ELSEVIER

Cognition 56 (1995) 165–193

COGNITION

Taking the intentional stance at 12 months of age

György Gergely*, Zoltán Nádasdy, Gergely Csibra, Szilvia Bíró

Institute for Psychology of the Hungarian Academy of Sciences, Budapest, PO Box 398, H-1394, Hungary

Received November 29, 1993, final version accepted December 22, 1994

Abstract

This paper reports a habituation study indicating that 12-month-old infants can take the "intentional stance" in interpreting the goal-directed spatial behavior of a rational agent. First, we examine previous empirical claims suggesting that the ability to attribute intentions to others emerges during the second half of the first year. It is argued that neither the perceptual evidence (concerning the early ability to discriminate agents), nor the behavioral data (indicating the use of communicative gestures for instrumental purposes) are sufficient to support such claims about the early appearance of a theory of mind, as there are alternative explanations for these phenomena in terms of simpler psychological processes. It is then suggested that to show that an infant indeed attributes an intention to interpret the goal-directed behavior of a rational agent, one needs to demonstrate that the baby can generate an expectation about the most rational future means action that the agent will perform in a new situation to achieve its goal. We then describe a visual habituation study that meets this requirement. The results demonstrate that based on the equifinal structure of an agent's spatial behavior, 12-month-old infants can identify the agent's goal and interpret its actions causally in relation to it. Furthermore, our study indicates that infants of this age are able to evaluate the rationality of the agent's goal-directed actions, which is a necessary requirement for applying the intentional stance. In closing, we discuss some of the theoretical and methodological implications of our study.

1. Introduction

During the last decade the burgeoning literature on the development of "theory of mind" has provided clear evidence that by the fifth year of life

* Corresponding author; E-mail: gergely@cogpsyphy.hu; fax: (36-1) 269-2972.

SSDI 0010-0277(95)00661-3

the young child applies a relatively sophisticated mentalistic interpretational strategy to explain and predict the behavior of other agents (Astington, Harris, & Olson, 1988; Perner, 1991; Wellman, 1990; Whiten, 1991). Briefly: it seems that young children (and possibly great apes, see Premack and Woodruff, 1978; Whiten, 1991) are so-called "belief-desire psychologists" (Fodor, 1987; Wellman, 1990); they attribute intentional mental states (such as desires, goals, and beliefs) to others as the causes of their actions. There is also a growing body of evidence concerning the developmental unfolding of the child's naive theory of mind between the second and the fifth year. For example, numerous studies have demonstrated that 3-year-olds perform poorly on a variety of tasks that require attributing false beliefs to others (Baron-Cohen, Leslie, & Frith, 1985; Perner, Leekam, & Wimmer, 1987; Wimmer & Perner, 1983), while children even in their third year seem able to attribute apparently simpler mental states such as desires (Wellman, 1990, 1991; Wellman & Woolley, 1990). Some authors took the position that the 3-year-old's problem stems from her lack of understanding the representational nature of beliefs (Gopnik & Wellman, 1994; Perner, 1991; Wellman, 1990) – a competence which, they suggest, develops only by the fifth year of life. In contrast, there are theorists (Fodor, 1992; Leslie, 1993) who argue that the young child's apparent inability to attribute false beliefs to others is, in fact, due to performance difficulties in identifying the correct belief content, and not to an actual lack of understanding propositional attitude concepts such as "belief". For example, according to Leslie's recent theory of agency (Leslie, 1994) the infant is innately equipped with a domain-specific modular representational system to represent intentional mental states of agents. Leslie argues that this aspect of the infant's "theory of mind mechanism", which sets up metarepresentational structures involving a core set of primitive propositional attitude concepts such as "believe" and "pretend", develops already between 18 and 24 months of age as evidenced by the emergence of the ability to produce and understand pretend play (see Leslie, 1987, 1988a).

However, as we look even further back in time inquiring about the origins or precursors of a theory of mind during the preverbal phase, we find a relative scarcity of relevant empirical data on the one hand, and a variety of theoretical speculations on the other, which interpret the available evidence in rather divergent ways. In this paper we shall briefly review the literature on the origins of a theory of mind in infancy and argue that the empirical evidence on which current claims about the attribution of intentionality to others during the first year of life are based, while suggestive, is not sufficient to support such interpretations. We shall then describe the kind of data that we believe would be necessary to corroborate such conjectures and report a habituation study that meets these requirements. Based on the results of our study, we shall argue that 12-month-old babies do indeed take an "intentional stance" (Dennett, 1987) in developing expectations about the future behavior of objects that they perceive as rational agents.

2. The origins of a theory of mind in infancy

There is a growing body of evidence indicating that already during the first year of life infants have the capacity to represent and reason about inanimate physical objects (Baillargeon, 1991; Leslie, 1994; Leslie & Keeble, 1987; Spelke, Breinlinger, Macomber, & Jacobson, 1992). Such a naive theory of physics (Spelke, 1990) seems highly adaptive for the infant when dealing with events of the physical world. However, when it comes to comprehending events involving the interactions of a significant subclass of physical objects, namely, *agents*, the infant's knowledge of physical objects is of rather limited value. This is so because the behavioral properties of agents are in many respect significantly different from those of other material objects (Gelman & Spelke, 1981; Leslie, 1994; Mandler, 1992), and so attempts to understand their behavior in terms of the infant's naive theory of physics is likely to fail. As Dennett (1987) has argued, a much more successful interpretational strategy to predict and explain the behavior of agents, is to consider them *rational* and to take an *intentional stance* towards them, which involves the attribution of intentional states (beliefs, desires, goals) as the mental causes of their actions. One may ask then: how early and to what degree are preverbal infants capable of taking the intentional stance towards agents to comprehend and predict their behavior?

There are two major classes of evidence that have been discussed in the literature in relation to the infant's emerging capacity to attribute causal intentional states to others. The first kind is *perceptual*: it has to do with the infant's early ability to *discriminate agents* from other objects. The second kind of data is *behavioral*: it refers to the infant's emerging behavioral capacities to engage in interactions with other agents in ways that suggest an ability to comprehend and manipulate the other's intentional mental states. Below we wish to argue, however, that neither of the above kinds of evidence is sufficient to support the claim for an early presence of a theory of mind in the preverbal infant.

3. The perceptual discrimination of agents

There are a number of stimulus properties that differentiate agents from inanimate objects (e.g., Gelman & Spelke, 1981; Mandler, 1992). One class of stimulus characteristics has to do with different aspects of biological motion, such as self-propelled movement (Premack, 1990), autonomous nonrigid transformations of an object's surface (Gibson, Owsley, & Johnston, 1978), and irregular path of movement (Mandler, 1992). All of these may indicate an internal and renewable source of energy (see Leslie, 1993) that can cause changes in the object's behavior without the impact of some external causal force. Another class of stimulus properties that differentiate agents from physical objects is their involvement in "causation at a distance" (Leslie, 1994; Mandler, 1992): their behavior can be affected by

distal stimuli (implying perception) and they can act to influence another agent from a distance (implying communication).

There is a growing body of evidence indicating that infants can differentiate agents on the basis of such stimulus properties already in the preverbal period. For example, Bertenthal's results (Bertenthal, Proffitt, Spetner, & Thomas, 1985; Bertenthal, 1993) show an early sensitivity to biomechanical movement; Gibson et al. (1978) demonstrated that 5-month-olds can discriminate between nonrigid versus rigid transformations of objects; Leslie's (1982, 1988a) habituation studies indicate that babies less than 6 months of age have differential expectations about the effects on another object of the actions of a human hand versus an inanimate object; Watson (1979, 1984) showed that 4-month-olds are differentially sensitive to high but imperfect degrees of contingent reactivity of objects – a pattern that is characteristic of interacting social objects; and Carlson (1980) provided evidence that 10-month-olds show an expectation that animate (but not inanimate) objects can be influenced at a distance.

There have been several proposals in the literature suggesting that the infant's early sensitivity to the stimulus properties of agents is innately related to the perception of intentionality. Thus, Premack (1990) hypothesized that "the perception of intention, like that of causality, is a hard-wired perception based not on repeated experience but on appropriate stimulation" (p. 2). Leslie (1994), who provided evidence for the direct perception of mechanical causation in 6-month-olds (Leslie & Keeble, 1987), also took a modularist position as to the perception of agency arguing that "the infant's perceptual systems are hardwired for the direct perception of intention and goal-directedness" (Leslie & Happé, 1989, p. 210).

The empirical problem, of course, is that while there may very well be an innate basis for attributing intentionality to agents, the data showing the early perceptual discrimination of agents, while compatible with, are not in themselves sufficient to support such an interpretation. The infant may have developed a simpler (nonintentional) concept of animate agent that can move without an external cause. For example, Poulin-Dubois and Shultz (1988) argue that the understanding of independent agency does not necessarily imply the attribution of intentionality; in fact, they hypothesize that "intentionality could perhaps be regarded as a more advanced and more refined analysis of how agents generate their own behavior" (p. 114), which develops only gradually and later.

4. Intentional communication for instrumental purposes

The second kind of evidence that has been brought to bear on the infant's emerging theory of mind consists of observations of certain types of interactive behaviors infants exhibit towards agents. For example, Trevarthen (1977) interpreted the turn-taking "protoconversational" structure

of caretaker–infant interactions during the second and third months as showing an innate capacity for "primary intersubjectivity" which involves mutual intentionality and communication of intentional states between mother and infant. However, Trevarthen's (1977) notion of "primary intersubjectivity" has been criticized on several grounds (e.g., Golinkoff, 1983; Poulin-Dubois & Shultz, 1988; Stern, 1985) generally pointing out that the interactive turn-taking structure of early mother–infant behavioral exchanges can be understood without attributing to the baby knowledge of the intentional states of the caretaker.

A further type of behavioral evidence consists of a family of new kind of interactive behaviors which emerge in the infant during the last quarter of the first year (Bates, Camaioni, & Volterra, 1975; Bretherton & Bates, 1979; Bretherton, McNew, & Beeghly-Smith, 1981; Bruner, 1975). These behaviors include the systematic use of *communicative gestures for instrumental purposes* such as pointing and gaze alteration (Bates, 1979; Murphy & Messer, 1977; Butterworth & Grover, 1990; Butterworth & Jarrett, 1991) to establish shared attention and reference, as well as the emergence of *social referencing* (Campos & Stenberg, 1981; Klinnert, Campos, Sorce, Emde, & Svejda, 1983; Bretherton, 1984) wherein infants start to use their caretaker's facial emotion expressions to appraise an ambiguous situation. Bretherton (1991) has repeatedly argued (see also Bretherton, 1984; Bretherton & Bates, 1979; Bretherton & Beeghly, 1982) that "the most parsimonious explanation of these phenomena is that, by the end of the first year, infants have acquired a rudimentary ability to impute mental states of self and other (what Premack & Woodruff, 1978, called a theory of mind) and, further, that they have begun to understand that one mind can be interfaced with another" (p. 57).

However, a number of other researchers have resisted the temptation to attribute a "theory of interfaceable minds" to infants on the basis of the behavioral data cited. For example, Butterworth and Jarrett's (1991) work on pointing and gaze alteration demonstrates the sequential development of three spatial ("ecological", "geometrical", and "representational") mechanisms that control the establishment of joint attention during the first 18 months of life, in ways which allow "unrelated minds . . . [to] experience the same object" (p. 69). They explicitly argue that "none of these mechanisms require the infant to have a theory that others have minds; rather the perceptual systems of different observers 'meet' in encountering the same objects and events in the world" (p. 55).[1] The phenomenon of social referencing could also be explained without assuming that the infant

[1] In contrast, it has been argued on conceptual grounds that the emergence of "protodeclarative" pointing, which involves commenting or remarking on an object or event to another person, is an example of "ostensive communication" and, as such, *does* imply the representation of the other person's intentional (attentional) mental state (see Leslie & Happé, 1989; Baron-Cohen, 1991).

necessarily attributes a mental state to the other. For example, based on previous operant experiences (such as the infant seeing her mother's frightened face just before touching the hot stove) the mother's facial emotion expressions may have become discriminative stimuli signalling to the baby the likely positive or negative consequences of the infant's approach behavior. Alternatively, observing the mother's emotion expression may induce the corresponding emotion in the infant who can then proceed to appraise the ambiguous situation on the basis of her *own* felt emotion.

In fact, one can argue that using the infant's intentional instrumental use of others to achieve her goals as evidence for attributing mental states to persons is problematic in general. While such behaviors do indicate that the infant herself is an intentional agent, they clearly do not imply that she must also perceive the other person as an agent whose actions are caused by intentional mental states that she manipulates by her communicative gestures. Influencing the other's behavior through distal gestures can always be explained as a case of social tool use where the child experiences that her gestures exert "magical" power over the other's behavior without any awareness on the infant's part of mediating causal intentions being induced in the other (see Golinkoff, 1983; Shatz, 1983). Based on similar considerations, Poulin-Dubois and Shultz (1988) concluded that currently "no *direct* evidence is available for an implicit attribution of intentions to people by infants. To what extent infants are aware of intentional states remains an open question until appropriate methodologies are created" (p. 120).

5. Identifying goals and predicting actions: the role of equifinality and rationality

What kind of evidence would one need then to support the hypothesis that infants attribute intentions to others? Let us start by asking a more general question: why do we attribute intentional states to agents in the first place? As Dennett (1987) has emphasized: adopting the intentional stance towards others is a useful evolutionary strategy because it allows one to *predict* their future behavior in new situations. Therefore, we suggest that to show that the infant attributes an intention to interpret the observed goal-directed behavior of an agent, we need to demonstrate that she will generate expectations about the particular means actions the agent is likely to perform in a new situation to achieve his goal.

In the past, researchers investigating the infant's early understanding of different aspects of the physical world, used the visual habituation paradigm (Spelke, 1985) successfully to demonstrate that the baby can develop expectations about the state of affairs behind an occluding screen (e.g., Baillargeon, 1991), that she infers on the basis of her naive theory of physics. In the same vein, the present study will attempt to apply the

habituation technique to demonstrate that infants can also infer on the basis of their naive theory of mind the likely future actions of an agent as a function of the intention attributed to him. Before turning to the experimental study, however, we have to discuss two further related questions: when taking the intentional stance (a) on what stimulus basis does the infant identify the purpose or goal that she attributes to an agent; and (b) how does she generate an expectation as to the likely future behavior of the agent on the basis of the attributed intention?

First, it should be pointed out that while a certain stimulus property, such as self-initiated movement (Leslie, 1993; Premack, 1990), might provide a direct perceptual cue to agency, it is, nevertheless, insufficient to identify the content of the particular intention that governs the agent's behavior, since any self-propelled action can be caused by a number of different intentions (Anscombe, 1957). Clearly, however, in order to predict the future behavior of an agent one needs to attribute to him an intention with a specific content; simply perceiving the other as being "generally purposeful" or just acting without an external cause will not do. Heider (1958) argued that one stimulus basis for identifying the particular goal of an agent lies in the *equifinal structure of his actions*: the goal of a rational agent can be discerned from observing that under varying environmental conditions his different actions result in one and the same consequence.

Thus, one may hypothesize that the perception of a direct stimulus cue to agency, such as self-propelled movement, could lead the infant to start monitoring the agent's actions over time to discover the presence of an equifinal outcome, if there is one. The observed equifinal outcome could provide then the specific content of the intention to be attributed to the agent. This hypothesis is in line with Leslie's (1993) proposal which suggests that during the second half of the first year a new representational system develops in the infant to capture the actional properties of agents allowing her "to learn about some immediate goals Agents may have by watching for outcomes. The outcome state of affairs can then be entered into the action representation as the goal state of affairs." We would like to point out, however, that while Leslie is probably right in stating that "outcome information can be useful . . . for construing later actions of Agents that are directed to the same kind of goal", identifying and attributing a goal to an agent on the basis of the equifinal outcome[2] of his behavior will not allow, in and of itself, the infant to anticipate the agent's specific future action in a new situation. This is so because knowing the agent's goal will provide no information as to which of the multiple possible means actions that could lead to the goal the agent will choose to perform. As Dennett (1987)

[2] Note also that equifinality of action in itself is likely to be an insufficient cue for attributing intentionality to an object, as there are many cases of equifinal behavior in the inanimate realm as well (such as the effects of gravitational or magnetic force on objects).

argued, taking the intentional stance allows one to generate specific action predictions only on the basis of a general assumption of rationality: one assumes that a rational agent will choose to perform that particular instrumental action which, given his beliefs about the situation, will lead to his goal in the most rational manner. Therefore, we wish to argue that the infant's theory of agency, which represents the actional properties of agents, must also have some means to represent the *rationality* of an action in relation to its goal, if it is to be used for action prediction. Thus, we hypothesize that apart from monitoring the agent's actions to identify their equifinal outcome, the infant will also evaluate whether the agent's goal-directed actions are rational in the given environmental situation or not.

However, the fact that the rational choice of a particular means action is a joint function of the agent's goal *and his beliefs* about the situation, raises a potential problem for the case of infancy. Several theory of mind researchers hold the view (e.g., Astington & Gopnik, 1991; Gopnik, 1993; Wellman, 1991) that before reaching their third year children are not yet able to attribute beliefs to others to guide their action predictions. In fact, it is precisely on this ground that Gopnik (1993; see also Astington & Gopnik, 1991) concluded that before 3 years of age the child does not have a concept of intentionality, where intentionality is defined as a complex mental state mediating between beliefs, desires, and actions. The problem, therefore, is this: if action prediction from the intentional stance is based on the principle of rationality, and the rationality of an agent's action is a function of his beliefs, does the presumed lack of understanding beliefs in others before 3 years of age imply that the young child will be unable to generate predictions about an agent's future behavior on the basis of an attributed intention?[3]

We wish to argue (and demonstrate) that this is *not* the case. We hypothesize that the infant's theory of agency contains, as one of its foundational component, an assumption of *rationality of action*. This early concept of rational action may be initially restricted to the (spatial–topological) domain of pathways through which agents move in relation to other objects in space. There is a growing body of evidence indicating that by the end of the first year the infant has a relatively sophisticated understanding of the physical (Spelke, 1990; Spelke et al., 1992), causal (Leslie & Keeble, 1987), gravitational (Kim & Spelke, 1992; Spelke et al., 1992) and

[3] Of course, the above position that the understanding of the attitude concept of "belief" is lacking from the 3-year-old's competence, has been criticized on several grounds (see Fodor, 1992, Leslie, 1993; but see Wimmer & Weichbold, 1993). Note, however, that even if one adopts Leslie's alternative theory according to which the metarepresentational ability to represent attitude concepts appears already between 18 and 24 months, the problem of action prediction based on the principle of rationality still remains to be resolved. This is so as it pertains to the developmentally earlier period at the end of the first year, which corresponds to the second level of Leslie's theory of agency ("actional agency") where the attribution of attitude concepts is not yet present even under his theoretical account.

biomechanical (Bertenthal et al., 1985; Bertenthal, 1993) constraints on the spatial movement of objects and agents. We believe that the representational domain of spatial pathways over which the rationality of the goal-directed movement of agents is evaluated is structured by these implicit naive theories of physics and agency. In other words, if the infant represents agents and other objects in space as obeying the principles and constraints of his naive theories of the world, she will be able to compute what spatial pathway would correspond to the most rational approach route towards a goal in a given situation.

Thus, we assume that an infant who observes the equifinal as well as rational approach behavior of an agent towards a given spatial location, will attribute that location as the goal of the agent's actions. Furthermore, we hypothesize that if the spatial arrangement of the agent's position relative to that of his goal is altered, the infant will be able to generate a specific expectation as to the most likely future pathway through which the agent will approach his goal in this new situation. In particular, depending on whether the criteria for rational movement are specified in purely perceptual (Mandler, 1992) or force dynamic (Leslie, 1994) terms, the infant will expect the agent to get to its spatial goal location through the shortest available pathway or through the pathway requiring least effort, respectively.[4]

Let us illustrate this hypothesis with a concrete example to be used later in our habituation study. Imagine an infant observing the behavior of two simple figures (a small circle and a large circle) positioned at a distance from each other with a rectangular figure placed in between them (Fig. 1a). First, the large circle expands then it contracts, regaining its original size. This is immediately followed by a similar expansion–contraction sequence performed this time by the small circle. This sequence of events is then repeated again, providing a contingent turn-taking structure for the stimulus event. After this "exchange", the small circle starts to move towards the large circle (as indicated by the horizontal arrow).

Note that the changes of state the two circles exhibit provide the infant with several types of direct stimulus cues potentially indicating agency: the nonrigid transformation of the surface of the two figures (expansion–contraction); the contingent reactivity of the two circles; and the self-propelled movement of the small circle. According to our hypothesis (see also Leslie, 1994; Premack, 1990), these stimulus cues of agency (or some subset of them) may result in the perceptual categorization of the two circular figures as agents, and may induce the infant to monitor the small circle's actions over time to discover the potential presence of an equifinal outcome.

The event continues by the small circle starting to approach the large

[4] Note that given the gravitational and anatomical constraints on agents' movements, in certain circumstances these two formulations can provide differential predictions.

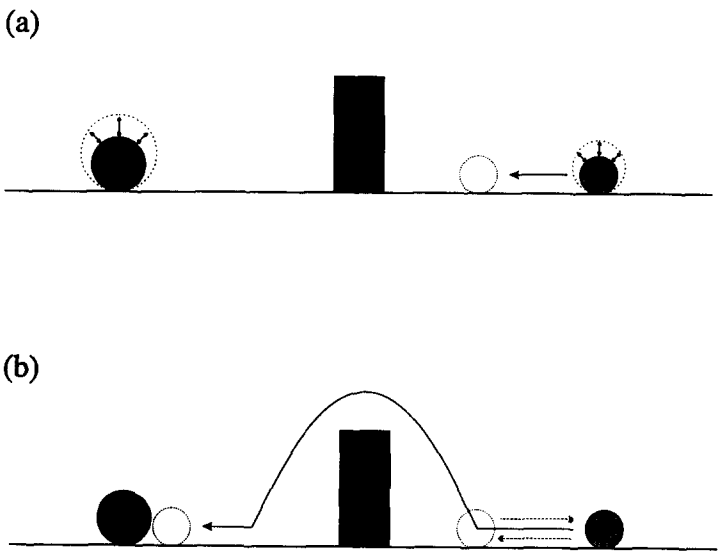

Fig. 1. An illustration of the habituation events for the rational approach group. (a) The large circle expands then contracts, regaining its original size, and it is immediately followed by a similar expansion–contraction sequence performed by the small circle. This sequence of events is then repeated again and then the small circle starts to move towards the large circle. (b) It stops in front of the rectangular figure and then retreats to its original position and starts out again towards the large circle. This time it jumps over the obstacle and, landing in front of the large circle, continues to approach it until they make contact. When they touch each other, the large circle exhibits again the contraction–expansion routine, which is immediately reciprocated by an identical response performed by the small circle, and this interchange is repeated a second time.

circle, following the shortest pathway that could connect them (Fig. 1b). However, it stops in front of the rectangular figure (the "obstacle") which blocks its path to the large circle. The small circle then retreats to its original position and starts out again towards the large circle. However, this time it jumps over the obstacle and, landing in front of the large circle, it continues to approach it until they make contact. When they touch each other, the large circle exhibits again the contraction–expansion routine, which is immediately reciprocated by an identical response performed by the small circle, and this interchange is repeated a second time.[5]

When this event sequence is presented to an infant repeatedly (with the

[5] When we show this visual event to adults, they typically report a dramatic story of intentional social interaction between agents (cf. Heider & Simmel, 1944). For example, they would describe a mother (the large circle) calling her baby (the small circle) to come over to her. The baby immediately responds and then starts running towards her mother, but stops when seeing the obstacle. The baby then retreats but only to try a second time: this time, however, smartly avoiding the obstacle by jumping over it. Finally, when she reaches the mother, they hug each other or kiss happily.

left–right positioning of the two circles also varied), she is clearly in the position to identify the equifinal outcome of the small circle's actions (i.e., the spatial location next to the large circle), and so to attribute it as the small circle's goal.

However, above we argued that equifinality of actions is not likely to be a sufficient condition to attribute intentionality to an agent. A further condition that needs to be fulfilled is that the pathway of the agent's approach should seem rational in the given situation from the point of view of the infant's naive theories of physics and agency. Reaching the goal location by jumping over the obstacle that is blocking the most direct pathway to it, seems a rational means action to be performed by an agent capable of biomechanical movement. By hypothesis, then, repeatedly observing the sequence of events depicted in Fig. 1 should allow the infant to interpret the small circle's behavior as that of a rational agent with an intention to approach its goal.

In contrast, if the same (equifinal) sequence of actions is performed under different environmental conditions, the small circle's behavior may cease to qualify as an instance of rational approach of a goal. Such a situation is depicted in Fig. 2 (nonrational approach). Note that the actions of the two circles are identical to those illustrated in Fig. 1, but the rectangular figure is this time placed *behind* the small circle, rather than in between the two. Of course, in this position it is not an "obstacle" any more as it does not block the shortest pathway between the two circles. While the equifinality of the

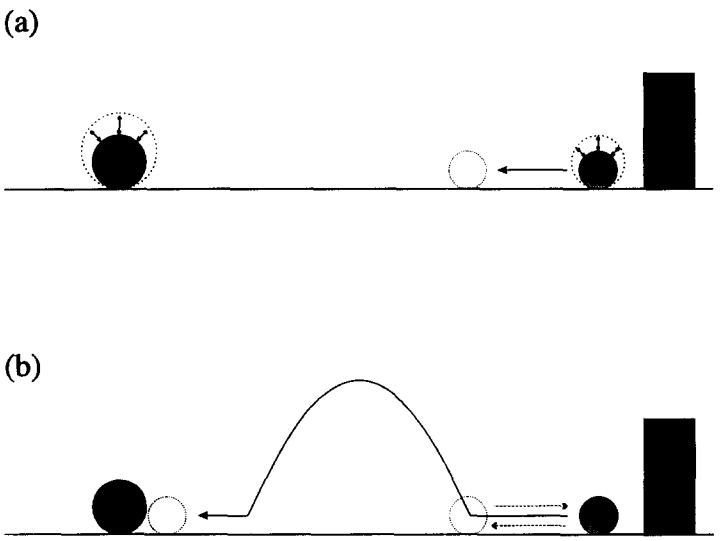

Fig. 2. Habituation events for the nonrational approach group. The sequence of the actions is identical to those for the rational approach group (see Fig. 1), but the rectangular figure is placed behind the small circle, rather than in between the two.

observed actions, just as in the case of Fig. 1, may induce the infant to attribute a goal to the small circle, for the event depicted in Fig. 2 it becomes very hard to coherently interpret the agent's behavior as a case of rational approach of the goal. This is so because in this situation there is a more rational means action available (i.e., approaching the goal through the shortest straight pathway leading to it) than the rather more complex and apparently unmotivated action (approach, retreat, and jump) that is actually performed by the agent. Therefore, we believe it to be more likely that, in the case of Fig. 2, the infant will abandon her[6] interpretation of the small circle's behavior as that of a rational agent.

EXPERIMENT

To test the above hypothesis concerning the infant's ability to evaluate the rationality of an agent's approach of a spatial goal, we performed a habituation study using two groups of 12-months-old infants as subjects. One group (rational approach group) was habituated to the sequence of events depicted in Fig. 1, while the other group (nonrational approach group) was presented with the event sequence illustrated in Fig. 2. Above it was argued that if the infant interpreted the visual event during the habituation trials as the goal-directed activity of a rational agent, when faced with a new situation she would be able to predict from the intentional stance the agent's most likely future behavior.

To test this hypothesis, following the habituation phase we presented both groups of subjects with two kinds of test events (Fig. 3) representing two different actions (Fig. 3a new action vs. Fig. 3b old action) the agent may perform in a new situation in which the rectangular figure (the "obstacle") *has been removed*. We are now ready to formulate predictions concerning the infants' expectations about the agent's future behavior in the new situation.

Let us consider first the rational approach (Fig. 1) condition. We hypothesized that since the agent's behavior in the habituation event satisfies both the conditions of equifinality and rationality of action, the infants observing this display will take the intentional stance and attribute the equifinal outcome of the agent's actions as the goal of the agent. When faced with the new situation, in which the obstacle has been removed (Fig. 3), the infant will be able to predict, based on the principle of rationality, that the small circle will approach its goal in the most rational manner now available; that is, through the shortest straight pathway leading to it, which also requires the least effort. This situation is depicted in Fig. 3(a) (new

[6] The authors' intention in using 'she' in reference to infants and 'he' to agents is to secure an equal division of anaforic labor for the pronouns.

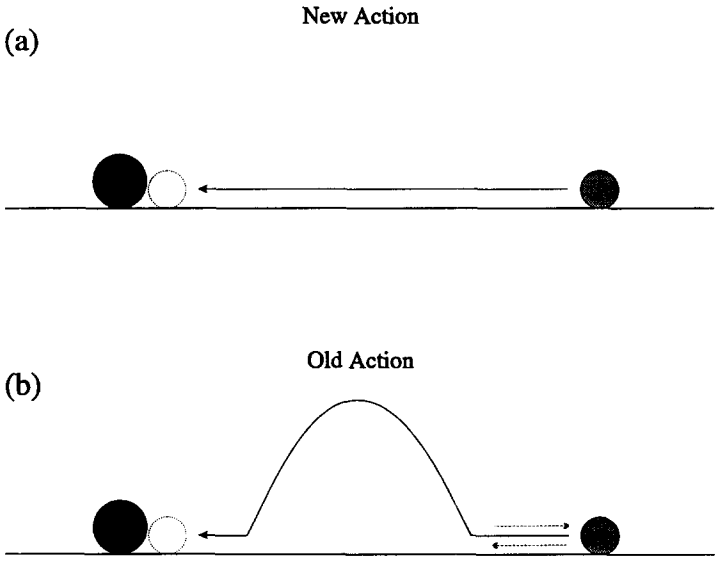

Fig. 3. Test events in the dishabituation phase. (a) The small circle approaches the large one through the shortest straight pathway (new action). (b) The small circle exhibits the same behavior as in the habituation phase (see Figs. 1 and 2) in the absence of the rectangular figure (old action).

action event), where instead of jumping as before, the agent approaches its goal through a new pathway: a horizontal straight line.

In contrast, the sight of the event depicted in Fig. 3(b) (old action) should be quite unexpected for the infant, *even though* the pathway of approach is identical to that observed during the habituation phase (Fig. 1). This is so because the complex (and causally unmotivated) approach behavior exhibited in the absence of the obstacle violates the rationality assumption of the intentional stance. Therefore, we predict that infants who are habituated to the rational approach condition (Fig. 1) will dishabituate significantly more to the nonrational jumping action depicted in Fig. 3b, than to the new but rational straight-line approach illustrated in Fig. 3(a). In fact, we expect these subjects to show little, if any, dishabituation to the latter kind of stimulus event as they can anticipate it from the intentional stance they have taken.

Note, furthermore, that if, contrary to our hypothesis, the infant does not interpret the habituation event as the intentional goal-directed activity of a rational agent, one could expect the opposite pattern of results on the basis of the relative degree of perceptual dissimilarity between the habituation stimuli versus the two types of dishabituation event. In such a case, one would predict relatively more dishabituation for the straight line approach depicted in Fig. 3(a), which is more dissimilar (being a new action) compared to the habituation stimulus than is the jumping approach of Fig.

3(b), where the small circle's actions are, in fact, identical to those in the habituation display.

Consider now the nonrational approach condition (Fig. 2). Here we hypothesized that since the agent's equifinal actions do not satisfy the requirement of rationality, the infant will abandon the intentional stance and will not consider the small circle to be a rational agent. Therefore, we predict that in the new situation (Fig. 3), where the rectangular figure is removed from behind the small circle, the infant will develop no specific expectations as to the most likely approach route that the small circle will follow. Thus, we expect that, unlike in the case of the rational approach habituation condition (Fig. 1), here the straight-line approach of Fig. 3(a) will *not* result in less dishabituation than the jumping approach of Fig. 3(b). A related prediction concerns the relative degree of dishabituation for the straight-line approach of Fig. 3(a) between the two groups of subjects: we expect less dishabituation for the event depicted in Fig. 3(a) in the rational approach condition (Fig. 1) than in the nonrational approach condition (Fig. 2).

6. Method

6.1. Subjects

We used as subjects infants who were brought in by their mothers for a medical examination to the Family Planning Service in Budapest. One hundred and twenty infants were randomly assigned to two groups: rational approach group and nonrational approach group. Due to fussiness, falling asleep, crying, and experimental errors 17 infants from the rational approach group and 27 from the nonrational approach group were rejected. A further 13 subjects from the rational approach and 11 from the nonrational approach group were excluded due to too short fixation times during the dishabituation trials (see section 6.3). This rather high rate of subject loss is likely to have been due to the fact that before they participated in the experiment, the infants had gone through a thorough medical examination which might have made a number of them too tired to complete the habituation study successfully. At the end, we were left with 30 subjects in the rational approach and 22 in the nonrational approach group. These infants were between 45 and 64 weeks old with a mean age of 52.8 weeks with a standard deviation of 4.5 weeks.

6.2. Stimuli

The stimulus events were presented on an 18 cm × 24 cm color monitor of a personal computer. Fig. 1 illustrates the habituation stimuli for the rational approach group, while Fig. 2 shows the habituation events pre-

sented to the nonrational approach group. The two kinds of habituation stimuli were identical in every respect except one: in the rational approach group the rectangular figure was placed in the middle of the screen, blocking the shortest pathway between the two circles on the two sides of the screen (forming an "obstacle"), while in the nonrational approach group the rectangular figure appeared behind the small circle near to the edge of the screen. The behavior of the two circles during the habituation events were identical in both conditions. First, the large circle expanded, then it contracted (regaining its original shape). This was immediately reciprocated by the same action carried out this time by the small circle. This "exchange" took 0.9 s and was then repeated a second time. After this the small circle started to move towards the large one but it stopped before reaching the middle of the screen. (In the rational approach condition (Fig. 1) this meant stopping in front of the "obstacle".) Then the small circle moved backwards, returning to its original position where it momentarily stopped. This forward–backward motion sequence took about 0.5 s. Finally, it started out again towards the large circle, but this time with faster speed culminating in a jump over the middle of the screen (following a parabolic trajectory), landing in front of the large circle, and then continuing its approach horizontally until the two circles made contact. (In the rational approach condition (Fig. 1) this involved jumping over the "obstacle" in the middle the screen.) The duration of this action was 1.4 s. Upon contact the two circles repeated their reciprocal expansion–contraction routine again (1.8 s).

Fig. 3 depicts the two types of test stimuli (a: new action; b: old action) which were the same for both the rational approach and the nonrational approach groups. Both kinds of test stimuli differed from the habituation stimuli in that the rectangular figure was absent. In the old action event (Fig. 3b) the behavior of the two circles was identical to that in the habituation conditions. The new action event (Fig. 3a), however, differed in that after the repeated exchanges of the expansion–contraction displays between the two circles, the small circle moved across the screen with constant speed through the shortest horizontal pathway leading to the large circle. Upon contact the two circles repeated their reciprocal expansion–contraction routine once again (as in the other conditions).

The large circle (1.39 cm diameter) was red, the small (0.94 cm diameter) was yellow, the rectangular figure (3.76 cm × 1.13 cm) was black, and the background was light green in all events. The horizontal velocity of the small circle's motion was 10 cm/s. Each event started with the simultaneous appearance of the two circles at the two sides of the screen (and with the rectangular figure in the middle or at the edge of the screen in the two habituation conditions, respectively). Each event lasted 5.50 s, then the figures disappeared from the screen. After a 1 s break the stimulus event started again. During the habituation phase the presentation of the event or its mirror image was randomly varied; that is, the small circle approached the large one from left to right or vice versa with equal frequency.

6.3. Procedure

The experiment was carried out in a darkened room. The babies sat in their mothers' lap looking at the monitor placed at eye level from a distance of 1 m. The monitor appeared in a window that was cut in a large black occluding screen, which ensured that the child's attention was not drawn to other objects in the room. The experimenter stood behind the screen throughout the experiment and watched the baby's face through a peephole cut in the screen right above the middle of the monitor to determine whether the baby was attending to the display or not. He controlled the presentation of the stimuli and the registration of looking times by pressing keys on the computer keyboard.

At the beginning of each trial the baby's attention was directed to the display when necessary by flashing lights on the monitor a few times and sounding a tone. When the baby looked at the monitor the experimenter started the presentation of the habituation stimulus event, which was repeated continuously until the subject looked away for more than 2 s. When the infant looked away, the experimenter hit a key on the keyboard, and if he did not signal within 2 s by hitting another key that the baby looked back again, the computer program stopped the timer and the stimulus display. After this the experimenter attracted the baby's attention to the monitor again by flashing the lights, and the next trial started.

The computer program averaged the fixation times for the first three habituation trials and compared this value on-line with the average of the last three fixation times. We applied a relatively stringent criterion of habituation: the average fixation time for the last three trials was required to be less than half of the average looking times for the first three habituation trials and this requirement had to be met twice in a row. Thus the minimal number of habituation trials was seven.

After the habituation criterion was reached a 30 s break was introduced during which the mother, who was sitting in a swivel chair, was asked to turn with her baby away from the monitor. When they turned back and the test trials started, we instructed the mothers to close their eyes so that they could not inadvertently bias their child's reaction to the dishabituation displays. The test trials were delivered in the same way as the habituation trials. Each subject saw four test trials: two new action events (Fig. 3a) and two old action events (Fig. 3b), with the two kinds of event presented in an alternating order. The experimenter was blind to the order in which the test stimuli were presented. For half of the subjects the first test trial was a new action display followed by an old action, while the other half of the subjects received the same stimuli in the opposite order. Independent of this, for half of the subjects the first two test trials showed the small circle approaching the large one from left to right, while the other half saw the approach taking place in the opposite direction. Thus, the order of presentation of event types and direction of approach was randomly intermixed throughout the test trials.

We wanted to ensure that the subjects' dishabituation scores reflect their reaction to the nature of the stimulus event, and so we had to make certain that they had a chance to identify which kind of event was presented to them. Therefore, since each event lasted for more than 5 s, to ensure that during the dishabituation trials the subjects were exposed to the full event structure, we excluded from the analysis all those subjects who watched either of the first two test trials for less than 4 s. The third and fourth test trials were left out of the analysis altogether for two reasons: (a) we were interested in the contrast between the new versus the old action events, and this comparison could be evaluated properly on the basis of the first two dishabituation trials alone; and (b) most subjects produced at least one fixation time below 4 s during the last two dishabituation trials.

7. Results

Fig. 4 shows the mean looking times of the first three and the last four habituation trials for the rational approach and the nonrational approach group. One-way ANOVAs indicated significant between-group differences in the average looking time for both the first three habituation trials ($F(1, 50) = 6.86$; $p < .05$) and the last three habituation trials ($F(1, 50) = 6.10$; $p < .05$). This result shows that the rational approach group looked longer at the habituation events than did the nonrational approach group.

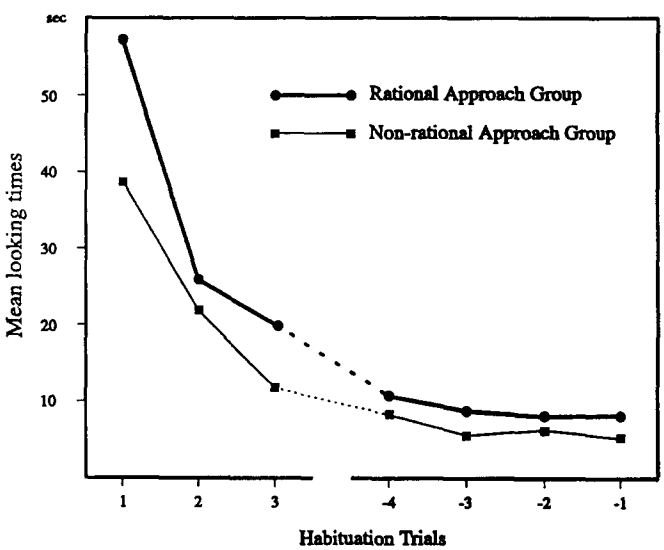

Fig. 4. Mean looking times of the rational approach and nonrational approach group in the first three and last four trials in the habituation phase. (Note that the last four habituation trials are numbered backward from the trial in which the habituation criterion was reached.)

However, the average number of habituation trials was similar in the two groups (7.70 and 7.64, respectively).

The analyses of the looking times in the test phase of the experiment were based on the dishabituation times defined as the length of stimulus fixation during the test phase *minus* the mean looking time of the last three habituation trials. In the ANOVAs three between-subject and one within-subject factors were used: the condition factor refers to the two kinds of habituation events (rational vs. nonrational approach); the order factor refers to the order of the test events (new first vs. old first); the side factor refers to the direction of the small circle's approach in the first two test events (right to left vs. left to right); while the within-subject event type factor refers to the two types of test events (new action vs. old action). Because a four-way ANOVA of the dishabituation times (Condition × Order × Side × Event type) did not result in either a side main effect or in any interaction including the side factor, this factor was eliminated from the further analyses.

Fig. 5 and Table 1 illustrate the mean dishabituation times for the first two test stimuli. The three-way ANOVA (Condition × Order × Event type) showed a significant three-way interaction ($F(1, 48) = 5.70$; $p < .05$). Therefore, we performed separate two-way ANOVAs for the two conditions. In the rational approach group the two-way ANOVA (Order × Event type) resulted in an event type main effect ($F(1, 28) = 7.97$; $p < .01$), which was

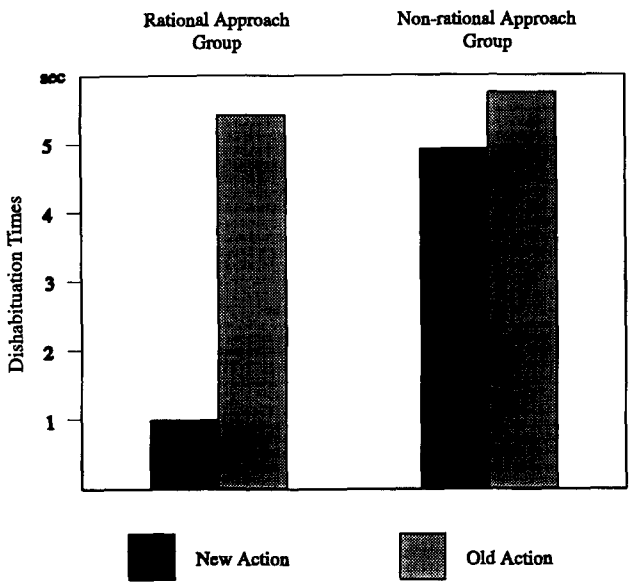

Fig. 5. Mean dishabituation times for the first two test events. Dishabituation times were calculated within-subjects as the difference of the looking time of the test trials and the mean looking time of the last three habituation trials.

Table 1
Mean dishabituation times (ms) to the test events. Subgroups "First new" and "First old" refer to the infants who were presented first with the new action and then with the old action test stimuli, or vice versa, respectively

	Group:	Rational approach		Nonrational approach	
	Subgroup:	First new	First old	First new	First old
Event:	New action	634	1263	8356	2963
	Old action	6831	3606	4752	6284

due to longer dishabituation times for the old action than for the new action event, irrespective of the order of these events (see Fig. 5). In the nonrational approach group a similar ANOVA showed a significant interaction ($F(1, 20) = 4.60$; $p < .05$) without any main effects. This result is due to the fact that in this group it was always the first test event that caused higher dishabituation (Table 1), irrespective of which event type it represented. Because the dishabituation times were not normally distributed, we confirmed these results by sign-tests. In the rational approach group 22 infants (73%) out of a total of 30 showed higher dishabituation for the old action than for the new action test event. This proportion is significant ($p < .05$). In contrast, in the nonrational approach group only 9 infants (41%) out of a total of 22 showed more dishabituation for the old action than for the new action event. This proportion is not significant.

To test our prediction concerning the relative amount of dishabituation for the new action event between the rational approach and the nonrational approach group, we performed a two-way (Condition × Order) ANOVA of the dishabituation times for the new action event. This revealed only a condition main effect ($F(1, 48) = 8.79$; $p < .01$), showing that infants in the nonrational approach group were dishabituated more than those in the rational approach group.

Finally, to evaluate the further prediction that the infants in the rational approach group would not dishabituate to the new action event, we performed a t-test which indicated that the mean dishabituation time (907 ms) to the event did not differ significantly from zero ($t(29) = 1.03$). In contrast, the mean dishabituation time (5434 ms) to the old action event in this group showed a significant difference ($t(29) = 3.17$; $p < .01$, two-tailed). In the nonrational approach group both the new and the old action events caused significant dishabituation (4924 ms, $t(21) = 3.45$ and 5727 ms, $t(21) = 3.15$, respectively; $p < .01$, two-tailed in both cases).

8. Discussion

In the Introduction we reviewed several types of converging behavioral evidence (such as the emergence of pointing and gaze alteration, or social

referencing) that are interpreted by numerous researchers as indicating the appearance at the end of the first year of the ability to attribute causal intentional states to others. While, as we pointed out, it is always possible to develop separate accounts in terms of simpler psychological processes for the different kinds of new behaviors in question, we agree that their convergence at the last quarter of the first year is highly suggestive, pointing at the possibility that it is during this developmental phase that the infant starts to adopt a mentalistic strategy to interpret and predict the behavior of other agents. In fact, we wish to argue that the results of the present habituation study provide independent empirical support for the general conjecture that by the end of the first year infants are indeed capable of taking the intentional stance (Dennett, 1987) in interpreting the goal-directed behavior of rational agents.

This conclusion is supported by the pattern of dishabituation responses (see Fig. 5) observed in the rational approach group whose subjects were habituated to the stimulus event depicted in Fig. 1. The structure of the visual habituation event in this condition fulfilled all the requirements that were hypothesized necessary to support an intentional causal analysis of the display: (a) the small circle exhibited several types of stimulus features (nonrigid transformation of surface, contingent reciprocal reactivity at a distance, self-propelled movement) that may indicate agency; (b) its actions resulted in an equifinal outcome which may provide a cue as to the goal of the agent; and (c) the sequence of actions performed by the small circle could be coherently interpreted as a case of rational approach of the goal.

We hypothesized that if the 12-month-old subjects were able to use these stimulus cues to interpret the small circle's actions as those of a rational agent with a specific intention, when faced with a new situation represented by the test stimuli, they would be able to generate an expectation about the most rational new means action the agent would be likely to perform to achieve its goal. That this was indeed the case is shown by the fact (see Fig. 5) that during the test phase the old action display (Fig. 3b) resulted in significantly more dishabituation than did the new action display (Fig. 3a).

Recall that in the two kinds of test stimuli presented to the rational approach group the obstacle blocking the shortest pathway between the agent and its goal was removed. In one case (Fig. 3a: new action) the agent approached its goal in this new situation through the shortest pathway that now became available to it; that is, through a straight line. In this way it performed a *new* action that was not witnessed by the subjects during the habituation phase. In the other case (Fig. 3b: old action) the agent approached the goal by performing exactly the same sequence of actions (the jumping event) as during the habituation phase. Therefore, the new action display was perceptually more dissimilar to the habituation stimulus than the old action display was. The fact that in spite of this subjects showed *less* surprise when seeing the new action display, indicates that from the intentional stance they could predict the new action of the agent as the most

rational means action it is likely to perform in the new situation. This conclusion is further supported by the fact that the amount of actual dishabituation for the new action display was, in fact, negligible (Fig. 5): fixation times for the new action stimulus were not significantly longer than those for the last three habituation stimuli. This then clearly indicates that the subjects in the rational approach group were not surprised to see the new means action performed by the agent in the new action display in spite of its perceptual dissimilarity to the habituation display.

One could, of course, wonder whether the larger dishabituation to the old action display is not due simply to its higher degree of perceptual complexity when compared to the new action display, rather than to its unpredictability from the intentional stance. Such an account is ruled out, however, by the pattern of dishabituation results found in the nonrational approach group in which subjects were habituated to the event depicted in Fig. 2. Since the two kinds of test stimuli in the nonrational approach group were identical to those in the rational approach group, the complexity account would predict higher dishabituation to the old action display in this condition as well. In contrast, there was no significant difference in dishabituation between the new versus the old action displays in the nonrational approach group.

In fact, in the nonrational approach group it was the order of presentation of the dishabituation stimuli, and not their type, which had an effect on the degree of dishabituation: it was the display presented first, irrespective of its type, which resulted in more dishabituation as indicated by the significant Order × Event type interaction. This order effect (see Table 1) may have been due to two factors: (a) the 30 s break and the turning away from the display may have facilitated an initial recovery of interest to the test stimuli which (b) at the same time showed some degree of perceptual dissimilarity in comparison to the habituation stimuli. No such order effect was observable in the rational approach group due to the presence of the active expectation that infants generated concerning the most likely means action on the part of the agent in the new situation.

Therefore, it seems safe to conclude that the differential degree of dishabituation between the old versus the new action displays found in the rational approach group cannot be attributed to the higher perceptual complexity of the old action stimulus. In fact, it seems clear that the difference is due to the lack of dishabituation for the new action display which could be predicted as the most likely new means action from the point of view of the intentional stance that subjects have taken during the habituation trials. In contrast, the subjects in the nonrational approach group were habituated to events (Fig. 2) in which the equifinal activity of the agent does not meet the requirement of rationality of action that was hypothesized necessary for the intentional stance. Therefore, we predicted that in this condition the infants would have to abandon the intentional stance and, consequently, would not generate specific expectations about the most likely future actions of the (nonrational) agent. This prediction is

borne out by the finding that the new action display resulted in significantly more dishabituation in the nonrational approach group than in the rational approach group.

Thus, we can conclude that the difference in the degree of dishabituation evoked by the new action display in the two groups is fully attributable to the consequences of the *differential interpretation* that subjects have imposed on the habituation events. An unanticipated aspect of the results provide further support for this view. As Fig. 4 shows, the subjects of the rational approach group looked significantly longer at the habituation events (Fig. 1) than did the subjects in the nonrational approach group (Fig. 2). This difference in overall fixation times is clearly in line with our hypothesis that in the case of the rational approach condition the small circle's behavior was interpreted from the intentional stance as the goal-directed actions of a rational agent, while when watching the nonrational approach display infants had to abandon the intentional stance as the behavior of the agent did not meet the requirements of rationality.

The differential results in our rational versus nonrational conditions also have some bearing on recent modularist proposals which suggest that "the perception of intention ... is a hard-wired perception based ... on appropriate stimulation" (Premack, 1990, p. 2). The empirical question that arises in relation to such a hypothesis of a modular "intentionality detector" (Baron-Cohen, 1994) concerns the specification of the stimulus input that triggers the attribution of an intention. For example, building on Premack's original proposal, Baron-Cohen (1994) suggested that apart from self-propelled movement, stimulus direction also triggers the attribution of goal (where "goal" is defined as "the target an action is directed towards", p. 516).[7] Also, Premack has recently further elaborated his modularist theory, suggesting that there are three features that are essential to the attribution of goal to a self-propelled object: (a) "all the object's actions must be directed at the same item"; (b) "the object must repeat its action"; and (c) "the object's repeated acts must not be repeated perfectly" (Premack & Premack, 1994, p. 151). Furthermore, according to Premack and Premack (1995) "infants attribute a goal to objects that they consider intentional when such objects ... contact another intentional object" (p. 209).

It is clear, however, that these proposals cannot easily accommodate the differential results in our rational versus nonrational approach conditions. Note that although *all* of the stimulus cues suggested above characterized both of our experimental groups equally, it was only in the rational approach condition that we found evidence for goal attribution. Therefore, we believe that while the proposed cues might indeed be relied on by the interpretative processes that attempt to ascribe intentional content to an agent, they, nevertheless, do not seem to be mandatory, as shown by the fact that when the spatial approach of the agent does not meet the

[7] See Gergely and Csibra (1994) for critical arguments concerning this proposal.

requirements of rationality, the intentional analysis of the object's behavior is abandoned.[8]

At this point we must address the thorny question of how to characterize the requirements for rationality of action. The principle of rationality, that we proposed to be part of the infant's theory of agency, assumes that whenever there are multiple available means actions that an agent could perform to achieve his goal, he will choose to carry out the one that is the most rational; that is, the action that will let him achieve his aim in the most optimal manner relative to a set of background conditions. In general, then, an observer will evaluate the rationality of an action in relation to a goal as a function of her assumptions concerning the agent's beliefs about the situation, his available repertoire of means, his current state of resources, the relative importance of his goal in relation to other, possibly conflicting, priorities, etc. (Dennett, 1987; Fodor, 1992). In the case of infancy, the set of background conditions is likely to be much more restricted: we hypothesized that they consist of the infant's knowledge of the physical, causal, gravitational and biomechanical constraints on the spatial movements of objects and agents.

When generating our hypothesis that the new action, but not the old action, test display would be evaluated by the infant as rational, we based our judgement on our *own* intuition without specifying an explicit algorithm that the infant might apply to compute the rationality of the action on the basis of the assumed background conditions. Dennett (1987) has argued that "the concept of rationality is systematically pre-theoretical" and "when one leans on our pre-theoretical concept of rationality, one relies on our shared intuitions – when they *are* shared, of course – about what makes sense" (p. 98). It is, of course, an empirical question whether our adult intuitions about the rationality of an agent's spatial goal-approach are shared by the infants or not. Recently, several theorists of cognitive development (e.g., Fodor, 1992; Leslie, 1988b; Spelke et al., 1992) have argued that "young infants' reasoning accords with principles at the center of mature, common-sense conceptions" (Spelke et al., 1992, p. 606). If this view is correct, and the domain-specific, core architecture of the initial state theories of the physical and social world that characterize the infant's competence, also forms the foundation of later, more enriched and refined adult conceptual structure, then one may expect to find that adult intuitions about rationality of spatial goal-approach might coincide with those of an infant. Therefore, one may argue that the fact that our intuitively based prediction concerning the rationality of the new action test display has been borne out by the infant data, is in line with the general view of the nature of cognitive development

[8] Note that from our study it cannot be ascertained that the evaluation of rationality of spatial approach is a function of the previous categorization of the moving object as an intentional agent. In fact, it is conceivable that rationality of approach is evaluated independently and can itself act as a cue to agency.

which holds that "initial conceptions form the core of many later conceptions" (Spelke et al., 1992, p. 606).

In closing, we wish to discuss three unresolved theoretical issues concerning our claim that 12-month-old infants take the intentional stance when perceiving the goal-directed actions of a rational agent. First, we would like to make explicit an inherent ambiguity in our interpretation of the results of our study. Up until now we have used descriptions such as: to predict or explain the other's behavior "the infant attributes a causal intention to the agent" versus "the infant takes the intentional stance towards the agent" interchangeably. However, behind these uses there may lie two distinct empirical possibilities. Strictly speaking, what our study implies is not more than the fact that the 1-year-old causally interprets the actions of the agent in relation to a goal identified on the basis of the equifinal outcome of the agent's observed actions. In other words, by taking the intentional stance the infant can come to represent the agent's action as intentional without actually attributing a mental representation of the future goal state to the agent's mind. This more conservative interpretation of our results seems to correspond to the kind of representation of the actional properties of agents that is generated by the representational system at the second level of Leslie's recent tripartite theory of Agency (Leslie, 1994). Thus, although the present findings do not rule out the possibility that the infant might attribute the goal state to the agent's mind as a causal intention, they seem sufficiently explained by the hypothesis that the infant applies a paradigm of "teleological causality" (Leslie, 1993) to interpret the action of the perceived agent. In this case the agent's action is represented as causally related to the future goal state without specifying the mechanisms that mediate between this state and the actual behavior.

The second theoretical issue concerns the status of the principles and constraints embodied in the infant's theories of physics and agency, that we suggested structure the domain over which she evaluates the rationality of the agent's goal-oriented actions. We hypothesized that these assumptions about the nature of objects and agents play the same *functional role* in evaluating the rationality of action in the restricted domain of spatial movement of agents, as do beliefs later in the child's theory of mind after 3 years of age (e.g., Gopnik, 1993). There are two points we would like to clarify in this regard.

First, just as in the case of goals discussed above, we do not believe that our results necessarily imply that the infant actually attributes to the other's mind the principles of her naive theories of physics and agency as mentally represented knowledge structures. While such an attribution might, in fact, take place, it seems equally possible that the infant simply assumes that a rational agent's behavior is *constrained* by the principles embodied in his naive theories of reality. Clearly, such an assumption would be sufficient to enable the infant to evaluate the rationality of the agent's equifinal behavior. A related point we wish to emphasize is that the constraints

embodied in the infant's naive theories of the world can function similarly to later beliefs in providing background assumptions for evaluating the rationality of an action, *without* having to have the status of a truth-functional propositional attitude concept as later beliefs have (e.g., Leslie, 1987, 1988a).

The third issue we would like to address concerns a further ambiguity in our empirical demonstration of the presence of the intentional stance at one year of age. Up until now we freely referred to the process underlying the lack of dishabituation to the new action event of our rational approach group as being due to the presence of a *prediction* generated from the intentional stance about the most likely future action that the rational agent is expected to perform in the new situation. Again, we wish to make clear that the observed lack of dishabituation may be caused by two distinct empirical processes.

On the one hand, it is clearly possible that as soon as the subject perceives the new situation (i.e., the disappearance of the obstacle) represented in the test display, he will proceed to generate from the intentional stance a predictive mental representation about the most rational *future* action that the agent is likely to perform. In this case, the lack of dishabituation to the new action event can be interpreted as due to the *match* between the predictive action representation generated and the actual action perceived.

On the other hand, the lack of a surprise reaction can also be explained as being due to a process of "retrospective integration" that takes place *after* the test event has been perceived. According to this interpretation, the infant does not project future states of affairs; rather, when faced with a new event involving previously experienced participants, she will attempt to integrate it into her already existing interpretation of the nature of the situation. In other words, having interpreted from the intentional stance the habituation event as the goal-directed behavior of a rational agent, when seeing the novel action of the agent, the subject will attempt to extend her previous interpretation to account in a coherent manner for the new event as well. If she succeeds in integrating the new event retrospectively into the framework of her previous interpretation, there will be no surprise reaction while observing the new event. However, if the test event involves a nonrational action on the part of the agent (as in the case of the old action test event), the attempt to extend the previous interpretation of the (rational) agent's behavior to the new event will fail, as the new action cannot be construed as that of a rational agent. This failure to keep up the continuity of the previous interpretation by coherently integrating the new event into it will then result in the dishabituation to the old action test display. In general, whether dishabituation results can be interpreted as being due to predictive inferencing or retrospective integration seems to us an open empirical question awaiting further research to be resolved.

In conclusion: the present study provides evidence that already at 1 year of age preverbal infants can take the intentional stance in interpreting the

goal-directed spatial behavior of a rational agent. Thus, based on the equifinal structure of the agent's behavior 12-month-old babies could identify the agent's goal and analyze its actions causally in relation to it. In particular, our study demonstrates that infants of this age are able to evaluate the rationality of an agent's goal-directed actions, at least, within the representational domain of spatial pathways through which agents move in relation to other objects in space. Dennett (1987) has argued that at the core of the intentional stance lies the basic notion of *rationality* which he proposes to keep "at the foundation of belief and desire attribution" (p. 94). The present findings from infancy provide additional empirical support for this strong emphasis on the foundational role of rationality in the intentional analysis of behavior.

Acknowledgements

This research was supported by grant #4692 from the National Research Foundation of Hungary (OTKA). We wish to express our thanks to Richard Aslin, András Vargha and John S. Watson for their valuable comments, and to Endre Czeizel and the Family Planning and Genetic Counseling Centre in Budapest for their support.

Reference

Anscombe, E. (1957). *Intention.* Oxford: Blackwell.

Astington, J.W., & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology, 9,* 7–31.

Astington, J.W., Harris, P.L., & Olson, D.R. (Eds.) (1988). *Developing theories of mind.* New York: Cambridge University Press.

Baillargeon, R. (1991). The object concept revisited: new directions in the investigation of infants' physical knowledge. In H.W. Reese (Ed.), *Advances in child development and behavior* (Vol. 23). New York: Academic Press.

Baron-Cohen, S. (1991). Precursors to a theory of mind: understanding attention in others. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 233–251). Oxford: Basil Blackwell.

Baron-Cohen, S. (1994). How to build a baby that can read minds: cognitive mechanisms in mindreading. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 13* (5), 513–552.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition 21,* 37–46.

Bates, E. (1979). Intentions, conventions and symbols. In E. Bates, L. Benigni, I. Bretherton, L. Camaioni, & V. Volterra (Eds.), *The emergence of symbols* (pp. 69–140). New York: Academic Press.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly, 21,* 205–226.

Bertenthal, B.I. (1993). Infants' perception of biomechanical motions: intrinsic image and knowledge-based constraints. In C. Granrud (Ed.), *Visual perception and cognition in infancy: Carnegie-Mellon symposia on cognition.* Hillsdale, NJ: Erlbaum.

Bertenthal, B.I., Proffitt, D.R., Spetner, N.B., & Thomas, M.A. (1985). The development of infant sensitivity to biomechanical motions. *Child Development, 56*, 531–543.

Bretherton, I. (1984). Social referencing and the interfacing of minds: a commentary on the views of Feinman and Campos. *Merrill-Palmer Quarterly, 30*, 419–427.

Bretherton, I. (1991). Intentional communication and the development of an understanding of mind. In D. Frye & C. Moore (eds.), *Children's theories of mind* (pp. 49–75). Hillsdale, NJ: Erlbaum.

Bretherton, I., & Bates, E. (1979). The emergence of intentional communication. In I.C. Užgiriš (Ed.), *Social interaction and communication during infancy*. San Francisco, CA: Jossey Bass.

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: the acquisition of an explicit theory of mind. *Developmental Psychology, 6*, 906–921.

Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication: when do infants acquire a "theory of mind"? In M.E. Lamb & L.R. Sherrod (Eds.), *Infants' social cognition*. Hillsdale, NJ: Erlbaum.

Bruner, J.S. (1975). The ontogenesis of speech acts. *Journal of Child Language, 2*, 1–19.

Butterworth, G.E., & Grover, L. (1990). Joint visual attention, manual pointing and preverbal communication in human infancy. In M. Jeannerod (Ed.), *Attention and performance* (Vol. XIII, pp. 605–624). Hillsdale, NJ: Erlbaum.

Butterworth, G.E., & Jarrett, N. (1991). What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology, 9*, 55–72.

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

Carlson, V. (1980). *Differences between social and mechanical causality in infancy*. Paper presented at the International Conference on Infant Studies. New Haven, CT.

Dennett, D.C. (1987). *The intentional stance*. Cambridge MA: Bradford Books/MIT Press.

Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press/Bradford Books.

Fodor, J.A. (1992). A theory of the child's theory of mind. *Cognition, 44*, 283–296.

Gelman, R., & Spelke, E.S. (1981). The development of thought about animate and inanimate objects: implication for research on social cognition. In J.H. Flawell & L. Ross (eds.), *Social cognitive development: Frontiers and possible futures*. London: Cambridge University Press.

Gergely, G., & Csibra, G. (1994). On the ascription of intentional content. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 13*, no. 5, 584–589.

Gibson, E.J., Owsley, C.J., & Johnston, J. (1978). Perception of invariants by five-month-old infants: differentiation of two types of motion. *Developmental Psychology, 14*, 407–416.

Golinkoff, R.M. (1983). The preverbal negotiation of failed messages. In R.M. Golinkoff (Ed.), *The transition from prelinguistic to linguistic communication* (pp. 57–78). Hillsdale, NJ: Erlbaum.

Gopnik, A. (1993). How we know our minds: The illusion of 1st-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*, 1–14.

Gopnik, A., & Wellman, H.M. (1994). The theory theory. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243–259.

Kim, I.K., & Spelke, E.S. (1992). Infants' sensitivity to effects of graivity on visible object motion. *Journal of Experimental Psychology: Human Perception and Performance, 56*, 531–543.

Klinnert, M.D., Campos, J.J., Sorce, J.F., Emde, R.N., & Svejda, M. (1983). Emotions as behavior regulators: social referencing in infancy. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience* (Vol. 2). New York: Acedemic Press.

Leslie, A.M. (1982). The perception of causality in infants. *Perception*, *11*, 173–186.

Leslie, A.M. (1987). Pretense and representation: the origins of "theory of mind". *Psychological Review*, *94*, 412–426.

Leslie, A.M. (1988a). Some implications of pretense for mechanisms underlying the child's theory of mind. In J.W. Astington, P.L. Harris, & D.R. Olson (Eds.), *Developing theories of mind* (pp. 19–46). New York: Cambridge University Press.

Leslie, A.M. (1988b). The necessity of illusion: perception and thought in infancy. In L. Weiskrantz (Ed.), *Thought without language* (pp. 185–210). Oxford: Clarendon Press.

Leslie, A.M. (1993). *A theory of agency*. Rutgers University Center for Cognitive Science, Technical Report TR-12.

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

Leslie, A.M., & Happé, F. (1989). Autism and ostensive communication: the relevance of metarepresentation. *Development and Psychopatology*, *1*, 205–212.

Leslie, A.M., & Keeble, S. (1987). Do six-month-olds perceive causality? *Cognition*, *25*, 265–288.

Mandler, J.M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, *99*, 587–604.

Murphy, C.M., & Messer, D.J. (1977). Mothers, infants and pointing: a study of a gesture. In H.R. Schaffer (Ed.), *Studies in mother–infant interaction*. London: Academic Press.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., Leekam, S.R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief. *British Journal of Developmental Psychology*, *5*, 125–137.

Poulin-Dubois, D., & Shultz, T.R. (1988). The development of the understanding of human behavior: from agency to intentionality. In J.W. Astington, P.L. Harris, & D.R. Olson (Eds.), *Developing theories of mind*. New York: Cambridge University Press.

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

Premack, D., & Premack, A.J. (1994). Moral belief: form versus content. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Premack, D., & Premack, A.J. (1995). Origins of human social competence. In M.S. Gazzaniga (Ed.), *The cognitive neurosciences*. Cambridge, MA: MIT Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526.

Shatz, M. (1983). Communication. In J.H. Flawell & E.M. Markman (eds.), *Handbook of child psychology. Vol. 3: Cognitive Development* (pp. 841–889). New York: Wiley.

Spelke, E.S. (1985). Preferential looking methods as tools for the study of cognition in infancy. In G. Gottlieb & N. Krasnegor (eds.), *Measurement of audition and vision in the first year of post-natal life*. Hillsdale, NJ: Erlbaum.

Spelke, E.S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Spelke, E.S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605–632.

Stern, D.N. (1985). *The interpersonal world of the infant*. New York: Basic Books.

Trevarthen, C. (1977). Descriptive analyses of infant communicative behavior. In H.R. Schaffer (Ed.), *Studies in mother–infant interaction*, London: Academic Press.

Watson, J.S. (1979). Perception of contingency as a determinant of social responsiveness. In E. Thoman (Ed.), *The origins of social responsiveness* (pp. 33–64). Hillsdale, NJ: Erlbaum.

Watson, J.S. (1984). Memory and learning: analysis of three momentary reactions of infants. In R. Kail & N. Spear (Eds.), *Comparative perspectives on the development of memory* (pp. 159–179). Hillsdale, NJ: Erlbaum.

Wellman, H.M. (1990). *The child's theory of mind*. Cambridge, MA: Bradford Books/MIT Press.

Wellman, H.M. (1991). From desires to beliefs: acquisition of a theory of mind. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 19–38). Oxford: Basil Blackwell.

Wellman, H.M., & Woolley, J.D. (1990). From simple desires to ordinary beliefs: the early development of everyday psychology. *Cognition, 35*, 245–275.

Whiten, A. (1991). The emergence of mindreading: steps towards an interdisciplinary enterprise. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 319–331). Oxford: Basil Blackwell.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Wimmer, H., & Weichbold, V. (1993). *Children's theory of mind: Fodor's heuristics or understanding informational causation*. Unpublished manuscript, University of Salzburg.