PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Teleological and referential understanding of action in infancy

Gergely Csibra

Phil. Trans. R. Soc. Lond. B 2003 **358**, 447-458
doi: 10.1098/rstb.2002.1235

THE ROYAL SOCIETY

Published online 18 February 2003

Teleological and referential understanding of action in infancy

Gergely Csibra

Centre for Brain and Cognitive Development, School of Psychology, Birkbeck College, Malet Street, London WC1E 7HX, UK (*g.csibra@bbk.ac.uk*)

There are two fundamentally different ways to attribute intentional mental states to others upon observing their actions. Actions can be interpreted as *goal-directed*, which warrants ascribing intentions, desires and beliefs appropriate to the observed actions, to the agents. Recent studies suggest that young infants also tend to interpret certain actions in terms of goals, and their reasoning about these actions is based on a sophisticated teleological representation. Several theorists proposed that infants rely on motion cues, such as self-initiated movement, in selecting goal-directed agents. Our experiments revealed that, although infants are more likely to attribute goals to self-propelled than to non-self-propelled agents, they do not need direct evidence about the source of motion for interpreting actions in teleological terms. The second mode of action-based mental state attribution interprets actions as *referential*, and allows ascription of attentional states, referential intents, communicative messages, etc., to the agents. Young infants also display evidence of interpreting actions in referential terms (for example, when following others' gaze or pointing gesture) and are very sensitive to the communicative situations in which these actions occur. For example, young infants prefer faces with eye-contact and objects that react to them contingently, and these are the very situations that later elicit gaze following. Whether or not these early abilities amount to a 'theory of mind' is a matter of debate among infant researchers. Nevertheless, they represent skills that are vital for understanding social agents and engaging in social interactions.

Keywords: infancy; referential action; goal-directed action; teleology; theory of mind

1. INTRODUCTION

People cannot help but interpret each other's actions in terms of hidden mental states like beliefs, desires, etc. The attribution of these intentional mental states helps them to explain observed, and predict future, behaviour of others, and also enables them to influence what social partners will do. According to philosophers (e.g. Dennett 1987), the distinctive aspect of intentional mental states is their 'aboutness'—they are 'about' certain states of the world. There are two fundamentally different ways to interpret actions as indicating mental states that are about something. First, we can interpret an observed behaviour as a *goal-directed action*. A goal-directed action is 'about' the end state of that action; it is performed in order to make the end state occur. If you see someone drilling a corkscrew into the cork of a bottle of wine, you will know what this action is about: an open bottle of wine. Because goal-directed actions are seen to be determined by their end state, this kind of action interpretation is essentially *teleological*. Teleological interpretation of an action enables us to attach a goal to the action (an open bottle), but it also allows us to attribute to the agent desires (e.g. a desire to drink wine), beliefs (e.g. a belief that there is wine in the bottle) and possibly further mental states as well. The second way to reveal what an action is about is to interpret it as a *referential action*. A referential action is about the state of the world that it highlights. If you see someone pointing to a car, you will know what this action is about: it is about the car. Interpretation of an action as referential emphasizes some aspect of the world as connected to the agent, and allows attribution of referential intentions (drawing your attention to the car), communicative messages (e.g. 'this is the car I was talking about'), and other mental states to the agent.

During the past decade, several studies have been published that attempted to see whether young, preverbal infants are engaged in these kinds of action interpretations and whether they attribute intentional mental states to others. The answer to these questions bears relevance to the debates on the origins and the nature of the human 'theory of mind' and may also help us to understand developmental disorders that are characterized by deficits in social cognition. This paper reviews some of the evidence relevant to these questions. I shall argue that this body of evidence unambiguously shows that infants younger than 1 year of age employ both kinds of action interpretation and do this in a more and more sophisticated way as they approach their first birthday. At the same time, I shall also argue that neither of these action interpretation systems is necessarily mentalistic; both can function without attribution of representational mental states. In other words, goals can be attached to actions without understanding desires, and referents can be identified without figuring out the meaning of the underlying communicative act. I propose that these two action

One contribution of 15 to a Theme Issue 'Decoding, imitating and influencing the actions of others: the mechanisms of social interaction'.

interpretation systems operate independently in early infancy. The independence of these cognitive mechanisms is supported by various facts: they rely on different types of computations, they are triggered by different stimulus conditions, and they can be dissociated in animal behaviours and developmental disorders. Crucially, there is no evidence for transfer between these action interpretation systems (for example attribution of goals on the basis of referential actions) in infants at, or below, 1 year of age. I suggest that the combining of these action interpretation systems into a higher-order mentalistic representation of actions takes place during the second year of life.

2. TELEOLOGICAL UNDERSTANDING OF ACTIONS

Understanding goals requires connecting actions not to their antecedents but to their consequents. Some studies suggest that infants as young as six months of age are sensitive to the end state of observed events. Woodward (1998), for example, repeatedly presented infants with an action in which a hand reached towards and grasped one of two toys. When the infants had been habituated to this event, she swapped the toys and the hand either grasped the same toy in the new location, or the other toy in the same location. Looking times for these two events were markedly different: infants looked longer at the hand grasping the 'new' toy at the 'old' location, as if they expected the hand to reach towards the same toy. This result shows that infants are more likely to associate a grasping hand with the grasped object than with its location, which is useful for interpreting the grasping action as directed to the goal of acquiring a specific object. Nothing in this experiment, however, demonstrates that infants do not simply discriminate between end states. It does not show that they relate actions to end states in an 'in-order-to' clause: the hand performed the reaching and grasping sequence *in order to* acquire a certain object.

An action is goal directed if it is performed not for itself but to achieve an end; in other words, if it is an instrumental action. Whether an action is instrumental in relation to an end state or not can be intuitively tested by considering if we would expect it to be performed when it is not needed for goal achievement. Pulling out a cork from a wine bottle, for example, can be an instrumental action to access the wine, but only when the bottle is not empty. Even when the bottle is not empty, we would not expect to see this action if all the glasses were filled with wine. As this example suggests, interpreting something as an instrumental action that has been performed to achieve, and gets its meaning from, a particular goal state requires consideration of many things—not just the action and its end state but also the environment in which it occurs. We have demonstrated that nine-month-old infants are already capable of doing this trick and, depending on the environment, interpret observed actions as goal directed.

These studies (Gergely *et al.* 1995; Csibra *et al.* 1999) presented infants with computer animations. Since the seminal work by Heider & Simmel (1944) several experiments provided evidence that high-level, sophisticated social interpretation of actions does not require featural identification of agents (for a review see Scholl & Tremoulet 2000). Triangles and circles moving on a two-dimensional surface are readily interpreted as if they were people engaged in various kinds of social interactions. One explanation of this phenomenon is that during development the motion patterns of human behaviour are abstracted away from their usual context and from human features, and when these patterns are recognized in artificially created animations, we tend to project further human attributes to those abstract figures. An alternative explanation, which I defend here, suggests that certain motion patterns that allow interpretation of the behaviour of the abstract figures as goal-directed actions will always attract such an interpretation and this tendency does not depend on extensive experience with human behaviour. This explanation predicts that young infants will as readily attribute goals to animated figures as they do to human agents.

Our experiments repeatedly presented infants with a simple animation (figure 1) in which a ball approached and contacted another ball by jumping over an obstacle. In this event, the jumping action can be interpreted as an instrumental action because it is necessary to achieve the end state (the spatial position beside the other ball). Whether or not the infants arrived at the same interpretation was tested by showing them two modified versions of the event and measuring their looking time. In the test events, we removed the obstacle, which changed the relevant aspects of the environment in a way that made the jumping action unnecessary for goal achievement. One of these events (old action) displayed the same jumping approach, which, however, was no longer necessary and was therefore inefficient, while the other event (new action) displayed a straight-line approach to the same position, which was an efficient action in the new situation. If infants interpreted the original event as a goal-directed action, they should find the new instrumental action more compatible with their interpretation and should respond with longer looking at the old, inefficient action.

This is exactly what we found in 9- and 12-month-old infants (see figure 1). Further control studies were performed to verify whether these looking time differences were indeed attributable to the interpretation of the habituation event. The control studies involved the same test events but they were preceded by a habituation event, which differed only slightly from the original habituation event but did not show a proper instrumental action. In this event the 'obstacle' was positioned not in between the two balls but behind the moving ball (see figure 2). In this environment the jumping action cannot be considered to be an instrumental action, and the infants were not expected to attribute the end position as the goal of the moving ball's action. Their looking times confirmed this prediction.

How can we explain infants' early emerging ability to attribute goals to animated shapes? We have proposed (Gergely & Csibra 1997; Csibra *et al.* 2003) that, watching these animations on a computer screen, infants adopt a *teleological stance*. The teleological stance is akin to the *intentional stance* of Dennett (1987) in that it represents an interpretational strategy that seeks to construe an event in terms of goals (see Keil (1995) for a different version of the teleological stance). It is, however, different from the intentional stance in that it does not attribute mental states to the agents (I will return to the relation between these two stances later). Note that the teleological stance (and the intentional stance) is not an explicit inferential system but a bias: a tendency to construe events in accord

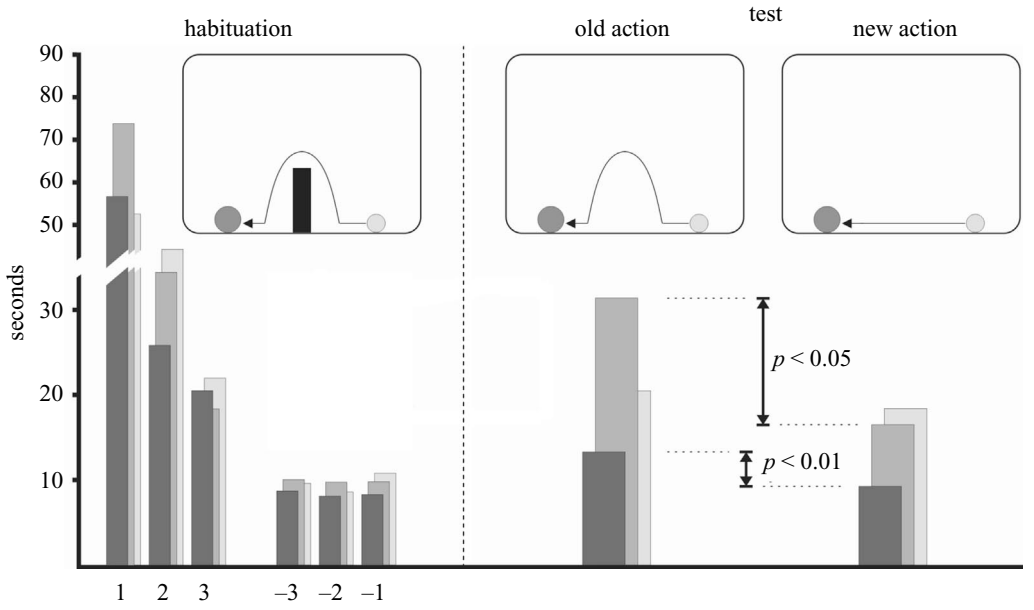

Figure 1. The experimental events and the looking time results in Gergely *et al.* (1995) and Csibra *et al.* (1999). Light grey bars, six-month-old infants; medium grey bars, nine-month-old infants; dark grey bars, 12-month-old infants.

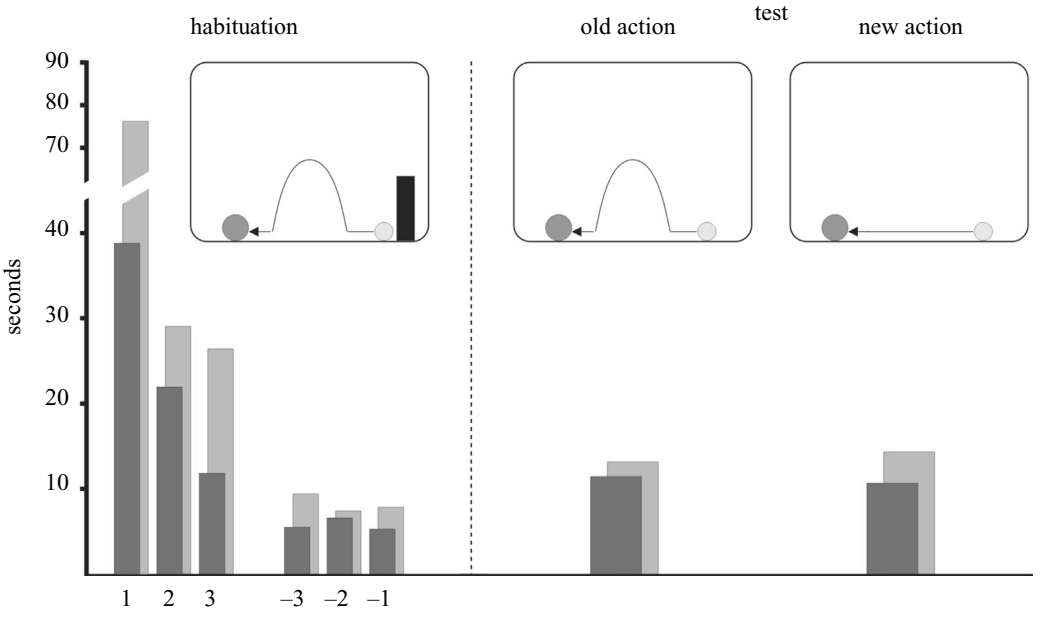

Figure 2. The control events and the the looking time results in Gergely *et al.* (1995) and Csibra *et al.* (1999). Medium grey bars, nine-month-old infants; dark grey bars, 12-month-old infants.

with a certain formal structure. Construing an action as goal-directed implies, as we have seen, relating at least three different aspects of the observed event to each other: the behaviour, its physical context and the end state (see figure 3). These three elements will create a well-formed *teleological representation* of the event if, and only if, the behaviour is an efficient action towards the end state in the given physical environment. The habituation event in our study (figure 1) met this criterion (jumping over the obstacle was the most efficient action towards the end position in that environment) while the habituation event in the control study (figure 2) did not (jumping over nothing was not an efficient action towards the end position in that environment). Thus, the habituation event represented in figure 1 allowed goal attribution, while the habituation event represented in figure 2 did not, and this difference was reflected in the differential patterns of looking time by infants.

As figure 3 suggests, goal attribution and efficiency evaluation are inseparable in the teleological stance. But perhaps this is only true for behaviours of abstract figures, and infants may be willing to attribute goals to inefficient actions as well, if they are performed by real human beings. Recent studies suggest that this is not the case. Woodward & Sommerville (2000), for example, presented infants with two transparent boxes that contained two different toys. They habituated 12-month-old infants to an action in which a hand first touched one of the boxes, then opened it and grasped the toy inside. After habituation, the toys were swapped between the boxes. During the test

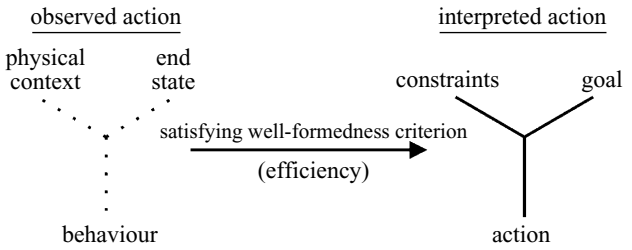

Figure 3. The teleological representation for actions.

event the hand either touched the same box as before (which, however, now contained the other toy) or it touched the other box (which contained the same toy that had been previously grasped). Infants looked longer at the former action, indicating that they did not expect the hand to perform the familiar action seen before, as that was no longer necessary, and in fact would have been an inefficient way to obtain the goal object (i.e. the toy that had been grasped during habituation). In a control study Woodward and Sommerville habituated the infants to the same hand actions (first touching the box, then grasping the toy) with the exception that the toys were not inside but in front of the boxes; hence, opening the box could not have been considered as an efficient instrumental action for grasping the toy. In this condition, the infants did not develop any specific expectation about which box the hand should touch when the toys were swapped and, in fact, they even looked slightly longer when the hand touched the other box. Note that, just as in our study described above, the only difference between the two conditions was the efficiency of the action: opening the box is an efficient instrumental action if, and only if, the target object is in the box.

A more recent study has also demonstrated that infants do care about efficiency when they interpret actions in terms of goals. Onishi's (2001) studies also confirmed that actions (removing obstacles) that make a target object accessible to a hand are interpreted as goal-directed by 10-month-old infants, but the same actions do not lead to goal attribution if they are not justified by the relative positions of the obstacles and the target object. Furthermore, evaluating the efficiency of actions is not restricted to looking-time context either. Gergely *et al.* (2002) provided evidence that infants modulate their imitative behaviour according to the justifiability of the goal-directed actions performed by a model. These authors replicated the well-known demonstration of Meltzoff (1988) that 14-month-old infants tend to imitate a behaviour that they have never seen before and would not perform spontaneously. In Meltzoff's study, infants watch a model leaning forward and touching an object on a table with his head, causing it to light up. A week later, when they are brought back to the laboratory, the majority of the infants perform the same action. In an additional condition, however, Gergely *et al.* slightly modified the model's behaviour. Before touching the object with her forehead, the experimenter covered her shoulder with a blanket, which then she held onto tightly with her hands—then she performed the same action. Or was it the same action? Touching the object with her head was a perfectly reasonable action in a situation where the actor's hands were unavailable, while in the other situation, where she kept her hands free, it just seemed unjustifiable. If infants are not sensitive to this difference, they should imitate the head-action equally. In fact, only a minority of them imitated the head-action in the 'hands occupied' version, which suggests that they interpreted this action as an instrumental action that one does not have to copy if he or she is free from the constraints that affected the model. In other words, they could attempt to achieve the same effect in the most efficient manner that was available to them: by touching the object with their hands. Infants do care about efficiency of perceived actions.

We have seen that nine-month-old infants will attribute the end state as the goal of the action if it is an efficient instrumental action, but people do not have to observe a complete action to attribute a goal to it. We can figure out what the goal of an action could be by considering what end state that action would be instrumental to. Pulling out the cork from a bottle is most likely to be carried out in order for the agent be able to access the content of the bottle. When we make these kinds of inferences, we have information about the agent's behaviour and the physical environment in which it takes place. We then adopt the teleological stance, and try to fill out the missing third element (the end state) of the teleological representation in order to satisfy the well-formedness criterion of efficiency (see figure 3). Can infants do the same trick?

Meltzoff (1995) let 18-month-old infants watch a model who performed apparently failed actions on novel objects. When these babies had a chance to imitate the model, they did not copy the failure but performed the complete intended action that led to the goal state that one could have inferred from the model's behaviour. This is a clear demonstration that 18-month-olds do not have to see the goal realized in order to be able to attribute it to an action. Twelve-month-old infants, however, fail in this task (Bellagamba & Tomasello 1999), suggesting that they may not be able to extract the goal from the observation of failed attempts. It is possible, though, that this test requires too much from babies: they not only have to attribute an unseen goal to the agent, they also have to ignore the observed end state of the action and replace it with an inferred one. In other words, Meltzoff's task requires a kind of counterfactual reasoning that may exceed 12-month-old infants' capabilities even if they were able to figure out goals for unfinished actions.

To avoid the complications inherent in the Meltzoff (1995) study, we created a computer animation that allows teleological interpretation of an action even though the goal is never seen achieved (Csibra *et al.* 2003). The animation shows a simple chase event in which a bigger ball follows a smaller one (see figure 4, habituation event). When the small one passes through a narrow gap between two barriers, the big ball takes a detour around the barriers and then continues its path in the direction where the small one left the screen. The goal of the big ball can easily be identified: catching the small ball. Note, however, that this goal is never seen achieved and can only be inferred from the evaluation of the big ball's behaviour. We performed two different tests on two different groups of 12-month-old infants to check whether they interpret the unfinished chase event as a goal-directed action. In the first test (figure 4), we changed the physical environment by enlarging the gap between the barriers and presented

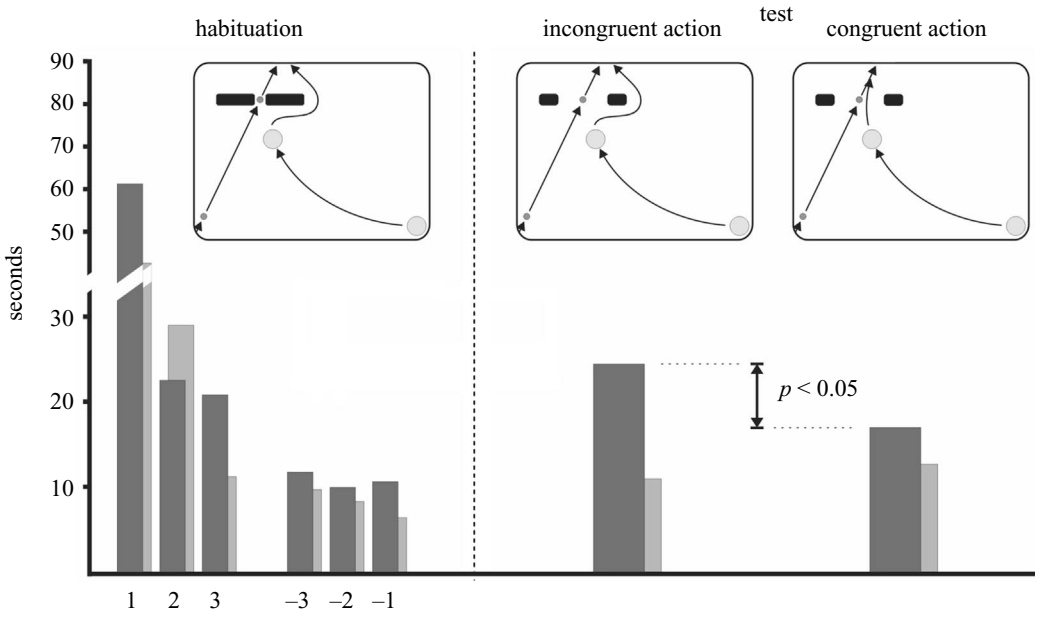

Figure 4. Evidence for attribution of an unseen goal in 12-month-old infants. Medium grey bars, nine-month-old infants; dark grey bars, 12-month-old infants. (Data from Csibra *et al.* (2003), experiment 1.)

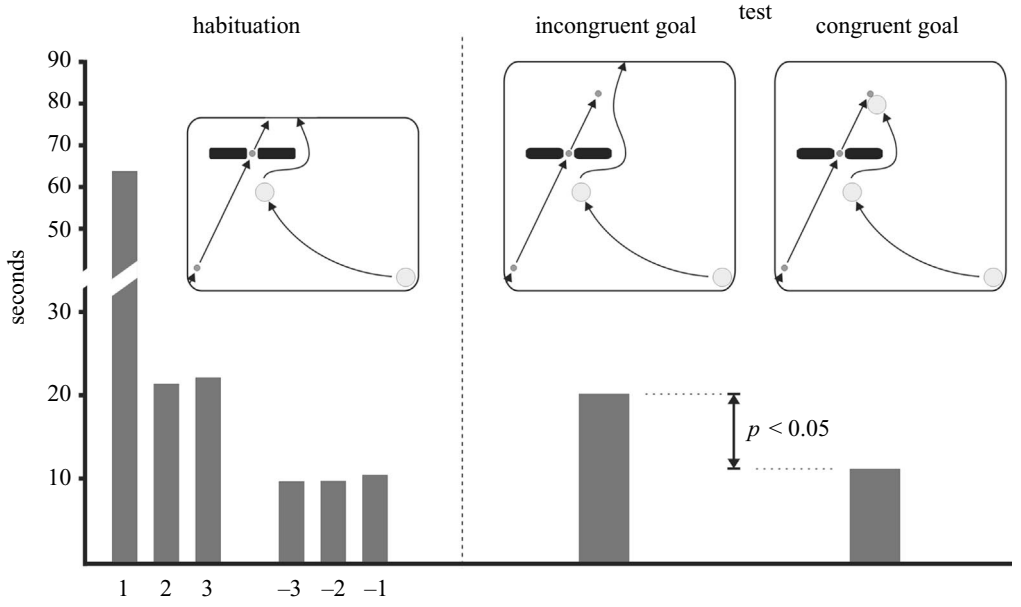

Figure 5. Evidence for attribution of an unseen goal in 12-month-old infants. (Data from Csibra *et al.* (2003), experiment 1A.)

two actions: the big ball either adjusted its path to the new constraints and followed the small ball through the gap (congruent action), or took the same detour as before, which, however, was no longer an efficient action towards the same goal (incongruent action). In the second test (figure 5), we opened up the previously hidden part of the scene where the balls had left and confronted the infants with two different endings for the story: the small ball halted and the big ball either stopped next to it (congruent goal) or changed its path, travelled past the small one, and left the scene (incongruent goal). Seeing an incongruent action or an incongruent goal resulted in longer looking times than seeing the corresponding congruent action or congruent goal events, indicating that 12-month-old infants took the teleological stance and were able to figure out the unseen goal of an agent. This result was recently replicated by Wagner & Carey (2002).

The teleological representational schema for actions (figure 3) allows a further type of inference as well. So far we have seen that, knowing the goal and the physical constraints, infants can predict new actions, and, knowing the physical constraints and the actions, they can attribute goals. A third type of inference that one can logically derive from this representational format would allow figuring out some invisible physical constraints on the basis of the observed action and its end state. This inference would be drawn on the same basis as the previous inferences: filling in the missing element (in this case, the physical constraints) of the schema with something that makes it a well-formed representation, i.e. with something

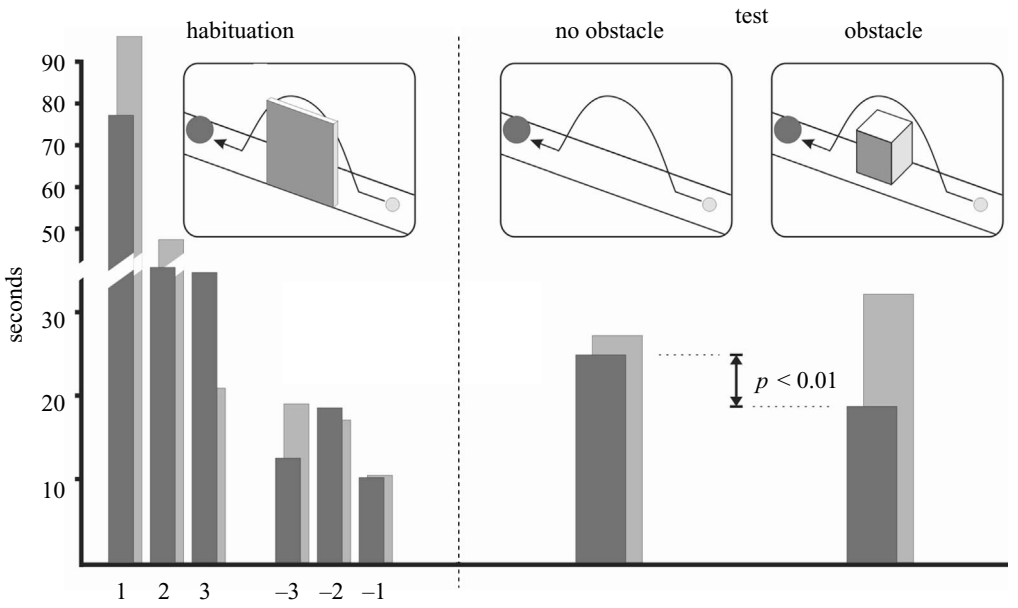

Figure 6. Evidence for the inference of an invisible action constraint in 12-month-old infants. Medium grey bars, nine-month-old infants; dark grey bars, 12-month-old infants. (Data from Csibra *et al.* (2003), experiment 2.)

that satisfies the efficiency principle. To elaborate our example, if someone takes a corkscrew and starts drilling it into a bottle, we will spontaneously assume that there is a cork in the bottle, because otherwise this action would not make any sense, i.e. it could not be interpreted as an efficient action towards the known goal (accessing the wine in the bottle).

To test whether infants make similar inferences, we presented them with a computer-animated event (Csibra *et al.* 2003), which was similar to those we used in our earlier studies (see figure 1): a ball approached another ball by a jumping action. The event here, however, differed from the original studies in two respects: we made the animation three dimensional, and occluded the part of the space that the ball jumped over (figure 6). In the test phase, the occluder was removed and it either revealed an object or an empty space. If infants justify the observed jumping action by inferring the presence of an obstacle behind the occluder, seeing the obstacle would confirm, while seeing the empty space would violate, their expectation, which should be reflected by longer looking time in the latter case. This is exactly what we found. Twelve-month-old infants inferred the presence of an obstacle on the sole basis of the behaviour of the ball. Note that the absence of the obstacle does not violate any physical knowledge; it does not *have to be* there. But its absence violates our expectation that the object approaches its goal efficiently and questions the interpretation that the action is performed *in order to* achieve an end, i.e. that it is a goal-directed action.

All these results indicate that, at least by their first birthday, if not earlier, infants rely on a quite sophisticated teleological representational system when they interpret behaviours, and they use this system productively to figure out invisible aspects of actions (Csibra *et al.* 2003). Does this conclusion entail that they use a 'theory of mind', i.e. do they attribute representational mental states, such as intentions, desires and beliefs to the agents? Infants may, indeed, interpret the observed behaviours in mentalistic

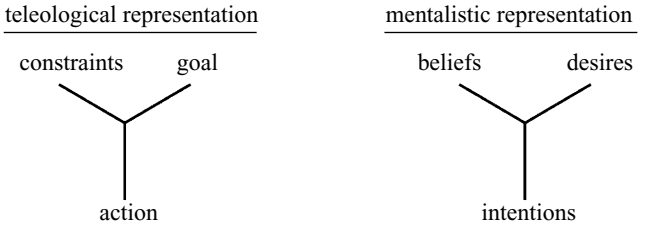

Figure 7. The teleological and mentalistic action explanations.

terms, but this is not a necessary implication of these results. No doubt, one could account for these findings by assuming that the infants interpreted the observed event as 'the ball wanted to touch the other ball, believed that the obstacle was impenetrable and decided to jump over it'. In this interpretation, the elements of the teleological schema are projected into the agent's mind as contents of his/her mental states: goals become desires, constraints become beliefs, and actions become intentions (figure 7). Note, however, that, within the particular context of action interpretation, there is no benefit gained from the computationally more demanding mental state attribution—one can predict exactly the same actions from goals and physical constraints as from contents of desires (i.e. goals) and contents of beliefs (i.e. physical constraints). The additional benefit of relying on mental states comes from situations where the mental states are attributed independently of the actual action and are used in action predictions ('he opens the bottle but he does not like wine, so he wants to offer it to someone else'). But none of the studies I have reviewed above required such inferences and therefore they did not provide conclusive evidence for mental state attribution. In fact, we have argued earlier (Csibra & Gergely 1998) that a non-mentalistic, purely teleological action interpretation system (the teleological stance) developmentally precedes the later emerging mature theory of mind (the intentional stance).

A further question arises from the fact that interpreting an action as goal-directed is not a causal inference but a result of a specific *stance* (whether it is teleological or intentional). What makes infants decide that an observed behaviour is to be interpreted from this stance, i.e. it is to be evaluated in terms of its efficiency? We all take the intentional stance when dealing with fellow human beings, but we also apply mentalistic terms to animals, to natural phenomena, and even to machines (see Dennett 1987), and neither Heider & Simmel's (1944) triangles nor our jumping balls had any animate, let alone human, features. A plausible assumption is therefore that we (and infants) rely on behavioural, rather than featural, cues to identify agents that are possibly engaged in goal-directed actions. The most direct hypothesis about these cues was put forward by David Premack (1990). According to him, infants will treat as intentional (hence goal-directed) any agent that appears to be self-propelled. This idea has been incorporated into several theories of infant development (Mandler 1992; Baron-Cohen 1994; Carey & Spelke 1994; Leslie 1994) and was originally one of the hypotheses that inspired our studies (Gergely *et al.* 1995). Surprisingly, however, not much research was devoted to verifying this hypothesis. Some studies (e.g. Kaufman 1998) suggest that infants make a distinction between self-propelled and externally driven objects from an early age and develop different kinds of expectations towards the two classes of objects (Luo & Baillargeon 2002). Our recent studies indicate that 12-month-old infants are more likely to attribute goals to animated agents that appear to move by themselves than to agents that are launched by other objects (Gergely & Csibra 2003). This indicates that self-motion does indeed work as a cue for goal-directedness.

But this is not the whole story. Infants in our study were more reluctant to attribute goals to externally propelled objects but still tended to take the teleological stance towards them as long as their behaviour appeared efficient. Even when we removed all cues of self-motion and animacy but left enough information to evaluate the efficiency of goal approach (Csibra *et al.* 1999), 9- and 12-month-old infants were willing to attribute a goal to the observed action. These results suggest that, although self-propulsion works as cue, it is not obligatory and there may also be other cues for goal-directedness. Such cues can be derived from the efficiency principle itself. The principle requires, for example, that behaviours directed to the same goal be adjusted in relation to the relevant aspects of the environment in which they occur. Consequently, the perception of behavioural adjustment that is a function of situational constraints may serve as the triggering condition for analysing the behaviour as goal directed.

The left side of table 1 summarizes the proposed specifics of the cognitive system that allows infants to interpret actions as goal-directed.

3. REFERENTIAL UNDERSTANDING OF ACTIONS

Understanding an action as referential requires linking the actor's behaviour to specific objects or to specific aspects of the environment. These actions, such as pointing to or looking at an object, normally occur in communicative contexts, direct the observer's attention to that object, and may help to secure a referent for other communicative signals, such as verbal utterances.

Studies on early language acquisition indicate that there is a special context where referential understanding of actions is indispensable. Young children have been shown to be specifically sensitive to where someone is looking or pointing at when uttering a new word and interpret the word and the observed action as referring to the same thing (Baldwin 1993; Tomasello 1999; Bloom 2000). In other words, they use referential interpretation of non-verbal actions to disambiguate the new word's referent. This is a clear demonstration that 18-month-old infants can understand actions in referential terms. Unfortunately, linguistic tests are difficult to administer before 18 months of age so they are not feasible for testing the understanding of referential actions in younger infants.

To find earlier referential contexts that do not rely on word learning, Louis Moses and his colleagues turned to another phenomenon, known as 'social referencing' (Moses *et al.* 2001). When infants are confronted with a new situation or with a novel object, they tend to check their parents' or other adults' face before approaching it, and modulate their behaviour accordingly (Campos & Stenberg 1981). For example, they cross a visual cliff, if their mother is smiling at the other side, but refrain from crawling over if she looks worried (Sorce *et al.* 1985). Moses and his colleagues tested whether this modulation of behaviour is specific to the object that the adult was looking at when she expressed a certain emotion. They arranged situations where an experimenter and the infant were focusing on either the same or different objects, and then the experimenter expressed either a positive or a negative affect both verbally and by facial expressions. In response to this, infants always looked at the adult's face and checked their line of regard. If infants understand the adult's emotional expressions in referential terms, they should modulate their behaviour towards the object that the experimenter was looking at with that emotion, even when their own attention was engaged by another object. And this is precisely what they found. Twelve-month-old infants explored the target object longer when it had been associated with positive affect, even when it was not in their own attentional focus. This is a clear evidence for referential understanding of looking.

This referential understanding of looking behaviour is assisted by infants' tendency to follow the gaze of other human beings. If you make eye contact with a 12-month-old infant and then conspicuously look at some other object in the environment, she will follow your gaze and many times she will rest her gaze on the same object. This behaviour, which is often called 'joint attention', emerges during the second half of the first year and its accuracy develops rapidly (Butterworth & Jarrett 1991).

But sensitivity to the gaze direction of others can be demonstrated even earlier in laboratory situations. If three- to six-month-old infants are presented with a target object on one side of a computer screen, they are more inclined to orient towards it if they perceive a gaze-shift on a face to the same direction just milliseconds earlier (Hood *et al.* 1998). This phenomenon has been shown to be partly explainable by sensitivity to motion cues provided by the perceived shift of pupil position (Farroni *et al.* 2000), but this does not account for all aspects of the

Table 1. The main characteristics of the two-action interpretation systems.

	goal-directed action	referential action
'about'	a future state	a current state
criteria of application	self-propelled agent; contingent adjustment; biological motion	communicative situations (eye contact, contingent reactivity)
initial principle	efficiency	directionality
mature principle	rationality	relevance
representation	constraints, goal → action	agent → action → referent

results. Recent studies revealed that pupil motion cues are effective in eliciting shifts of attention in infants if, and only if, they are preceded by a period of eye contact (Farroni *et al.* 2003). Motion cues do not elicit attentional shifts if the perceived gaze of the face on the computer screen is moving from the side to the centre (i.e. from an averted to an eye-contact position), or when the face is presented upside down. This makes sense because, at least in humans, making eye contact is the simplest way of establishing a communicative situation, and referential actions usually occur in communicative contexts.

An interesting hypothesis that one can draw from these results is that referential understanding of actions is assisted by infants' sensitivity to two kinds of cues: those that indicate a communicative situation and those that indicate spatial directions. The combination of these two tendencies ('look for communicative situations and, if you find one, follow the direction') may ensure that infants will find the referent of a communicative act in most cases. Again, the sensitivity to communicative situations does not imply an understanding of communication. Rather, it represents a bias in processing the information available in the infant's environment. This hypothesis provided us with the prediction that even the youngest infants must be sensitive to the best cue for a communicative situation, i.e. eye contact. We tested this hypothesis with newborns in a simple preferential looking paradigm (Farroni *et al.* 2002). Seventeen 1–5-day-old newborns were shown two faces: one that looked straight at them (direct gaze) and one that looked away (averted gaze). All but two of them looked longer at and all of them looked more times towards the face with the direct gaze (figure 8). This early, and most probably innate, preference for eye contact can be interpreted in various theoretical frameworks that were developed to explain early sensitivity to social cues. It fits well into the eye-direction detection mechanism hypothesized by Baron-Cohen (1994) and it can also be adapted into Morton & Johnson's (1991) 'CONSPEC' mechanism, which orients babies to faces. The fact that faces with direct gaze engage four-month-old infants' brain circuits that are associated with face perception stronger than do faces with averted gaze (Farroni *et al.* 2002) is more consistent with the latter theory. But whatever the exact mechanism is, the fact that newborns are sensitive not just to faces (Johnson *et al.* 1991) and eyes (Batki *et al.* 2000) but also to eye contact gives them a kick-start towards understanding referential actions.

However, eye contact is not the only method for establishing a communicative situation. Contingent responses from a source may also indicate that the source is communicating with you. John Watson argued 30 years ago that very young infants' well-known sensitivity for objects that respond to their own actions with high but imperfect contingency is not simply a preparedness for operant conditioning but a way to find social partners in the world (Watson 1972). Suzan Johnson and colleagues have demonstrated that 12-month-old infants will follow the 'gaze' of a non-human object if it reacts to the children's actions and vocalizations contingently (Johnson *et al.* 1998; see also Johnson 2003). This is a very clear example of interpreting a behaviour as a referential action when the only cue for treating the object's behaviour meaningfully is the communicative situation established by contingent reactivity. And contingent reactivity can indicate a communicative context for 10-month-old infants even when it comes from a clearly mechanical object, like a robot (Movellan & Watson 2002).

This example suggests that early understanding of referential actions evident in 'gaze following' phenomena originates not from a rich comprehension of the link between a mental state of an agent and its referent but from a blind tracking of motion cues in communicative contexts. In other words, referential interpretation of actions, just like teleological interpretation of actions, represents not a knowledge but a 'stance'. Taking this 'referential stance' triggers a search for referents on the basis of directional cues and should be initially restricted to well-defined communicative contexts. This interpretation of early capacities explains several aspects of the results in this field. Although infants tend to follow the gaze of others, initially they will not find the object looked at by the other person (Butterworth & Jarrett 1991). Even when nine-month-old infants follow pointing gestures, they would not necessarily associate the pointing action with the pointed object (Woodward & Guajardo 2002). It is not before they are 16 months of age that they tie directional motion cues to the line of regard of others correctly, for example, when they follow head turns that they observe from behind (Muir & Lee 2002; see also Johnson 2003).

What purpose does this rudimentary understanding of referential actions serve, if it does not specify what the actor's intentions have been behind his/her actions, i.e. if it does not provide mental state attribution? The answer seems obvious: referential understanding of actions might

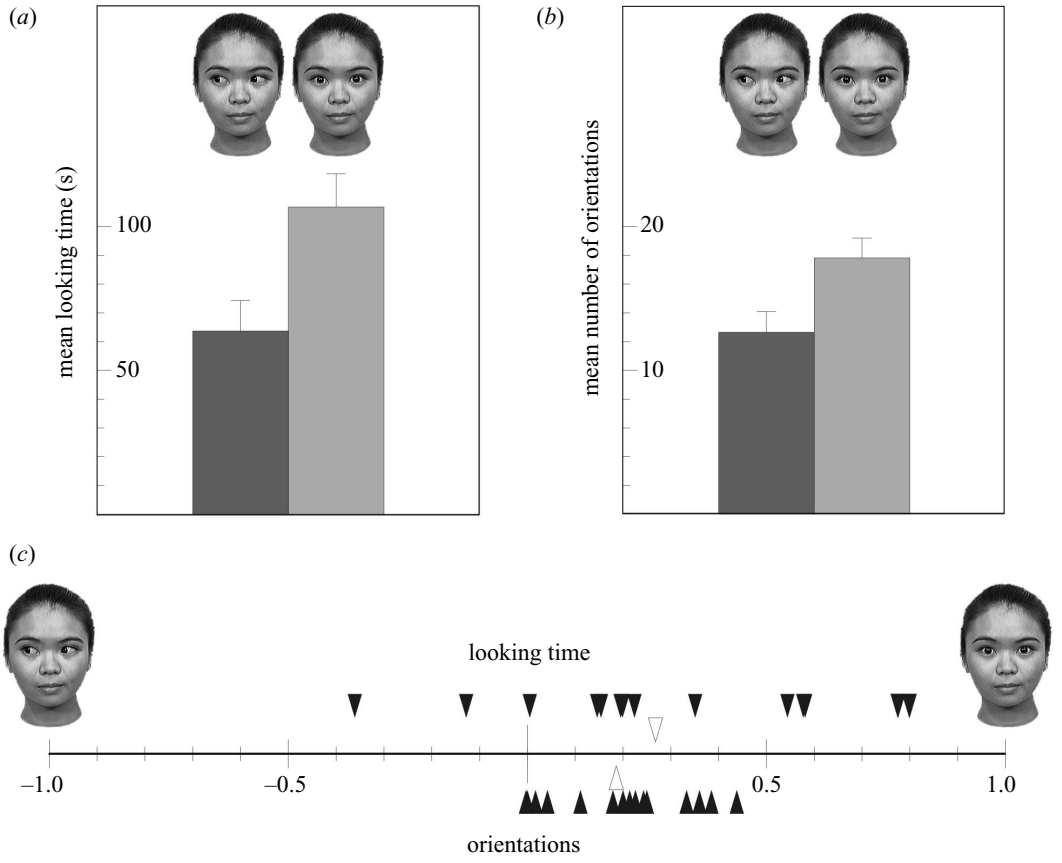

Figure 8. Newborn babies look longer at (*a*), and more times towards (*b*), faces with direct eye gaze. (*c*) Individual preference scores also show a bias towards direct eye gaze both in looking times and number of orientations. (Data from Farroni *et al.* (2002).)

have evolved to support children's word learning. If infants interpret linguistic utterances and non-verbal communicative actions, like looking or pointing, in referential terms, then they can assume that simultaneous references by the same person will refer to the same object. If non-arbitrary referential relationships, like the spatial relationships inherent in looking and pointing, help them find a referent, it may also help them establish arbitrary referential relationships, like the one between words and objects. (The study by Moses *et al.* (2001) that is mentioned above shows how this works in a non-linguistic context.) This role of action interpretation in word learning is well documented (Baldwin 1993; Tomasello 1999; Bloom 2000) but it is usually assumed that this interpretation must yield a mental state (i.e. the referential intention of the speaker), to which both the verbal and non-verbal actions can be mapped. This is, however, not required; direct mapping between words and non-verbally referred objects can also function as a bootstrapping mechanism into word learning. Recently, Sperber & Wilson (2002) have also proposed that the early processing of communicative signals may not be based on general-purpose mind-reading mechanisms but relies on a sub-module which evolved to support fast comprehension of ostensive stimuli. Understanding of referential intentions does indeed play an important role in language acquisition from 18 to 24 months of age. Nevertheless, the development of this essential cognitive skill suggests that an understanding of referential intentions is not a precondition for, but the product of, understanding referential actions. This hypothesis is parallel with the one I proposed for teleological action understanding. Just as the notion of pre-existing goals (i.e. desires) is derived developmentally from teleological understanding of actions, and not the other way around, I propose that the notion of pre-existing 'meaning' (i.e. the communicative message) is derived developmentally from referential understanding of actions, and not the other way around.

It is not yet clear how this development is achieved. However, early understanding of referential actions, at least in one respect, seems to be very similar to the principle that governs mature human communication, the relevance principle (Sperber & Wilson 1986). The application of both the relevance principle and the referential interpretation of actions depends only on the recognition of a communicative context, and both work as a presumption that cannot be violated. The specifics of the referential action interpretation system are summarized in table 1.

4. TELEOLOGICAL VERSUS REFERENTIAL UNDERSTANDING OF ACTIONS

The evidence reviewed so far suggests that 12-month-old or even younger infants readily interpret action as goal-directed or referential. What is the relationship between these two types of action understanding? Some researchers of infants tend to treat these two kinds of action interpretation as equivalent (e.g. Phillips *et al.*

2002; Woodward & Guajardo 2002), calling them both 'object-directed actions'. This is because in the majority of experiments with infants the agent's goal is to seize an object or the referred state of affairs is an object. I think, however, that 'object-directedness' is a misleading term because it relies on a surface similarity between these particular cases of action interpretation (namely, that both refer to a relationship between an agent and a distal object) and conceals the fact that they rely on different kinds of action understanding. Indeed, a comparison of the specifics of these two systems (see table 1) suggests that there is not much common in them: they are triggered by different cues, apply different representations and computations, serve different functions, and are likely to be implemented in separate mechanisms.

One may argue, however, that the distinction between these systems is artificial; after all, both represent intentional actions and the difference between them is simply determined by the content of the actual intention. In other words, they may be subsystems of a single 'theory of mind' mechanism (Leslie 1994). Nevertheless, it is not only formal arguments that support the claim that these two kinds of action interpretation reflect two distinct cognitive mechanisms. Various kinds of dissociations confirm that these action interpretation systems can operate independently without the help of a higher-level theory of mind. The first of these dissociations occurs in the animal kingdom. Chimpanzees, for example, seem to be able to attribute goals to observed actions (Uller & Nichols 2000), and can be trained to follow gaze alterations (Tomasello *et al.* 2001; Okamoto *et al.* 2002). They also understand what other individuals can see (Call 2001). However, there is no evidence that they conceive representational mental states at all (Call & Tomasello 1999), and indeed, their understanding of seeing does not reflect an understanding of reference (Povinelli *et al.* 1999). A second dissociation can be seen in autism, a developmental disorder characterized by severe difficulties in attributing mental states to others (Frith 2001). However, people with autism can attribute goals to animated shapes the same way as typical children (Abell *et al.* 2000; Castelli *et al.* 2002), suggesting that their teleological action interpretation system is intact. At the same time, their main difficulties seem to be rooted in a non-functioning referential interpretational system: they do not make eye contact (Phillips *et al.* 1992), fail to understand eye gaze (Baron-Cohen 1994), and as a result their acquisition of language is seriously delayed and impaired. These dissociations could hardly occur if the two action interpretation systems discussed in this paper were simply different manifestations of the same mechanism (for a more detailed discussion of these dissociations see Gergely (2002)).

A third type of dissociation is provided by the studies of human infants themselves. If a single, mentalistic action interpretation system existed, which could attribute intentions by both teleological and referential interpretation of actions, it would provide a link between these mechanisms, as it does in older children and adults. We can use a referential act (e.g. pointing) to figure out the likely goal of another person (e.g. obtaining the referred object); and can use a goal-directed act (e.g. searching) to figure out the referent of a word (Tomasello & Barton 1994). There is no evidence, however, that young infants would be able to make such inferences. Thus, I hypothesize that these two action interpretation systems initially represent separate mechanisms which will be integrated into a higher-order, mentalistic action interpretation during the second year of life.

Indeed, a suitable test for whether and when infants attribute mental states to others would be a demonstration of a transfer between these systems. At what age do infants pass this test? A recent study claimed to demonstrate such a transfer, i.e. goal prediction on the basis of looking behaviour, at 12 and 14 months of age. Phillips *et al.* (2002) habituated infants to a person looking at one of two objects with a positive emotion. In the test phase, infants displayed longer looking time when the person held in her hands the other object than when she held the same object. Phillips *et al.* concluded that infants were able to predict a goal-directed action (grabbing an object) from the referential relation (looking) between the person and one of the objects. Note, however, that no instrumental action was presented to the infants in these studies. They may have inferred that a certain action (grabbing the object) must have taken place between looking and holding behind the closed curtain, but they did not have to do that. All they needed to do was match two referential actions (looking, and holding *plus* looking) with their referent and notice the change of the referred object. In other words, they could have detected the incompatibility between habituation and test events without goal attribution and action prediction, i.e. entirely within the referential action interpretation system.

Other laboratories (e.g. Sodian & Thörmer 2000) have also been trying to establish the age when a link between teleological and referential understanding of actions can be demonstrated. Further research is needed to establish not just the timing but also the mechanism of the integration between these systems, as this step represents a major milestone in the development of a mature theory of mind.

I thank Teresa Farroni, György Gergely, Mark Johnson and an anonymous reviewer for their helpful comments on an earlier version of this paper. The author was supported by the UK Medical Research Council (programme grant G9715587).

REFERENCES

Abell, F., Happé, F. & Frith, U. 2000 Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cogn. Dev.* **15**, 1–16.

Baldwin, D. A. 1993 Early referential understanding: infants' ability to recognize referential acts for what they are. *Devl Psychol.* **29**, 832–843.

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

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J. & Ahluwalia, J. 2000 Is there an innate gaze module? Evidence from human neonates. *Infant Behav. Dev.* **23**, 223–229.

Bellagamba, F. & Tomasello, M. 1999 Re-enacting intended acts: comparing 12- and 18-month-olds. *Infant Behav. Dev.* **22**, 277–282.

Bloom, P. 2000 *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Butterworth, G. & Jarrett, N. 1991 What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *Br. J. Devl Psychol.* **9**, 55–72.

Call, J. 2001 Chimpanzee social cognition. *Trends Cogn. Sci.* **5**, 388–393.

Call, J. & Tomasello, M. 1999 A non-verbal false belief task: the performance of children and great apes. *Child Dev.* **70**, 381–395.

Campos, J. J. & Stenberg, C. R. 1981 Perception, appraisal, and emotion: the onset of social referencing. In *Infant social cognition* (ed. M. Lamb & L. Sherrod), pp. 273–314. Hillsdale, NJ: Lawrence Erlbaum Associates.

Carey, S. & Spelke, E. 1994 Domain-specific knowledge and conceptual change. In *Mapping the mind: domain specificity in cognition and culture* (ed. L. Hirschfeld & S. Gelman), pp. 169–200. New York: Cambridge University Press.

Castelli, F., Frith, C., Happé, F. & Frith, U. 2002 Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* **125**, 1839–1849.

Csibra, G. & Gergely, G. 1998 The teleological origins of mentalistic action explanations: a developmental hypothesis. *Devl Sci.* **1**, 255–259.

Csibra, G., Gergely, G., Bíró, S., Koós, O. & Brockbank, M. 1999 Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition* **72**, 237–267.

Csibra, G., Bíró, S., Koós, O. & Gergely, G. 2003 One-year-old infants use teleological representations of actions productively. *Cogn. Sci.* **27**, 111–133.

Dennett, D. C. 1987 *The intentional stance*. Cambridge, MA: MIT Press.

Farroni, T., Johnson, M. H., Brockbank, M. & Simion, F. 2000 Infants' use of gaze direction to cue attention: the importance of perceived motion. *Visual Cogn.* **7**, 705–718.

Farroni, T., Csibra, G., Simion, F. & Johnson, M. F. 2002 Eye-contact detection in humans from birth. *Proc. Natl Acad. Sci. USA* **99**, 9602–9605.

Farroni, T., Mansfield, E. M., Lai, C. & Johnson, M. H. 2003 Infants perceiving and acting on the eyes: tests of an evolutionary hypothesis. *J. Exp. Child Psychol.* (In the press.)

Frith, U. 2001 Mind blindness and the brain in autism. *Neuron* **32**, 969–979.

Gergely, G. 2002 The development of understanding self and agency. In *Blackwell's handbook of childhood cognitive development* (ed. U. Goswami), pp. 26–46. Oxford: Blackwell.

Gergely, G. & Csibra, G. 1997 Teleological reasoning in infancy: the infant's naive theory of rational action. A reply to Premack and Premack. *Cognition* **63**, 227–233.

Gergely, G. & Csibra, G. 2003 Perception of autonomous motion and goal attribution by one-year-old infants. (In preparation.)

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

Gergely, G., Bekkering, H. & Király, I. 2002 Rational imitation in preverbal infants. *Nature* **415**, 755.

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

Hood, B. M., Willen, J. D. & Driver, J. 1998 Adult's eyes trigger shifts of visual attention in infants. *Psychol. Sci.* **9**, 131–134.

Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. 1991 Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* **40**, 1–19.

Johnson, S. C. 2003 Detecting agents. *Phil. Trans. R. Soc. Lond.* B **358**, 549–559. (DOI 10.1098/rstb.2002.1237.)

Johnson, S. C., Slaughter, V. & Carey, S. 1998 Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Devl Sci.* **1**, 233–238.

Kaufman, L. E. 1998 Five-month-old infants distinguish between inert and self-moving inanimate objects. In *11th Biannual Int. Conf. on Infant Stud.*, April 1998, Atlanta, GA.

Keil, F. C. 1995 The growth of causal understanding of natural kinds. In *Causal cognition: a multidisciplinary debate* (ed. D. Sperber, D. Premack & A. J. Premack), pp. 234–262. Oxford University Press.

Leslie, A. M. 1994 ToMM, ToBy and Agency: core architecture and domain specificity. In *Mapping the mind: domain specificity in cognition and culture* (ed. L. Hirschfeld & S. Gelman), pp. 119–148. New York: Cambridge University Press.

Luo, Y. & Baillargeon, R. 2002 Does a self-moving box possess an 'internal force' and 'mind'? In *13th Biennial Int. Conf. on Infant Stud.*, April 2002, Toronto, Ontario, Canada.

Mandler, J. M. 1992 How to build a baby. II. Conceptual primitives. *Psychol. Rev.* **99**, 587–604.

Meltzoff, A. N. 1988 Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli. *Devl Psychol.* **24**, 470–476.

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

Morton, J. & Johnson, M. H. 1991 CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychol. Rev.* **98**, 164–181.

Moses, L. J., Baldwin, D. A., Rosicky, J. G. & Tidball, G. 2001 Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Dev.* **72**, 718–735.

Movellan, J. R. & Watson, J. S. 2002 The development of gaze following as a Bayesian systems identification problem. UCSD Machine Perception Laboratory Technical Reports 2002.01.

Muir, D. & Lee, K. 2002 Invisible gaze following: perspective taking at 12 months of age. In *13th Biennial Int. Conf. on Infant Stud.*, April 2002, Toronto, Ontario, Canada.

Okamoto, S., Tomonaga, M., Ishii, K., Kawai, N., Tanaka, M. & Matsuzawa, T. 2002 An infant chimpanzee (*Pan troglodytes*) follows human gaze. *Anim. Cogn* **5**, 107–114.

Onishi, K. H. 2001 Ten-month-old infants understand that a goal object can be obtained in multiple ways. In *The Soc. Res. Child Dev.*, Minneapolis, April 2001.

Phillips, A. T., Wellman, H. M. & Spelke, E. S. 2002 Infants' ability to connect gaze and emotional expression to intentional action. *Cognition* **85**, 53–78.

Phillips, W., Baron-Cohen, S. & Rutter, M. 1992 The role of eye contact in goal detection: evidence from normal infants and children with autism or mental handicap. *Dev. Psychopathol.* **4**, 375–383.

Povinelli, D. J., Bierschwale, D. T. & Cech, C. G. 1999 Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *Br. J. Devl Psychol.* **17**, 37–60.

Premack, D. 1990 The infant's theory of self-propelled objects. *Cognition* **36**, 1–16.

Scholl, B. J. & Tremoulet, P. D. 2000 Perceptual causality and animacy. *Trends Cogn. Sci.* **4**, 299–309.

Sodian, B. & Thörmer, C. 2000 Understanding the epistemic aspects of seeing. Precursors in infancy? In *12th Biennial Int. Conf. on Infant Stud.*, July 2000, Brighton, UK.

Sorce, J., Emde, R. N., Campos, J. J. & Klinnert, M. 1985 Maternal emotional signaling: its effect on the visual cliff behavior in 1-year-olds. *Devl Psychol.* **21**, 195–200.

Sperber, D. & Wilson, D. 1986 *Relevance: communication and cognition*. Oxford: Blackwell.

Sperber, D. & Wilson, D. 2002 Pragmatics, modularity and mond-reading. *Mind Lang.* **17**, 3–23.

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

Tomasello, M. & Barton, M. E. 1994 Learning words in non-ostensive contexts. *Devl Psychol.* **30**, 639–650.

Tomasello, M., Hare, B. & Fogleman, T. 2001 The ontogeny of gaze following in chimpanzees, *Pan troglodytes*, and rhesus macaques, *Macaca mulatta*. *Anim. Behav.* **61**, 335–343.

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

Wagner, L. & Carey, S. 2002 Coming to conclusions: 12-month-old infants form expectations about probable endings of motion events. In *13th Biennial Int. Conf. on Infant Stud.*, April 2002, Toronto, Ontario, Canada.

Watson, J. S. 1972 Smiling, cooing, and 'the game'. *Merril Palmer Q.* **18**, 323–339.

Woodward, A. L. 1998 Infants selectively encode the goal object of an actor's reach. *Cognition* **69**, 1–34.

Woodward, A. L. & Guajardo, J. J. 2002 Infants' understanding of the point gesture as an object-directed action. *Cogn. Dev.* **17**, 1061–1084.

Woodward, A. L. & Sommerville, J. A. 2000 Twelve-month-old infants interpret action in context. *Psychol. Sci.* **11**, 73–77.