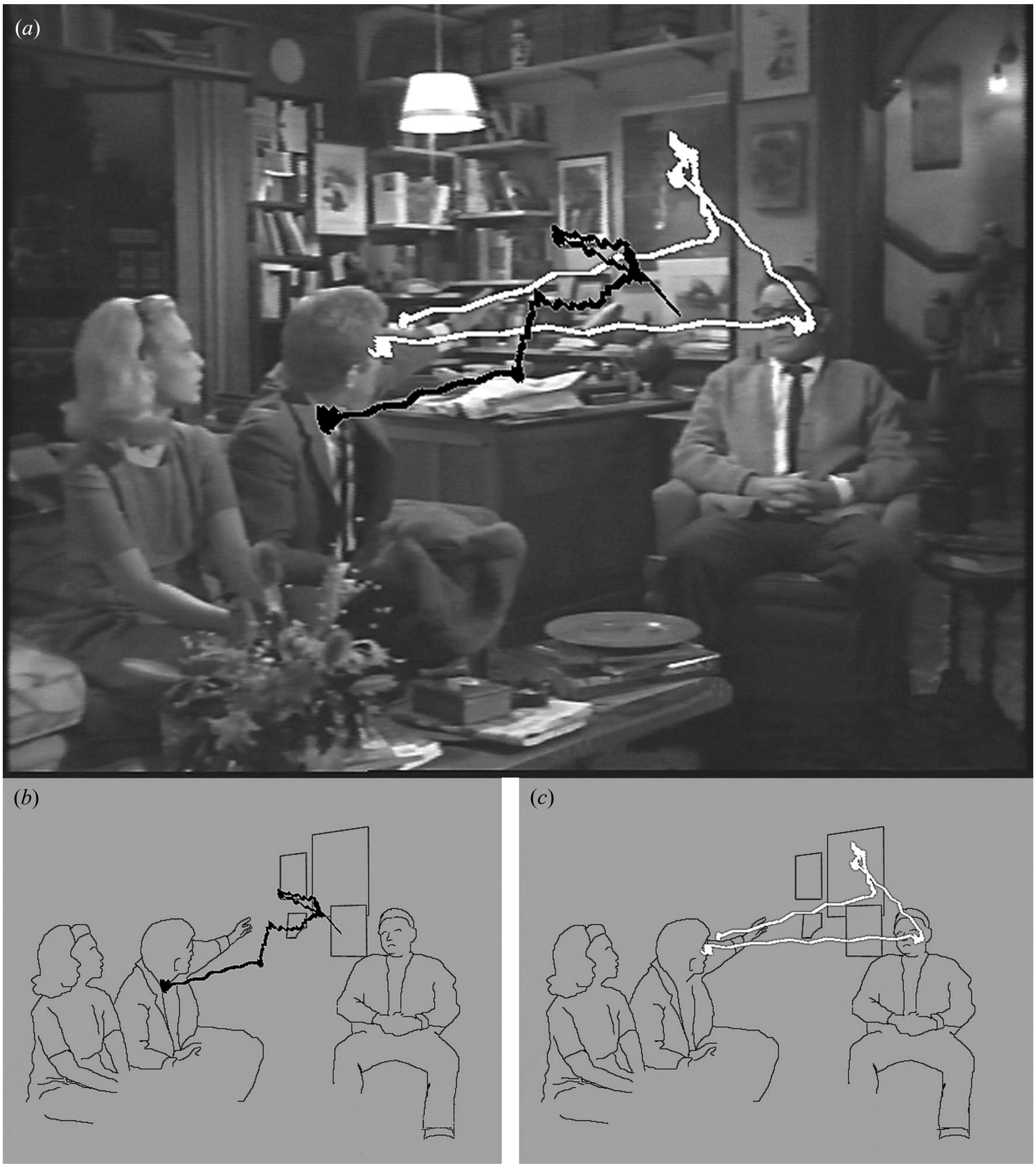

Figure 4. Scanning patterns in response to social visual versus verbal cues. Viewers with autism: black trace in (*a*) and (*b*); typically developing viewer: white trace in (*a*) and (*c*).

identified. The framework presented in this paper is called EM in order to highlight the central role of motivational predispositions to respond to social stimuli and a developmental process in which social cognition results from social action.

The emphases of the EM framework differ from those in computational models in a number of ways:

(i) instead of assuming a social environment that consists of a pre-given set of definitions and regularities, and a perceiving social agent (e.g. a child) whose mind consists of a pre-given set of cognitive capacities that can solve problems as they are explicitly presented to it, this framework proposes an active mind that sets out to make sense of the social environment and that changes itself as a result of this interaction (Mead 1924);

(ii) moving from a focus on abstracted competencies (what an organism can do), this framework focuses on the adaptive functions which are subserved by these competencies (i.e. how an agent engages in the process of acquiring such competencies in the first place) (Klin *et al.* 2000);

(iii) moving away from a focus on cognition, this framework rekindles a once more prominent role given to affect and predispositional responses in the process of socialization (Damasio 1999); and

(iv) it shifts the focus of investigation from what can be called 'disembodied cognition', or insular abstractions captured by computational cognition (e.g. algorithms in a digital computer) to 'embodied cognition', or cognitive traces left by the action of an organism upon an environment defined by species-specific regularities and by a species-specific topology of differential salience (i.e. some things in the environment are more important than others).

Of particular importance in this framework is the premise that agents may vary in what they are seeking in the environment, resulting in highly disparate 'mental representations' of the world that they are interacting with (Varela *et al.* 1991; Clark 1999). This process, in turn, leads to individual variation in neurofunctional specialization given that more prominence is given in this framework to the notion of the brain as a repository of experiences (LeDoux 2002); that is, our 'brain becomes who we are' or experience repeatedly.

Specifically, the EM approach is offered as an avenue to conceptualize phenomena deemed essential for understanding social adaptation, and which are typically not emphasized in research based on computational models of the social mind. These include the need to consider the complexity of the social world, the very early emerging nature of a multitude of social adaptive mechanisms and how these mechanisms contextualize the emergence of social cognition, as well as important temporal constraints on social adaptation. Our formulation of the EM framework is primarily based on Mead's Darwinian account of the emergence of mind (Mead 1924), the work of Searle *et al.* (1980) and Bates (1976, 1979) in respect of the underlying functions of communication, the philosophy of perception of Merleau-Ponty (1962), and, particularly, on a framework for cognitive neuroscience outlined by Varela *et al.* (1991), from which the term 'enactive mind' is borrowed. Excellent summaries of psychological and neurofunctional aspects of this framework have been provided by Clark (1999) and Iacoboni (2000*a*). Some of the views proposed here have long been part of discussions contrasting information processing and ecological approaches to every aspect of the mind, including attention and sensorimotor integration, memory and language, among other psychological faculties (Gibson 1963; Neisser 1997).

2. THE SOCIAL WORLD AS AN 'OPEN-DOMAIN TASK'

In the EM approach, a fundamental difference between explicit and naturalistic social tasks is captured in the distinction between 'closed domains' and 'open domains' of operation (Winograd & Flores 1986). Research paradigms based on computational models of the social mind often reduce the social word to a set of pre-given rules and regularities that can be symbolically represented in the mind of a young child. In other words, the social world is simplified into a 'closed domain task', in which all essential elements to be studied can be fully represented and defined. This is justified in terms of the need to reduce the complexity of the social environment into a number of easily tested problem-solving tasks. By contrast, the EM approach embraces the open-ended nature of social adaptation. The social world as an 'open domain task' implies the need to consider a multitude of elements that are more or less important depending on the context of the situation and the person's perceptions, desires, goals and ongoing adjustment. Successful adaptation requires from a person a sense of relative salience of each element in a situation, preferential choices based on priorities learned through experience, and further moment-by-moment adjustments. For example, if one were to represent the skills of driving a car successfully, one could define the 'driving domain' as involving wheels, roads, traffic lights and other cars. However, this domain is hardly complete without encompassing a host of other factors including attention to pedestrians (sometimes but not always), driving regulations (but these can be overridden by safety factors), local customs (in some cities or countries more than others), variable weather conditions, signals from other drivers, and so on. This rich texture of elements defines the 'background' of knowledge necessary to solve problems in the driving domain. Similarly, the social domain consists of people with age, gender, ethnic and individual differences, facial and bodily gestures, language and voice/prosodic cues in all of their complexity and context-dependent nature, posture, physical settings and social props, and situation-specific conventions, among a host of other factors. Successful driving, or social adaptation, would require more than knowing a set of rules—at times referred to as 'Knowing That'. Rather, it would require 'Knowing How', or a learning process that is based on the accumulation of experiences in a vast number of cases that result in being able to navigate the background environment according to the relative salience of each of the multitude of elements of a situation, and the moment-by-moment emerging patterns that result from the interaction of the various elements. In autism, one of the major limitations of available teaching strategies, including forms of social skills training (Howlin *et al.* 1999), is the difficulty in achieving generalization of skills; in other words, how to translate a problem-solving capacity learned in a closed-domain environment (e.g. therapeutic methods relying on explicit rules and drilling) into a skill that the person avails himself, or herself, of in an open-domain environment (e.g. a naturalistic social situation). This may also be the reason why individuals with autism have difficulty in spontaneously using whatever social cognitive skills they may have learned through explicit teaching. Incidentally, driving is an equally challenging task to individuals with autism.

In the EM approach, the child 'enacts the social world', perceiving it selectively in terms of what is immediately essential for social action, whereas mental representations of that individualized social world arise from repeated experiences resulting from such perceptually guided actions (Varela *et al.* 1991). In this way, the surrounding environment is reduced to perceptions that are relevant to social action; a great simplification if one is to consider the richness of what is constantly available for an agent to hear, see and otherwise experience. Similarly, the mental representations (i.e. social cognition) available for the child to reason about the social environment are deeply embedded in the child's history of social actions, thereby constituting a tool for social adaptation. Thus, there are two principles underlying the EM approach to naturalistic social situations as 'open-domain tasks'. First, the vast complexity of the surrounding environment is greatly simplified in terms of a differential 'topology of salience' that separates aspects of the environment that are irrelevant (e.g. light fixtures, a person turned away) from those that are crucially important (e.g. someone staring at you). Second, this topology of salience is established in terms of perceptually or cognitively guided actions subserving social adaptation.

Figure 5. Adult viewer with autism (white cross circled in black): focus on non-essential inanimate details.

Figure 6. Toddler viewer with autism: focus on non-essential inanimate details.

These principles imply, however, that the surrounding environment will be 'enacted' or recreated differently based on differences in predispositions to respond in a certain way (Maturana & Varela 1973). In autism, our eye-tracking illustrations are beginning to show what this social landscape may look like from the perspective of individuals with this condition. Consider, for example, the illustration in figure 5, showing the point of regard (signalled by the white cross in the centre of the black circle) of a normative-IQ adult with autism who is viewing a romantic scene. Rather than focusing on the actors in the foreground, he is foveating on the room's light-switch on the left. In figure 6, a 2-year-old boy with autism is viewing a popular American children's show. His point of regard on the video frame presented as well as his scan-path immediately before and after that frame (seen in black at the right-hand corner of the picture) indicate that rather than focusing on the protagonists of the show and their actions, this child is visually inspecting inanimate details on the shelves. By 'enacting' these scenes in this manner it is likely that, from the perspective of the two viewers with autism, the scenes are no longer social scenes, however clear their social nature might be to a typical viewer. It is also quite probable that if these viewers were explicitly asked or prompted to observe the social scenes and perform a task about them, they might be able to fare much better. The fact that they did not orient to the essential elements in the scene, however, suggests that were they to be part of such a situation, their adjustment to the environmental demands (e.g. to fit in the ongoing play taking place between the two child protagonists) would be greatly compromised.

3. DEVELOPMENTAL ELEMENTS IN THE EMERGENCE OF MENTAL REPRESENTATIONS

Computational models of the social mind make use of cognitive constructs that could help a child successfully navigate the social environment (Baron-Cohen 1995). There is less emphasis on how these constructs emerge within a broader context of early social development, which is a justifiable way of modelling the more specific, targeted social cognitive skills. By contrast, the EM approach depends on this broader discussion of early social predispositions to justify the need to consider complex social situations in terms of a differential 'topology of salience'. In other words, why should some aspects of the environment be more salient than others? In order to address this question, there is a need to outline a set of early social reactions that may precede and accompany the emergence of social cognitive skills.

In the EM approach, the perceptual make-up of typical human infants is seen as consisting of a specific set of somatosensory organs that are constantly seeking salient aspects of the world to focus on, particularly those that have survival value. To invoke the notion of survival value implies the notion of adjustment to, or action upon, the environment. In this context, the gravitation towards and engagement of conspecifics is seen as one of the important survival functions. Thus, social stimuli are seen as having a higher degree of salience than competing inanimate stimuli (Bates 1979; Klin *et al.* 2000). The possibility that, in autism, the relative salience of social stimuli might be diminished (Klin 1989; Dawson *et al.* 1998) could be the basis for a cascade of developmental events in which a child with this condition fails to enact a relevant social world, thus failing to accrue the social experiences suggested in the EM approach to be the basis for social cognitive development.

A large number of social predispositions have been documented in the child development literature, some of which appear to be greatly reduced in children with autism. To limit the discussion to early social orientation skills, we consider only infants' reactions to human sounds and faces. The human voice appears to be one of the earliest and most effective stimuli conducive of social engagement (Eimas *et al.* 1971; Mills & Melhuish 1974; Alegria & Noirot 1978; Eisenberg & Marmarou 1981), a reaction that is not observed in autism (Adrien *et al.* 1991; Klin 1991, 1992; Osterling & Dawson 1994; Werner *et al.* 2000). In fact, the lack of orientation to human sounds (e.g. when the infant hears the voice of a nearby adult) has been found to be one of the most robust predictors of a later diagnosis of autism in children first seen at the age of 2 years (Lord 1995). In the visual modality, human faces have been emphasized as one of the most potent facilitators of social engagement (Bryant 1991). For example, 2-day-olds look at their mother rather than at another unknown woman (Bushnell *et al.* 1989). Three-month-olds focus on the more emotionally revealing eye regions of the face (Haith *et al.* 1979), and 5-month-olds are sensitive to very small deviations in eye gaze during social interactions (Symons *et al.* 1998) and can match facial and vocal expressions based on congruity (Walker 1982). In autism, a large number of face perception studies have shown deficits and abnormalities in such basic visual social processing situations (Langdell 1978; Hobson *et al.* 1988; Klin *et al.* 1999) which, incidentally, were not accompanied by failure in developmentally equivalent tasks in the physical (non-social) domain. For example, one study demonstrated adequate visual processing of buildings as against faces (Boucher & Lewis 1992). Another study asked children with autism to sort people who varied in terms of age, sex, facial expressions of emotion and the type of hat that they were wearing (Weeks & Hobson 1987). In contrast to typical children who grouped pictures by emotional expressions, the participants with autism grouped the pictures by the type of hat the people were wearing. Such studies indicated not only abnormalities in face processing but also preferential orientation to inanimate objects, a finding corroborated in other studies (Dawson *et al.* 1998). In a more recent study (Dawson *et al.* 2002), children with autism failed to exhibit differential brain event-related potentials to familiar versus unfamiliar faces, but they did show differences relative to familiar versus unfamiliar objects.

While computational models of the social mind are often modular in nature (Leslie 1987), that is, certain aspects of social functioning could be preserved while others were disrupted, the EM approach ascribes importance to early disruptions in sociability because of its central premise that normative social cognition is embedded in social perception and experience. This principle states that social perception is perceptually guided social action, and social cognitive processes emerge only from recurrent sensorimotor patterns that allow action to be perceptually guided (hence the notion of 'embodied cognition'; Varela *et al.* 1991). The radical assumption of this framework, therefore, is that it is not possible to disentangle cognition from actions, and that if this happened (e.g. a child was taught to perform a social cognitive task following an explicit drill rather than acquiring the skill as a result of repeated social engagement and actions), the given skill would represent a 'disembodied cognition', or a reasoning skill that would not retain its normative functional value in social adaptation (Markman & Dietrich 2000). For example, an infant may be attracted to the face of his mother, seeking to act upon it, and in the context of acting upon it the infant learns a great deal about faces and mothers, although this knowledge is a function of the child's active experiences with that face. These experiences may include learning of contingencies (e.g. vocal sounds and lip movements go together; certain voice inflections go with certain face configurations such as smiles and frowns), and that these contingencies have pleasurable value (thus leading to approach or an attempt at re-enactment of the situation) or unpleasurable value (thus leading to withdrawal). Studies of infants' early social development have shown not only that they are sensitive to affective salience, but that they also act upon that salience through reactions that are appropriate to emotional signals (Haviland & Lelwica 1987). They react negatively to their mothers' depressed affect (Tronick *et al.* 1986), and appropriately to the emotional content of praise or prohibition (Fernald 1993). From an early age, they expect contingency between their actions and those of their partners (Tarabulsy *et al.* 1996). Fewer developmental phenomena have demonstrated this effect more clearly than studies using the 'still-face paradigm' (Tronick *et al.* 1978). When mothers, who have previously been stimulating their babies in a playful fashion, withdraw the smiles and vocalization and assume a still-face, infants as young as 2–3 months old first make attempts to continue the interaction but then stop smiling, avert their gaze, and may protest vigorously (Field *et al.* 1986; Gusella *et al.* 1988). One study of the still-face effect involving children with autism has failed to document this normative pattern of response (Nadel *et al.* 2000).

In summary, in the EM approach early social predispositions are thought to create the basis and the impetus for the subsequent emergence of mental representations that, because of their inseparability from social action (i.e. they are 'embodied'), retain their adaptive value. Infants do not build veridical models of the social world based on 'universals' or context-invariant representations. Rather, their models or expectations of the world follow their salience-guided actions upon an ever-changing environment that needs to be coped with in an adaptive, moment-by-moment and context-dependent manner (Engel *et al.* 2001).

4. CONTEXTUAL ELEMENTS IN THE EMERGENCE OF MENTAL REPRESENTATIONS

The classical computational model in cognitive science assumes that cognitive processes are rule-based manipulations of symbols representing the external environment (Newell 1991). Similarly, computational models of the social mind build on the notion that to operate socially is to execute algorithms involving mental representations (Baron-Cohen 1994). By contrast, the EM approach raises the non-trivial question of how a representation acquires meaning to a given child, the so-called 'mind–mind problem' (Jackendoff 1987). The question is, what is the relationship between computational states (e.g. manipulation of mental representations) and a person's experience of the real-life referent of the computational state? How do we progress from having a representation of a person's intention, to experiencing that intention by reacting to it in a certain way? In the computer world, we do know where the meaning of the computational algorithms comes from, namely the programmer. But how do mental representations acquire meaning to a developing child? In autism, individuals often acquire a large number of symbols and symbolic computations that are devoid of shared meaning with others, i.e. the symbols do not have the meaning to them that they have to typical children. Examples are: (i) hyperlexia (reading decoding skills go unaccompanied by reading comprehension; Grigorenko *et al.* 2002); (ii) echolalia and echopraxia (echoing of sounds or mimicry of movements; Prizant & Duchan 1981; Rogers 1999); (iii) 'metaphoric language' (e.g. neologisms, words used in idiosyncratic ways; Lord & Paul 1997); and (iv) prompt-dependent social gestures, routines or scripts (e.g. waving bye-bye without eye contact, staring when requested to make eye contact), among others. While it is difficult for one to conceive of a dissociation between knowing a symbol and acting upon it (e.g. knowing what is the meaning of the pointing gesture and spontaneously turning one's head when somebody is pointing somewhere), this actually happens in autism, as shown in figure 3 and in the other examples given above. We know that children with autism can learn associatively (e.g. a symbol becomes paired with a referent). This happens, for example, in vocabulary instruction using simple behavioural techniques. However, one of the big challenges for these children is often to pair a symbol with the adaptive action subsumed by the symbol (Wetherby *et al.* 2000).

In the EM approach, symbols or cognition in general have meaning to the child using them because they are 'embodied actions' (Johnson 1987; Clark 1999), meaning that 'cognition depends upon the experiences that come from having a body with various sensorimotor capacities', and that ' perception and action are fundamentally inseparable in lived cognition' (Varela *et al.* 1991, p. 173). An artificial separation of cognition from the other elements would render the given cognitive construct a 'mental ghost' once again. One can exemplify the inseparability of cognition and action through the classic studies of Held & Hein (1963) and Held (1965) of perceptual guidance of action. They raised kittens in the dark and exposed them to light only under controlled conditions. One group of kittens was allowed to move around normally, but each of them was harnessed to a carriage that contained a second group of kittens. While the groups shared the same visual experience, the second group was entirely passive. When the kittens were released after a few weeks of this treatment, members of the first group (the one that moved around) behaved normally, whereas members of the second group (the one that was passively carried by the others) behaved as blind, bumping into objects and falling off edges. These experiments illustrate the point that meaningful cognition of objects (i.e. the way we see them and adjust to them) cannot be formed by means of visual extraction alone; rather, there is a need for perceptual processes to be actively linked with action in order to guide further action upon these objects. Studies of adaptation of disarranged hand–eye coordination in humans (Held & Hein 1958), tactile vision substitution in blind humans (Bach-y-Rita 1983) and neural coding of body schema in primates (Iriki *et al.* 1996) among others (see Iacoboni 2000*b*) support this point. A striking example is provided in a study (Aglioti *et al.* 1996) of a patient with right-brain damage who denied the ownership of her left hand and of objects that were worn by her left hand (such as rings). When the same objects were worn by the right hand, the patient recognized them as her own. In infancy research, a wide range of phenomena, from haptic and depth perception (Bushnell & Boudreau 1993) to Piagetian milestones (Thelen *et al.* 2001) have began to characterize developmental skills as 'perception-for-action' systems, while neuroimaging studies have shown overlapping brain circuitry subserving action observation and action generation (Blakemore & Decety 2001).

Perception-for-action systems are particularly relevant to a discussion of social adaptation. Consider the skill of imitation, one of the major deficits in autism (Rogers 1999). It is interesting that while children with autism have great difficulty in learning through imitation, they do exhibit a great deal of 'mirroring' or 'copying' behaviours, both vocally (e.g. echoing what other people say) and motorically (e.g. making the same gesture as another person). However, these are typically devoid of the function that these behaviours serve to typical people displaying them. One theory derived from the EM approach would predict that this curious discrepancy originates from the aspect of the typical person's action that is most salient in the child's perception. Whereas typical children may see a waving gesture as a motion embedded in the act of communication or emotional exchange, children with autism may dissociate the motion from the social context, focusing on the salient physical facts and thus repeating the gesture in a mechanical fashion. This is not unlike what a typical child might do in a game of imitating meaningless gestures, or what a neonate might do when protruding his or her tongue in response to seeing an adult doing so (Meltzoff & Moore 1977). This theory originates from the notion that while perception-for-action may occur in the absence of social engagement (e.g. in neonates), in typical infants, around the middle of their second year of life, imitation is much more likely to serve social engagement and social learning than to occur outside the realm of social interaction, as in autism. Supporting this theory is a series of studies in which, for example, 18-month-old infants were exposed to a human or to a mechanical device attempting to perform various actions. The children imitated the action when it was performed by the human model, but not when it was performed by the mechanical device (Meltzoff 1995).

Figure 7. Series of static images of the human form rendered as point-light displays.

Perception-for-action systems are of particular interest in the context of survival abilities (e.g. responding to a threatening person or a lethal predator). A central example of such systems is the ability to perceive certain patterns of movement as biological motion. This system allows humans, as well as other species, to discern the motion of biological forms from motion occurring in the inanimate environment. In the wild, an animal's survival would depend on its ability to detect approaching predators and predict their future actions. In humans, this system has been linked to the emergence of the capacity to attribute intentions to others (Frith & Frith 1999). The study of biological motion has traditionally used the paradigm of human motion display created by Johansson (1973). In his work, the motion of the living body is represented by a few bright spots describing the motions of the main joints. In this fashion, the motion pattern is dissociated from the form of people's bodies. The moving presentation of this set of bright spots evokes a compelling impression of basic human movements (e.g. walking, running, dancing) as well as of social movements (e.g. approaching, fighting, embracing). Figure 7 illustrates a series of static images of the human form rendered as point-light animations. The phenomenon studied by

Johansson, however, can only be fully appreciated when the display is set in motion.

Using this paradigm, several studies have documented adult's abilities to attribute gender, emotions and even personality features to these moving dots (Koslowski & Cutting 1978; Dittrich *et al.* 1996). Even 3-month-old infants are able to discriminate between the moving dots depicting a walking person and the same dot displays moving randomly (Fox & McDaniel 1982). The presence of this ability at such a young age, as well as its presence in other species including monkeys (Oram & Perrett 1994) and birds (Regolin *et al.* 2000), and the demonstrated singularity of biological motion relative to other forms of motion from the perspective of the visual system (Neri *et al.* 1998) suggest that this is a highly conserved and unique system that makes possible the recognition of movements of others in order to move towards or away from them. Several neuroimaging studies have singled out the superior temporal sulcus as an important structure involved in the perception of biological motion (Grossman & Blake 2000; Grossman *et al.* 2000; Grezes *et al.* 2001), a region also associated with basic 'survival' reactions such as evaluating facial expressions and/or direction of eye gaze (Puce *et al.* 1998). A positron emission tomography study attempting to separate decontextualized human motions (point-light displays depicting a hand bringing a cup to one's mouth) from what can be seen as a more naturalistic human motion (a person dancing) showed that the perception of the latter also implicated limbic structures such as the amygdala (Bonda *et al.* 1996). This finding is consistent with a perception-for-action system that not only perceives to act, but one that is embedded in an approach/withdrawal, affective-based context (Gaffan *et al.* 1988).

Given the fundamental and adaptive nature of perception of biological motion, one would expect this system to be intact in even very disabled children. One study so far has shown the system to be intact in children with profound spatial deficits and a degree of mental retardation (Jordan *et al.* 2002). By contrast, our own preliminary data suggest, to date, that this system may be compromised in young children with autism. We used Johansson point-light displays to depict a series of social approaches that are part of the typical experiences of young children (e.g. an animated adult trying to attract the attention of a young toddler, 'pat-a-cake', 'peek-a-boo'). Scenes were presented in two formats simultaneously, one on each of the two horizontal halves of a computer screen. The scenes were identical except that one was oriented correctly and the other was upside-down. The child heard the corresponding sound effects of that social scene (e.g. the verbal approach of an adult). The experiment followed a visual preference paradigm in which the child looked at one of the two scenes presented. By requiring the child to choose between an upside-down and a correctly oriented animation matching the sound effects of the social interaction, we were able to test the child's ability to impose mental representations of human movement interactions on the ambiguous visual stimuli. This paradigm is illustrated in figure 8. Our preliminary data for 11 2-year-old toddlers, 5 with a diagnosis of autism and 6 typical children is given in figure 9. Overall, the typically developing toddlers demonstrated a marked preference for the correctly ori-

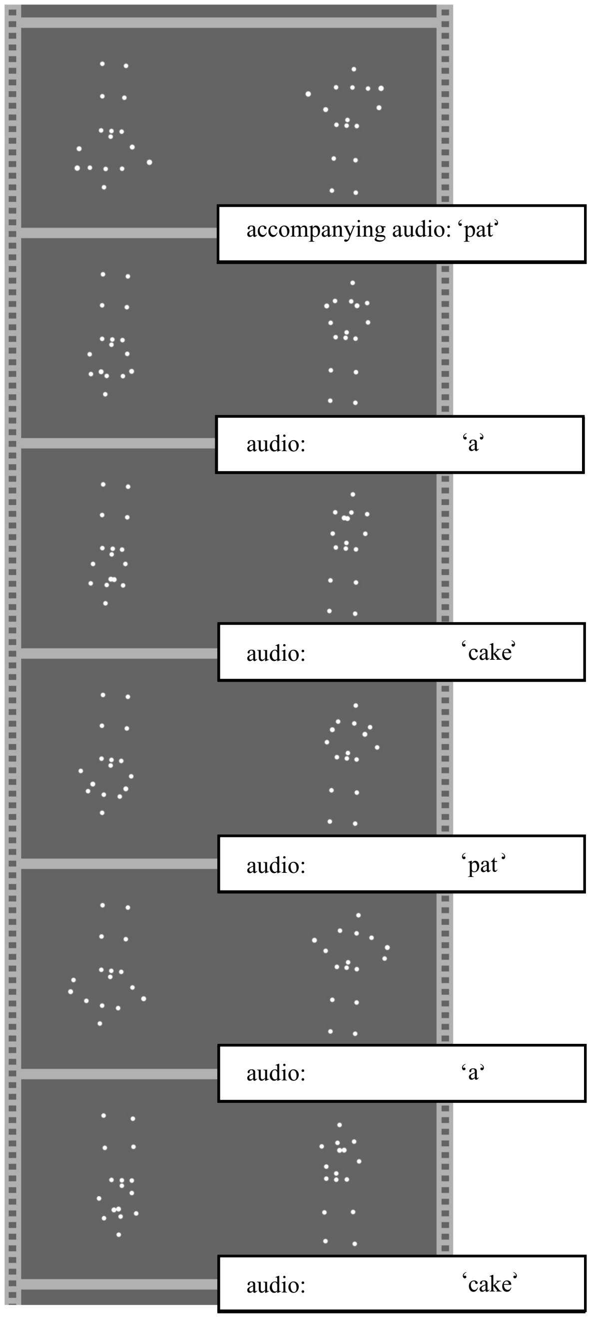

Figure 8. Cross-modal matching task with social animation stimuli.

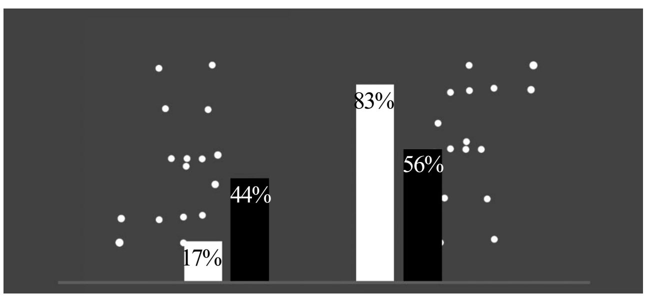

Figure 9. Percentage of total viewing time spent on upright versus inverted figures. Black bars: toddlers with autism; white bars: typically developing toddlers.

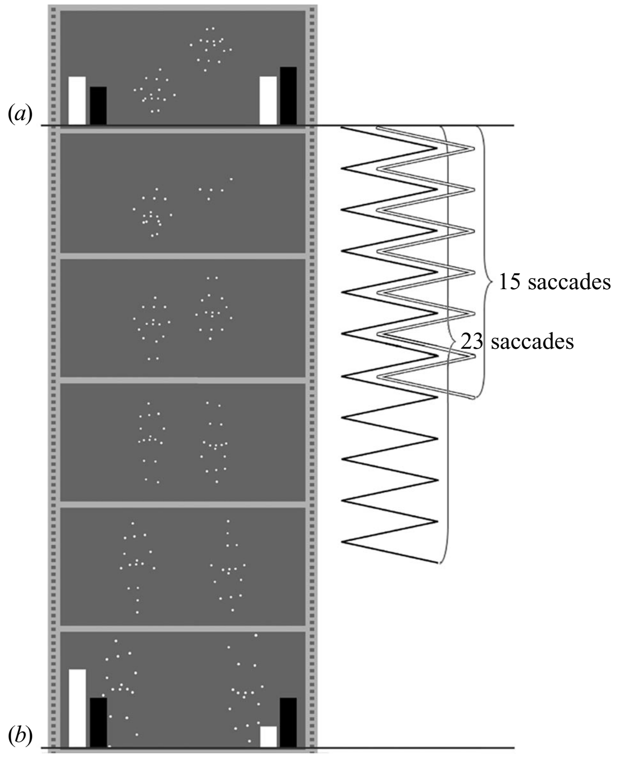

Figure 10. Initial and final fixation data, and number of saccades between upright and inverted figures. Toddlers with autism, filled bars; typically developing toddlers, open bars. (*a*) Initial fixation: toddlers with autism 40% upright, 60% inverted; typically developing toddlers 50% upright, 50% inverted. (*b*) Final fixation: toddlers with autism 50% upright, 50% inverted; typically developing toddlers 79% upright, 21% inverted. Number of saccades between upright and inverted figures: toddlers with autism 23 saccades min^{-1}, typically developing toddlers 15 saccades min^{-1}.

ented figure (83% of total viewing time versus 17% for upside-down display), while the toddlers with autism showed a pattern closer to a random choice (56% versus 44%). We also analysed initial fixations and final fixations (defined by the figure the child was focusing on at the end of the animation) as a rudimentary view of how understanding of the animation's content might progress during viewing. We recorded the number of times the toddlers with autism shifted their focus from the upright to the inverted figure, relative to typically developing controls. These results are depicted graphically in figure 10. While typically developing toddlers and toddlers with autism both exhibited initial fixations at chance or near-chance levels, the typically developing infants were focused on the upright figure at the end of more than 75% of all trials, while the toddlers with autism remained at chance level. Of similar interest are group differences in the pattern of shifting between the upright and inverted figures. Toddlers with autism shifted more frequently than typically developing toddlers, a trend suggestive of increased difficulty in adequately understanding either of the two displays. If corroborated in larger studies, this finding would point to a major disruption in a highly conserved skill that is thought to be a core ability underlying social engagement and, subsequently, the capacity to attribute intentionality to others.

5. TEMPORAL CONSTRAINTS ON MODELS OF SOCIAL ADAPTATION

Computational models of the mind place less emphasis on the temporal unfolding of the cognitive processes involved in a task (Newell 1991). This stance is justified when a given task is explicit and fully defined. However, in naturalistic situations there are important temporal constraints in social adaptation, as failure to detect an important but fleeting social cue, or a failure to detect temporal relationships between two social cues, may lead to partial or even misleading comprehension of the situation, which may in turn lead to ineffective adjustment to the situation. For example, if the viewer of a scene fails to monitor a non-speaker in a social scene who is clearly embarrassed by what another person is saying, the viewer is unlikely to correctly identify the meaning of that situation (Klin *et al.* 2002*a*). In this way, the EM approach sees social adaptation along the same principles currently being considered in research into 'embodied vision' (Churchland *et al.* 1994). This view holds that the task of the visual system is not to generate exhaustive mental models of a veridical surrounding environment but to use visual information to perform real-time, real-life adaptive reactions. Rather than creating an inner mirror of the outside world to formulate problems and then to solve them ahead of acting upon them, vision is seen as the active retrieval of useful information as it is needed from the constantly present and complex visual environment. From the organism's adaptive perspective, the topology of salience of this visual tapestry, from light reflections to carpet patterning, to furniture and clothing, to mouths and eyes, is far from flat. We would be overwhelmed and paralysed by its richness if we were to start from a position of equal salience to every aspect of what is available to be visually inspected. Rather, we actively retrieve aspects of the visual environment that are essential for quick, adaptive actions by foveating on sequential locations where we expect to find them. These 'expectations' are generated by a brain system dedicated to salience (a lion entering the room is more important than the light-switch next to the door), and an ever more complex (going from infancy to adulthood) understanding of the context of the situation, the so-called 'top-down' approach to vision (Engel *et al.* 2001).

A pertinent example of this view of vision is the analysis of a baseball game by Clark (1999, p. 346) in which an outfielder positions himself or herself to catch a fly ball: 'It used to be thought that this problem required complex calculations of the arc, acceleration and distance of the ball. However, more recent work suggests a computationally simpler strategy (McBeath *et al.* 1995). Put simply, the fielder continually adjusts his or her run so that the ball never seems to curve towards the ground, but instead appears to move in a straight line in his or her visual field. By maintaining this strategy, the fielder should be guaranteed to arrive in the right place at the right time to catch the ball' (p. 346). Piaget (1973) provided similar examples from children's play, and Zajonc (1980) provided similar examples from intersubjective adaptation. Consistent with these examples, the EM approach considers the 'social game' to be not unlike the outfielder's effort. A typical toddler entering a playroom pursues a sequence of social adaptive reactions to split-second environmental demands with moment-by-moment disregard of the vast majority of the available visual stimulation. Such a child is ready to play the social game. For individuals with autism, however, the topology of salience, defined as the 'foveal elicitation' of socially relevant stimuli (as exemplified in our eye-tracking illustrations and in studies of preferential attention to social versus non-social entities; see above), is much flatter. The social worlds enacted by individuals with autism and by their typical peers, if viewed in this light, may be strikingly different.

6. SOCIAL COGNITION AS SOCIAL ACTION

The radical assumption made in the EM approach is that mental representations as described in computational models of the mind are proxies for the actions that generated them and for which they stand (Varela *et al.* 1991; Thelen & Smith 1994; Lakoff & Johnson 1999). This counter-intuitive view can be traced back to the account of Mead (1924) of the social origins of mind. Mead saw the emergence of mind as the capacity of an individual to make a 'gesture' (e.g. bodily sign, vocal sound) that means to the other person seeing or hearing it the same as for the person making it. The meaning of the gesture, however, is in the reaction of the other. A gesture used in this way becomes a symbol, i.e. something that stands for the predicted reaction of the other person. Once a child has such a symbolic gesture, she can then uphold it as a representation for the reaction of the social partner, thus being able to take a step back from the immediate experience and then to contemplate alternatives of action using such symbols as proxies for real actions. In the EM approach, the fact that the emergence and evolution of a symbol are tied to actions of adaptation, which in turn are immersed in a context of somatosensory experiences, salience and perceptually guided actions, makes the symbol a proxy for these elements of the action. When we uphold and manipulate symbols in our mind, therefore, we are also evoking a network of experiences resulting from a life history of actions associated with that symbol.

This view, connecting social cognition with social action, is useful in our attempt to explore possible reasons why accomplishments in social reasoning in individuals with autism are not accompanied by commensurate success in social action. Consider an example from research on face perception. While face recognition deficits are very pronounced in young children with autism (Klin *et al.* 1999), the size of this deficit is much smaller in older and more cognitively able adolescents (Celani *et al.* 1999). The possibility that older individuals might perform such tasks using atypical strategies relative to their peers was investigated in our recent fMRI study of face recognition in autism (Schultz *et al.* 2000) in which normative-IQ individuals with autism and controls were presented with face versus object recognition tasks. In contrast to controls for whom face processing was associated with FG activation, in individuals with autism face processing was associated with activation in inferior temporal gyrus structures, an activation pattern that was obtained for controls when they were processing objects. These results indicated that individuals with autism did not rely on the normal neural substrate during face perception (Kanwisher *et al.* 1997) but rather engaged brain areas that were more important to non-face, object processing (Haxby *et al.* 1999). In other words, they failed to treat faces as a special form of visual stimulus, treating them instead as ordinary objects.

It would be tempting from these results to suggest that a circumscribed area of the brain, namely the FG, and the mechanism it represents, namely perception of face identity, were causally related to autism. Given the centrality of face perception in interpersonal interactions, this would be a plausible theory of autism. However, other recent studies (Gauthier & Tarr 1997; Gauthier *et al.* 1999) have suggested that the FG is not necessarily the brain site for face recognition, appearing instead to be a site associated with visual expertise, so that when a person becomes an expert on a given object category (say Persian carpets), selective activation of the FG occurs when the person is looking at an instance of that object. This notion suggests a reinterpretation of our face recognition results in autism. The FG was not selectively activated when individuals with autism were looking at faces because they were not experts on faces. By contrast, typically developing individuals have a lifetime to develop this expertise, a result of a very large number of recurrent experiences of focusing on and acting upon other people's faces beginning in very early infancy. As previously described, faces have little salience to young children with autism and would thus represent a much less frequent target of recurrent actions necessary to produce expertise.

If this interpretation is correct, if individuals with autism were to be asked to perform a visual recognition task using stimuli on which they had expertise, one might observe FG activation. Preliminary results supportive of this suggestion were obtained in an fMRI study of an individual with autism whose expertise area is *Digimon* characters (a large series of cartoon figures) (Grelotti *et al.* 2003). Interestingly, fMRI activations for *Digimon* characters in this individual with autism also included the amygdala, suggesting salience-driven rewards associated with the characters. Results such as these are beginning to delineate a developmental profile of functional brain maturation in autism in which hardwired social salience systems are derailed from very early on, following a path marked by seeking physical entities (not people) and repeatedly enacting them and thus neglecting social experiences (Klin *et al.* 2002*a*). This proposal is consistent with the notion of functional brain development as 'an activity-dependent process' that emphasizes the infancy period as a window of maximal plasticity (Johnson 2001). An interesting line of research supporting this theory is the case of people with a period of visual deprivation early in postnatal life due to bilateral congenital cataracts. Although early surgical correction was associated with rapid improvement of visual acuity, deficits in configural processing of faces remained even after many years post-surgery (Maurer *et al.* 1999; Le Grand *et al.* 2001). Configural processing of a class of visual stimuli (say, faces) represents a developmental shift from processing an object from its parts to processing objects in a Gestalt manner (Tanaka *et al.* 1998), which, in turn, is a mark of the acquisition of perceptual expertise (Diamond & Carey 1986; Gauthier & Nelson 2001). Thus, studies of early visual deprivation seem to highlight the effects of reduced early 'visual enactment' of a class of visual stimuli on later, automatic, and more efficient ways of processing that class of stimuli.

Returning to the fMRI example in which individuals with autism treated faces as objects (Schultz *et al.* 2000), it is of considerable interest that all participants could perform relatively well on the behavioural task of face recognition. They could correctly match faces, albeit using a strategy that differed markedly from controls. Thus, an analysis of results on the behavioural task by itself would have unveiled no significant differences between the two groups. One may, however, consider what would be the behavioural impact of failing to process faces as a special class of objects. Most people are able to recognize possibly thousands of faces very quickly, whereas their ability to recognize, say, pieces of luggage is much more limited. Thus, some of us are quite likely to mistake our bags when coming to pick them up from a luggage carousel at the airport, but we are very unlikely to mistake our mother-in-law rushing to greet us from the surrounding crowd.

The point illustrated in this example is the importance of developmental and contextual aspects of social development in making social cognitive accomplishments into tools of social action. Temporal constraints on social adaptation require skills to be displayed spontaneously and quickly, without the need for an explicit translation of what requirements are to be met in a given social task. There is a need to seek socially relevant information, and to maintain on-line, as it were, a continuous process of imposing social meaning to what is seen. This comes easily and effortlessly to typical individuals. By contrast, the most challenging task in the daily lives of individuals with autism involves the need to adjust to commonplace, naturalistic social situations. Consider, for example, an adolescent with autism entering a high school cafeteria. There is usually an array of interrelated social events taking place, each one consisting of a vast amount of social cues including language exchange, voice/prosody cues, facial and bodily gestures, posture and body movements, among many others. These cues are embedded in a complex visual and auditory setting, with some physical stimuli being relevant to the social events (i.e. representing specific social contexts—a cafeteria—or specific 'props'—a costume worn by one of the students), and other physical stimuli being entirely irrelevant (e.g. light switches or fixtures, number of doors, detailing in the walls). Such situations are so challenging because there is hardly any aspect of the social event that is explicitly defined. Faced with a highly complex and ambiguous social display that demands a reaction (e.g. where to sit down, how to insert oneself in an unfolding social event), they need to make sense of what they see and hear by imposing social meaning onto essential social aspects of the situation (e.g. facial expressions) while ignoring irrelevant stimuli (e.g. light fixtures).

In order to study how difficult it might be for individuals to make sense of such a situation, one can use an experimental metaphor that measures a person's spontaneous tendency to impose social meaning on ambiguous visual stimuli. More specifically, it measures how salient the social meaning of an array of ambiguous visual stimuli is to a viewer, and how socially relevant the viewer's thinking is when making an effort to make sense of the presented visual stimuli. The paradigm involves the presentation of a classic animation in which geometric shapes move and act like humans (Heider & Simmel 1944; figure 11). Typical viewers immediately recognize the social nature of the cartoon, and provide narratives that include a number of social attributions involving relationships portrayed there (e.g. being a bully, being a friend), the meaning of specific actions (e.g. trapping, protecting), and attributions of mental states (e.g. being shy, thinking, being surprised) to the characters. By contrast, cognitively able adolescents and adults with autism have great difficulty in doing so. In one study (Klin 2000), they were, on average, able to recognize only a quarter of the social elements deemed essential to understanding the plot of the story. A large proportion of them limited their narratives to faithful descriptions of the geometric events depicted in the cartoon, but without any social attributions. This was quite surprising considering that an inclusionary condition in this study required participants to 'pass' a relatively advanced social reasoning task (a second-order theory of mind task; Tager-Flusberg & Sullivan 1994). Thus, these individuals' ability to solve explicit social cognitive problems was no assurance that they would use these skills spontaneously. Some of them were unable to make any social attribution at all. Yet, such spontaneous attributions of intentionality to these geometric cartoons have been documented in infants (Gergely *et al.* 1995), and even primates (Uller & Nichols 2000). Some of the individuals with autism did, however, make a meaningful effort to make sense of the cartoon, but in doing so provided entirely irrelevant attributions, explaining the movements of the geometric shapes in terms of physical meaning (e.g. magnetic forces), not social meaning. Translated into a task of social adjustment to a naturalistic setting such as the high school cafeteria, the results of this study would suggest that some of these individuals might have no access to the social cues (not even noticing them), whereas others might search for causation relationships in the wrong domain, namely physical rather than social.

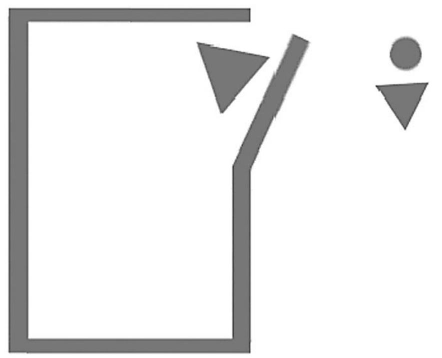

Figure 11. Screen shot showing cast of characters from a cartoon from Heider & Simmel (1944).

To impose social meaning on an array of visual stimuli is an adaptive reaction displayed by typical children, from infancy onwards, at an ever-increasing level of complexity. This spontaneous skill is cultivated in countless hours of recurrent social engagement. From discerning the meaning of facial expressions and detecting human motion and forms of human action, to attributing intentionality and elaborate mental states to others, the act of adjusting to social demands imbues social cognitive accomplishments with their functional value. It is in this light that the above examples suggest that in autism there is a breakdown in the process through which social cognitive skills and social action become inseparable.

7. CONCLUSIONS

This paper began with an intriguing puzzle posed by normative-IQ individuals with autism: how can they learn so much about the world and yet still be unable to translate this knowledge into real-life social adaptive actions? A framework different from the prevailing computational models of social cognitive development was offered—EM—as a way of exploring this puzzle. This framework is based on the emerging embodied cognitive neuroscience. EM views cognition as embedded in experiences resulting from a body's actions upon salient aspects of its surrounding environment. Social cognition is seen as the experiences associated with a special form of action, namely social interaction. These are tools of social adaptation that can be abstracted in the form of symbols and used to reason about social phenomena, although they retain their direct connection to the composite of enactive experiences that originated and shaped them over the lifetime of the child.

In autism, the EM approach proposes the theory that the above process is derailed from its incipience, because the typical overriding salience of social stimuli is not present. In its place is a range of physical stimuli, which attracts the child's selective attention, leading into a path of ever greater specialization in things rather than people. Clearly, individuals with autism are capable of acquiring language and concepts, and even a vast body of information on people. But these tools of thought are acquired outside the realm of active social engagement and the embodied experiences predicated by them. In a way, they possess what is, typically, the rooftop of social development. However, this rooftop is freestanding. The constructs and definitions are there, but their foundational experiences are not. The EM approach contends that without the set of embodied social cognitive tools required to produce moment-by-moment social adaptive reactions in naturalistic social situations, social behaviour becomes truncated, slow and inefficient.

A corollary of this theory is that individuals with autism learn about people in a way that departs from the normative processes of social development. The fact that cognitively able individuals with autism are able to demonstrate so much social cognitive understanding in some situations is as interesting as the fact that they fail to make use of these skills in other situations. The study of possible compensatory paths and the degrees to which they help these individuals to achieve more independence is as important a research endeavour as to document their social cognitive failures, but to do so there will be a need to go beyond results on explicit tasks. There will be a need both to explore more deeply the atypical processes used by these individuals to perform explicit tasks, and to increase our arsenal of methodologies capable of studying social adaptation in more naturalistic settings (Klin *et al.* 2002*a*).

REFERENCES

Adrien, J., Faure, M., Perrot, A., Hameury, L., Garreau, B., Barthelemy, C. & Sauvage, D. 1991 Autism and family home movies: preliminary findings. *J. Autism Devl Disorders* **21**, 43–49.

Aglioti, S., Smania, N., Manfredi, M. & Berlucchi, G. 1996 Disownership of left hand and objects related to it in a patient with right brain damage. *NeuroReport* **8**, 293–296.

Alegria, J. & Noirot, E. 1978 Neonate orientation behavior towards human voice. *Int. J. Behav. Dev.* **1**, 291–312.

Bach-y-Rita, P. 1983 Tactile vision substitution: past and future. *Int. J. Neurosci.* **19**, 29–36.

Baron-Cohen, S. 1994 How to build a baby that can read minds: cognitive mechanisms in mindreading. *Cah. Psychol. Cogn.* **13**, 513–552.

Baron-Cohen, S. 1995 *Mindblindness: an essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Baron-Cohen, S., Jolliffe, T., Mortimore, C. & Robertson, M. 1997 Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *J. Child Psychol. Psychiat.* **38**, 813–822.

Bates, E. 1976 *Language and context: the acquisition of pragmatics*. New York: Academic Press.

Bates, E. 1979 On the evolution and development of symbols. In *The emergence of symbols: cognition and communication in infancy* (ed. E. Bates), pp. 1–32. New York: Academic Press.

Blakemore, S.-J. & Decety, J. 2001 From the perception of action to the understanding of intention. *Nature Rev. Neurosci.* **2**, 561–567.

Bonda, E., Petrides, M., Ostry, D. & Evans, A. 1996 Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J. Neurosci.* **15**, 3737–3744.

Boucher, J. & Lewis, V. 1992 Unfamiliar face recognition in relatively able autistic children. *J. Child Psychol. Psychiat.* **33**, 843–859.

Bowler, D. M. 1992 'Theory of mind' in Asperger's syndrome. *J. Child Psychol. Psychiat.* **33**, 877–893.

Bryant, P. E. 1991 Face to face with babies. *Nature* **354**, 19.

Bushnell, E. W. & Boudreau, J. P. 1993 Motor development and the mind: the potential role of motor abilities as a determinant of aspects of perceptual development. *Child Dev.* **64**, 1005–1021.

Bushnell, I. W. R., Sai, F. & Mullin, J. T. 1989 Neonatal recognition of the mother's face. *Br. J. Devl Psychol.* **7**, 3–15.

Celani, G., Battacchi, M. W. & Arcidiacono, L. 1999 The understanding of the emotional meaning of facial expressions in people with autism. *J. Autism Devl Disorders* **29**, 57–66.

Churchland, P. S., Ramachandran, V. S. & Sejnowski, T. J. 1994 A critique of pure vision. In *Large-scale neuronal theories of the brain. Computational neuroscience* (ed. C. Koch & J. L. Davis), pp. 23–60. Cambridge, MA: MIT Press.

Clark, A. 1999 An embodied cognitive science? *Trends Cogn. Sci.* **3**, 345–351.

Dahlgren, S. O. & Trillingsgaard, A. 1996 Theory of mind in non-retarded children with autism and Asperger's syndrome. A research note. *J. Child Psychol. Psychiat.* **37**, 759–763.

Damasio, A. 1999 *The feeling of what happens: body and emotion in the making of consciousness*. New York: Harcourt Brace.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J. & Brown, E. 1998 Children with autism fail to orient to naturally occurring social stimuli. *J. Autism Devl Disorders* **28**, 479–485.

Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J. & Webb, S. J. 2002 Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay and typical development. *Child Dev.* **73**, 700–717.

Diamond, R. & Carey, S. 1986 Why faces are and are not special: an effect of expertise. *J. Exp. Psychol.* **115**, 107–117.

Dittrich, W. H., Troscianko, T., Lea, S. E. & Morgan, D. 1996 Perception of emotion from dynamic point-light displays represented in dance. *Perception* **25**, 727–738.

Eimas, P., Siqueland, E., Jusczyk, P. & Vigorito, J. 1971 Speech perception in infants. *Science* **171**, 303–306.

Eisenberg, R. & Marmarou, A. 1981 Behavioural reactions of newborns to speech-like sounds and their implications for developmental studies. *J. Infant Mental Hlth* **2**, 129–138.

Engel, A. K., Fries, P. & Singer, W. 2001 Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Rev. Neurosci.* **2**, 704–716.

Fernald, A. 1993 Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Dev.* **64**, 657–674.

Field, T., Vega-Lahar, N., Scafidi, F. & Goldstein, S. 1986 Effects of maternal unavailability on motion-infant interactions. *Infant Behav. Dev.* **9**, 473–478.

Fox, R. & McDaniel, C. 1982 The perception of biological motion by human infants. *Science* **218**, 486–487.

Frith, C. D. & Frith, U. 1999 Interacting minds: a biological basis. *Science* **286**, 1692–1695.

Gaffan, E. A., Gaffan, D. & Harrison, S. 1988 Disconnection of the amygdala from visual association cortex impairs visual-reward association learning in monkeys. *J. Neurosci.* **8**, 3144–3150.

Gardner, H. 1985 *The mind's new science: a history of the cognitive revolution*. New York: Basic Books.

Gauthier, I. & Nelson, C. A. 2001 The development of face expertise. *Curr. Opin. Neurobiol.* **11**, 219–224.

Gauthier, I. & Tarr, M. 1997 Becoming a 'greeble' expert: exploring mechanisms for face recognition. *Vis. Res.* **37**, 1673–1681.

Gauthier, I., Tarr, M., Anderson, A., Skudlarski, P. & Gore, J. 1999 Activation of the middle fusiform face area increases with expertise in recognizing novel objects. *Nature Neurosci.* **2**, 568–573.

Gergely, G., Nadasdy, Z., Csibra, G. & Biro, S. 1995 Taking the intentional stance at 12 months of age. *Cognition* **56**, 165–193.

Gibson, J. J. 1963 The useful dimension of sensitivity. *Am. Psychol.* **18**, 1–15.

Grelotti, D. J., Klin, A., Volkmar, F. R., Gauthier, I., Skudlarski, P., Cohen, D. J., Gore, J. C. & Schultz, R. T. 2003 FMRI activation of the fusiform gyrus and amygdala to cartoon characters but not faces in a boy with autism (Submitted.)

Grezes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C. & Decety, J. 2001 Does perception of biological motion rely on specific brain regions? *NeuroImage* **13**, 775–785.

Grigorenko, E. L., Klin, A., Pauls, D. L., Senft, R., Hooper, C. & Volkmar, F. R. 2002 A descriptive study of hyperlexia in a clinically referred sample of children with developmental delays. *J. Autism Devl Disorders* **32**, 3–12.

Grossman, E. & Blake, R. 2000 Brain activity evoked by inverted and imagined motion. *Vision Res.* **41**, 1475–1482.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G. & Blake, R. 2000 Brain areas involved in perception of biological motion. *J. Cogn. Neurosci.* **12**, 711–720.

Gusella, J. L., Muir, D. W. & Tronick, E. Z. 1988 The effect of manipulating maternal behavior during an interaction on 3- and 6-month olds' affect and attention. *Child Dev.* **59**, 1111–1124.

Hadwin, J., Baron-Cohen, S., Howlin, P. & Hill, K. 1997 Does teaching theory of mind have an effect on the ability to develop conversation in children with autism? *J. Autism Devl Disorders* **27**, 519–537.

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

Happé, F. G. 1994 An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Devl Disorders* **24**, 129–154.

Haviland, J. M. & Lelwica, M. 1987 The induced affect response: 10-week-old infants' responses to three emotional expressions. *Devl Psychol.* **23**, 97–104.

Haxby, J., Ungerleider, L., Clark, V., Schouten, J., Hoffman, E. & Martin, A. 1999 The effect of face inversion on activity in human neural systems for face and object perception. *Neuron* **22**, 189–199.

Heider, F. & Simmel, M. 1944 An experimental study of apparent behavior. *Am. J. Psychol.* **57**, 243–259.

Held, R. 1965 Plasticity in sensory-motor systems. *Sci. Am.* **213**, 84–94.

Held, R. & Hein, A. 1958 Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. *Percept. Motor Skills* **8**, 87–90.

Held, R. & Hein, A. 1963 Movement-produced stimulation in the development of visually guided behavior. *J. Comp. Physiol. Psychol.* **56**, 872–876.

Hobson, R. P., Ouston, J. & Lee, A. 1988 What's in a face? The case of autism. *Br. J. Psychol.* **79**, 441–453.

Howlin, P., Baron-Cohen, S. & Hadwin, J. 1999 *Teaching children with autism to mind-read*. New York: Wiley.

Iacoboni, M. 2000*a* Attention and sensorimotor integration: mapping the embodied mind. In *Brain mapping: the systems* (ed. A. W. Toga & J. C. Mazziotta), pp. 463–490. San Diego, CA: Academic Press.

Iacoboni, M. 2000*b* Mapping human cognition: thinking, numerical abilities, theory of mind, consciousness. In *Brain mapping: the systems* (ed. A. W. Toga & J. C. Mazziotta), pp. 523–534. San Diego, CA: Academic Press.

Iriki, A., Tanaka, M. & Iwamura, Y. 1996 Coding of modified body schema during tool use by macaque postcentral neurons. *NeuroReport* **7**, 2325–2330.

Jackendoff, R. 1987 *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Johansson, G. 1973 Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* **14**, 201–211.

Johnson, M. 1987 *The body in the mind: the bodily basis of meaning*. University of Chicago Press.

Johnson, M. 2001 Functional brain development in humans. *Nature Rev. Neurosci.* **2**, 475–483.

Jordan, H., Reiss, J. E., Hoffman, J. E. & Landau, B. 2002 Intact perception of biological motion in the face of profound spatial deficits: Williams syndrome. *Psychol. Sci.* **13**, 162–167.

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

Klin, A. 1989 Understanding early infantile autism: an application of G. H. Mead's theory of the emergence of mind. *Lond. School Economics Q.* **3**, 336–356.

Klin, A. 1991 Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *J. Autism Devl Disorders* **21**, 29–42.

Klin, A. 1992 Listening preferences in regard to speech in four children with developmental disabilities. *J. Child Psychol. Psychiat.* **33**, 763–769.

Klin, A. 2000 Attributing social meaning to ambiguous visual stimuli in higher functioning autism and Asperger syndrome: the social attribution task. *J. Child Psychol. Psychiat.* **41**, 831–846.

Klin, A., Sparrow, S. S., de Bildt, A., Cicchetti, D. V., Cohen, D. J. & Volkmar, F. R. 1999 A normed study of face recognition in autism and related disorders. *J. Autism Devl Disorders* **29**, 497–507.

Klin, A., Schultz, R. & Cohen, D. 2000 Theory of mind in action: developmental perspectives on social neuroscience. In *Understanding other minds: perspectives from developmental neuroscience*, 2nd edn (ed. S. Baron-Cohen, H. Tager-Flusberg & D. Cohen), pp. 357–388. Oxford University Press.

Klin, A., Jones, W., Schultz, R., Volkmar, F. & Cohen, D. 2002*a* Defining and quantifying the social phenotype in autism. *Am. J. Psychiat.* **159**, 895–908.

Klin, A., Jones, W., Schultz, R., Volkmar, F. R. & Cohen, D. J. 2002*b* Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiat.* **59**, 809–816.

Koslowski, L. T. & Cutting, J. E. 1978 Recognizing the sex of a walker from point-lights mounted on ankles: some second thoughts. *Percept. Psychophys.* **23**, 459.

Lakoff, G. & Johnson, M. 1999 *Philosophy in the flesh*. Cambridge, MA: MIT Press.

Langdell, T. 1978 Recognition of faces: an approach to the study of autism. *J. Child Psychol. Psychiat.* **19**, 255–268.

Le Grand, R., Mondloch, C. J., Maurer, D. & Brent, H. P. 2001 Early visual experience and face processing. *Nature* **410**, 890.

LeDoux, J. 2002 *Synaptic self: how our brains become who we are*. New York: Viking Penguin.

Leslie, A. 1987 Pretence and representation: the origins of 'theory of mind'. *Psychol. Rev.* **94**, 412–426.

Lord, C. 1995 Follow-up of two-year olds referred for possible autism. *J. Child Psychol. Psychiat.* **36**, 1365–1382.

Lord, C. & Paul, R. 1997 Language and communication in autism. In *Handbook of autism and pervasive developmental disorders* (ed. D. Cohen & F. Volkmar), pp. 195–225. New York: Wiley.

McBeath, M., Shaffer, D. & Kaiser, M. 1995 How baseball outfielders determine where to run to catch fly balls. *Science* **268**, 569–573.

Markman, A. N. & Dietrich, E. 2000 Extending the classical view of representation. *Trends Cogn. Sci.* **4**, 470–475.

Maturana, H. R. & Varela, F. G. 1973 *De máquinas y seres vivos*. Santiago, Chile: Editorial Universitaria.

Maurer, D., Lewis, T. L., Brent, H. P. & Levin, A. V. 1999 Rapid improvement in the acuity of infants after visual input. *Science* **286**, 108–110.

Mead, G. H. 1924 *Mind, self, and society*. University of Chicago Press.

Meltzoff, A. N. 1995 Understanding the intention of others: re-enactment of intended acts by 18-month-old children. *Devl Psychol.* **31**, 838–850.

Meltzoff, A. N. & Moore, M. K. 1977 Imitation of facial and manual gestures by human neonates. *Science* **198**, 75–78.

Merleau-Ponty, M. 1962 *Phenomenology of perception*. London: Routledge & Kegan Paul.

Mills, M. & Melhuish, E. 1974 Recognition of mother's voice in early infancy. *Nature* **252**, 123–124.

Mundy, P. & Neal, R. 2000 Neural plasticity, joint attention and autistic developmental pathology. In *International review of research in mental retardation*, vol. 23 (ed. L. Glidden), pp. 141–168. New York: Academic Press.

Nadel, J., Croue, S., Mattlinger, M.-J., Canet, P., Hudelot, C., Lecuyer, C. & Martini, M. 2000 Do children with autism have expectancies about the social behavior of unfamiliar people? A pilot study using the still face paradigm. *Autism* **4**, 133–146.

Neisser, U. 1997 The future of cognitive science: an ecological approach. In *The future of the cognitive revolution* (ed. D. M. Johnson & C. E. Erneling), pp. 247–260. New York: Oxford University Press.

Neri, P., Morrone, M. C. & Burr, D. C. 1998 Seeing biological motion. *Nature* **395**, 894–896.

Newell, A. 1991 *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Oram, M. W. & Perrett, D. I. 1994 Response of anterior superior temporal polysensory (STPa) neurons to 'biological motion' stimuli. *J. Cogn. Neurosci.* **6**, 99–116.

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

Ozonoff, S. & Miller, J. N. 1995 Teaching theory of mind: a new approach to social skills training for individuals with autism. *J. Autism Devl Disorders* **25**, 415–433.

Piaget, J. 1973 The affective unconscious and the cognitive unconscious. *J. Am. Psychoanal. Assoc.* **21**, 249–261.

Prizant, B. & Duchan, J. 1981 The functions of immediate echolalia in autistic children. *J. Speech Hearing Disorders* **46**, 241–249.

Puce, A., Allison, T., Bentin, S., Gore, J. C. & McCarthy, G. 1998 Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.* **18**, 2188–2199.

Regolin, L., Tommasi, L. & Vallortigara, G. 2000 Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure. *Anim. Cogn.* **3**, 53–60.

Rogers, S. 1999 An examination of the imitation deficits in autism. In *Imitation in infancy: Cambridge studies in cognitive perceptual development* (ed. J. Nadel & G. Butterworth), pp. 254–283. New York: Cambridge University Press.

Schultz, R. T., Gauthier, I., Klin, A., Fulbright, R., Anderson, A., Volkmar, F. R., Skudlarski, P., Lacadie, C., Cohen, D. J. & Gore, J. C. 2000 Abnormal ventral temporal cortical activity among individuals with autism and Asperger syndrome during face discrimination. *Arch. Gen. Psychiat.* **57**, 331–340.

Searle, J., Kiefer, F. & Bierwisch, M. 1980 *Speech act theory and pragmatics*. Boston, MA: D. Reidel.

Sparrow, S., Balla, D. & Cicchetti, D. 1984 *Vineland adaptive behavior scales*, expanded edn. Circle Pines, MN: American Guidance Service.

Symons, L. A., Hains, S. M. J. & Muir, D. W. 1998 Look at me: five-month-old infants' sensitivity to very small deviations in eye-gaze during social interactions. *Infant Behav. Dev.* **21**, 531–536.

Tager-Flusberg, H. & Sullivan, K. 1994 A second look at second-order belief attribution in autism. *J. Autism Devl Disorders* **24**, 577–586.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B. & Szechter, L. 1998 Face recognition in young children: when the whole is greater than the sum of its parts. *Vis. Cogn.* **5**, 479–496.

Tarabulsy, G. M., Tessier, R. & Kappas, A. 1996 Contingency detection and the contingent organization of behavior in interactions: implications for socioemotional development in infancy. *Psychol. Bull.* **120**, 25–41.

Thelen, E. & Smith, L. 1994 *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Thelen, E., Schoener, G., Scheier, C. & Smith, L. B. 2001 The dynamics of embodiment: a field theory of infant perseverative reaching. *Behav. Brain Sci.* **24**, 1–86.

Tronick, E., Als, H., Adamson, L., Wise, S. & Brazelton, T. B. 1978 The infant's response to entrapment between contradictory messages in face-to-face interaction. *J. Am. Acad. Child Adolesc. Psychiat.* **17**, 1–13.

Tronick, E. Z., Cohn, J. & Shea, E. 1986 The transfer of affect between mothers and infants. In *Affective development in* *infancy* (ed. T. B. Brazelton & M. W. Yogman), pp. 11–25. Norwood, NL: Ablex.

Uller, C. & Nichols, S. 2000 Goal attribution in chimpanzees. *Cognition* **76**, B27–B34.

van der Geest, J. N., Kemner, C., Verbaten, M. N. & van Engeland, H. 2002 Gaze behavior of children with pervasive developmental disorder toward human faces: a fixation time study. *J. Child Psychol. Psychiat.* **43**, 1–11.

Varela, F., Thompson, E. & Rosch, E. 1991 *The embodied mind: cognitive science and human experience*. Cambridge, MA: MIT Press.

Walker, A. S. 1982 Intermodal perception of expressive behaviors by human infants. *J. Exp. Child Psychol.* **33**, 514–535.

Weeks, S. & Hobson, R. 1987 The salience of facial expression for autistic children. *J. Child Psychol. Psychiat.* **28**, 137–151.

Werner, E., Dawson, G., Osterling, J. & Dinno, H. 2000 Recognition of autism spectrum disorder before one year of age: a retrospective study based on home videotapes. *J. Autism Devl Disorders* **30**, 157–162.

Wetherby, A. M., Prizant, B. M. & Schuler, A. L. 2000 Understanding the nature of communication and language impairments. In *Autism spectrum disorders: a transactional developmental perspective* (ed. A. M. Wetherby & B. M. Prizant), pp. 109–142. Baltimore, MD: Paul H. Brookes Publishing Co.

Winograd, T. & Flores, F. 1986 *Understanding computers and cognition*. Norwood, NJ: Ablex Publications.

Zajonc, R. 1980 Feeling and thinking: preferences need no inferences. *Am. Psychol.* **35**, 151–175.

GLOSSARY

EM: enactive mind
fMRI: functional magnetic resonance imaging
FG: fusiform gyrus