Teleological reasoning in infancy: the naïve theory of rational action

György Gergely[1] and Gergely Csibra[2]

[1]Institute for Psychological Research, Hungarian Academy of Sciences, Budapest, Hungary
[2]Centre for Brain and Cognitive Development, Birkbeck College, London, UK

Converging evidence demonstrates that one-year-olds interpret and draw inferences about other's goal-directed actions. We contrast alternative theories about how this early competence relates to our ability to attribute mental states to others. We propose that one-year-olds apply a non-mentalistic interpretational system, the 'teleological stance' to represent actions by relating relevant aspects of reality (action, goal-state and situational constraints) through the principle of rational action, which assumes that actions function to realize goal-states by the most efficient means available. We argue that this early inferential principle is identical to the rationality principle of the mentalistic stance – a representational system that develops later to guide inferences about mental states.

The evolutionary and ontogenetic origins of 'theory-of-mind' [1,2], our ability to explain and predict others' actions by attributing causal intentional mental states (beliefs, desires and intentions) to them (Box 1) have been at the center of interest in a wide range of fields within cognitive science including philosophy of mind, cognitive neuroscience, developmental psychology, artificial intelligence, robotics and evolutionary psychology. The full-fledged emergence of taking such a 'mentalistic' or 'intentional stance' [3] seems a relatively late developmental achievement: its clearest indicator (attributing false beliefs) emerging only around 4 years of age [4–6].

By contrast, recent research demonstrated a surprisingly sophisticated understanding of intentional, goal-directed actions already by the end of the first year [7–13]. Therefore, the question of how to explain this competence and how to account for the developmental gap between its early appearance and the later emergence of the mentalistic stance have become central and controversial issues of the field. In this opinion paper our aim is twofold: (a) we shall outline two general types of approaches that have been proposed to 'bridge the gap' (for reviews, see [14,15]), and (b) we shall contrast these with an alternative approach: the one-year-old's 'naïve theory of rational action' or the 'teleological stance' [16,17]. We shall summarize the supporting evidence our theory has

Box 1. Practical reasoning

As adults we routinely take the 'mentalistic stance' when interpreting others' actions in terms of intentional mental states we infer and attribute to the actor's mind. These cases of practical reasoning relate three kinds of mental states: **beliefs**, **desires** and **intentions**, with the help of the 'rationality assumption': given information about any two of the three elements of mentalistic action interpretations, one can infer (and predict) what the third element *ought* to be. Let's consider an example for each of the three types of inference:

Inferring intentions

Tom, a blind man, has learned to make a detour around the dinner table whenever he wants to get to the kitchen. Yesterday Sylvia had the table taken away for repairs, but forgot to tell Tom about this. Today she saw Tom enter the dining room announcing he wanted to go to the kitchen. Sylvia inferred and attributed to Tom the false belief that the table was still in the room. Knowing Tom's desire (to get to the kitchen) and his false belief about the constraints of the situation, she could infer Tom's *intention*: to make a detour around the (missing) table to realize his desired goal. Note that Tom's intention to perform the detour, although specifying a goal-approach that was inefficient in actual reality, was nevertheless explicable (and predictable) as an intention to carry out a justifiable and rational goal-directed action within the constraints of the counterfactual fictional world represented as true by his false belief.

Inferring beliefs

Mari, my assistant usually walks to work passing over the Danube through Margaret Bridge. Today, however, I saw her arrive from the opposite direction suggesting she took the longer route through Árpád Bridge. Knowing her desire (to arrive on time) and seeing her action (taking the longer route), I infer that Margaret Bridge is closed. By attributing to her a *belief* about this situational constraint that could rationalize her intention to walk through Árpád Bridge as the most sensible means available to realize her desired goal, I could justify her taking the longer route as a rational goal-directed action.

Inferring desires

Peter is cooking a pot of stew. His gaze shifts to his empty glass, then he unscrews a bottle of wine. What *desire* could Peter have that he intended to satisfy by opening the bottle? His glance to the empty glass leads me to attribute Peter the desire to drink some wine. This desire would justify his intention to open the bottle, as doing so seems a justifiable means towards realizing his goal of drinking wine. Note that the inference specifying the *content* of the desire was guided by an informative aspect of the situation in which his intention to open the bottle was carried out:, namely, that the glass he glanced at was empty. If the glass had been half filled with beer, I would have attributed to Peter a desire with a *different* content to justify his intention to open the bottle: say, that his goal was to add wine to the stew.

Corresponding authors: György Gergely (gergelyg@mtapi.hu), Gergely Csibra (g.csibra@bbk.ac.uk).

http://tics.trends.com 1364-6613/03/$ - see front matter doi:10.1016/S1364-6613(03)00128-1

generated [7,10,13,18] and outline our own proposal to explain the 'gap', spelling out how and to what degree the proposed non-mentalistic teleological interpretative system of the one-year-old is related to the later emerging mentalistic stance.

Interpretation of goal-directed actions by young infants

Early understanding of goal-directed actions has been demonstrated using a variety of paradigms: imitation [8,18–20], joint attention [8,11], and violation-of-expectation looking time studies [7,9,10,12,13]. Let us illustrate the complex nature of this understanding by one of our violation-of expectation studies [7]. Twelve-month-olds were habituated to a computer-animated goal-directed event (Fig. 1a) in which a small circle approached and contacted a large circle ('goal') by jumping over ('means act') an obstacle separating them ('situational constraint'). During the test phase we changed the situational constraint by removing the obstacle. Infants then saw two test displays: the same jumping goal-approach as before, or a perceptually novel straight-line goal-approach. They looked longer (indicating violation-of-expectation) at the old jumping action (maybe because it seemed to them an inefficient means to the goal now that there was no obstacle to jump over), but showed no dishabituation to the novel straight-line goal-approach (possibly because this action appeared to them the most efficient means to the goal in the new situation). Such results indicate that by 12 months infants can (1) interpret others' actions as goal-directed, (2) evaluate which one of the alternative actions available within the constraints of the situation is the most efficient means to the goal, and (3) expect the agent to perform the most efficient means available.

How is this early capacity related to the later emerging theory-of-mind?

Alternative 1. One-year-olds already take the mentalistic stance

One dominant approach has been to argue that there is, in fact, no qualitative 'gap' to explain as the one-year-olds' ability to interpret goal-directedness actually indicates an already genuinely mentalistic understanding of actions. According to this view [21], infants in the above study attributed to the small circle a *desire* to get to the large circle and a *belief* about the impenetrability of the obstacle. Exemplifying this approach are recent 'modularist' [22–25] and 'simulationist' [11,26] theories that both propose (different) innate mechanisms through which young infants can identify and attribute specific causal intentional mind states to interpret the actions of others.

Briefly, according to modularist theories mental state attributions are driven by innate stimulus cues (such as self-propulsion [22], or direction of gaze or movement [23]) that activate a prewired triggering mechanism whose direct output identifies the specific content (e.g. a future goal state) represented by the intentional mental state (a desire) that is attributed to the actor. By contrast, in simulationist theories the hypothesized mechanisms whereby the other's mental states and their represented contents are identified are processes of identification and/or imitation. Through these infants 'put themselves into the actor's place' and internally generate (simulate) those intentional mental states that they would have, were they acting like the actor. These subjective mental states are then introspectively accessed and attributed, by analogy, to the actor's mind.

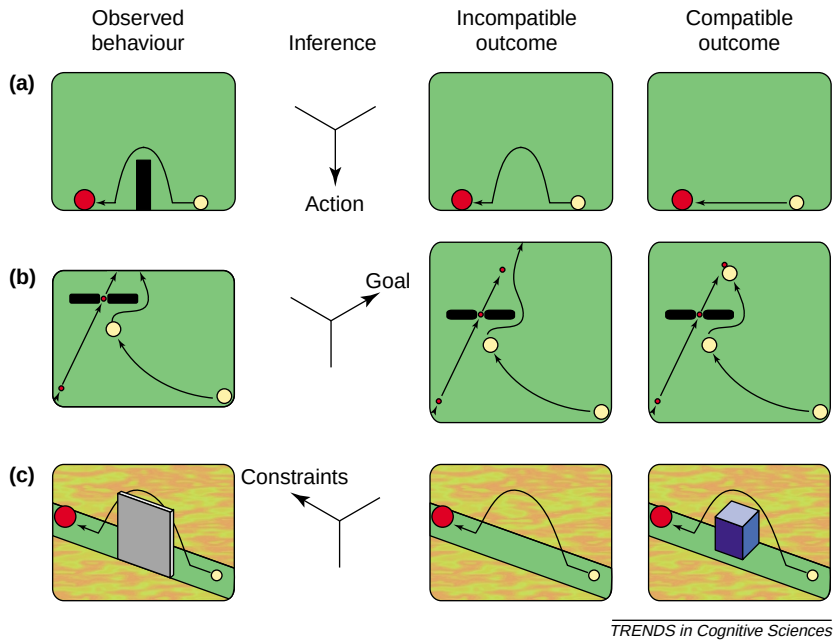

Fig. 1. Three types of inference that infants can draw based on a teleological representation of actions. One-year-old infants were habituated to the event depicted in the first column (Observed behaviour). Their interpretation of this event was tested by presenting them with two different outcomes, one of them being incompatible (second column), the other one being compatible (third column) with a possible inference based on a teleological representation of the event. Infants looked longer at the incompatible outcome than the compatible outcome events, indicating that they had made the assumed inference. (a) From the study in Ref. [7] and Experiment 1 in [10]. (b) Experiment 1A in [13]. (c) Experiment 2 in [13].

Alternative 2 The teleological stance in one-year-olds

According to our alternative proposal, one-year-olds can represent, explain and predict goal-directed actions by applying a non-mentalistic, reality-based action interpretational system, the 'teleological stance' [16,17]. This interpretational schema establishes a teleological (rather than causal [17]) explanatory relation among three relevant aspects of current and future reality: the *action*, the (future) *goal state*, and the current *situational constraints* (Fig. 2). Thus, in contrast to the modularist and simulationist accounts outlined above, by applying the teleological stance young infants can interpret goal-directed actions *without* attributing intentional mental states to the actor's mind. Rather, teleological action explanations make reference to the relevant aspects of reality as those are represented by the interpreting infant herself when observing the action unfold in its situational context.

Our second major disagreement concerns the type of *mechanism* that young infants apply to identify the specific aspects of current and future world states in terms of which they explain others' goal-directed actions. Although we agree that innate triggering cues and simulation processes might play some non-negligible role in this process, we argue that they are not sufficient in themselves to account for interpreting intentional actions [10,14]. By contrast, as philosophers [3] have argued persuasively, when taking the mentalistic stance the actor's mental states are typically *inferred* through the application of the 'rationality principle' that functions as the central inferential component of theory-of-mind. In fact, we shall propose that the same principle of rational action also forms the core inferential component of the non-mentalistic teleological stance of the one-year-olds.

The functions of the rationality principle in the mentalistic and teleological stance

The rationality principle captures our normative assumptions about the essentially functional nature of intentional actions. It serves both as a criterion of 'well-formedness' for mentalistic action interpretations and as an 'inferential principle' guiding and constraining the construction of such action interpretations (Box 1). In particular, the principle of rational action presupposes that (1) *actions* function to bring about future *goal states*, and (2) goal states are realized by the most rational action available to the actor within the *constraints of the situation*. Thus, the principle asserts that a mentalistic action explanation is well-formed (and therefore acceptable) if, and only if, the action (represented by the agent's *intention*) realizes the goal state (represented by the agent's *desire*) in a rational manner within the situational constraints (represented by the agent's *beliefs*) (Fig. 2).

In fact, the central insight that has guided our theorizing about the functional viability of a non-mentalistic teleological stance was the realization that when taking the mentalistic stance the rationality principle is always applied to the *contents* that the actor's mental states represent, and *not* to the intentional mind states themselves. Note that these mentally represented contents specify the relevant aspects of current and future states of reality in relation to which the efficiency of an action as means to a goal is evaluated. In fact, the rationality principle is applied to these specified aspects of

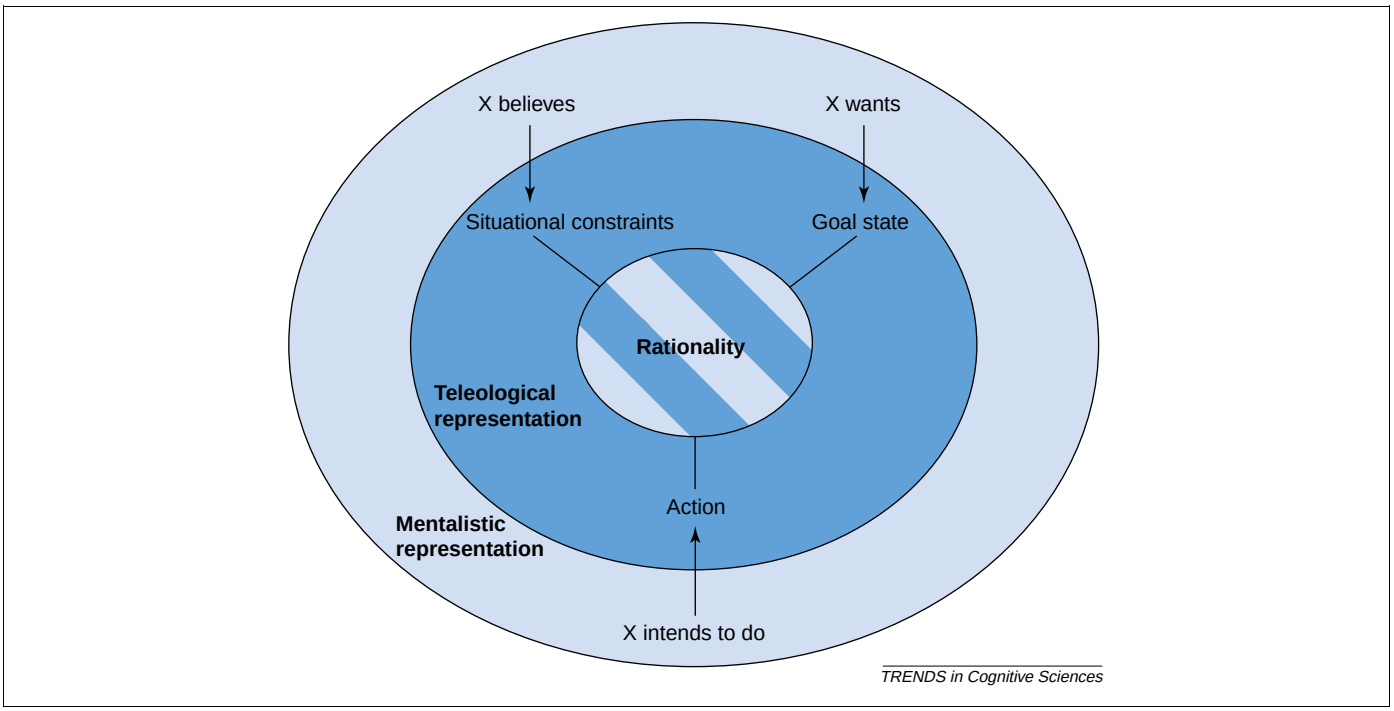

Fig. 2. Teleological and mentalistic representations of actions. Teleological representations relate three aspects of the real world to each other via the rationality principle, which provides explanations and predictions for observed actions. Mentalistic action representations involve three types of intentional mental states attributed to an agent (X). The contents of these mental states correspond to the elements of the teleological representations. There are several differences between these action explanations, including the direction of the explanation (causal versus teleological), or the ontological status of the elements (real versus fictional worlds). Note, however, that the principle of rational action applies equally to both kinds of representation.

reality irrespective of whether they correspond to actual reality (as contents of true beliefs do), or whether they correspond to counterfactual or just fictional realities (as do contents of false beliefs or pretense, respectively).

This raises the possibility of a representationally less sophisticated organism that – although unable to represent intentional mental states – could nevertheless have evolved a reality-based interpretational strategy to represent goal-directed actions. This 'mindblind' creature could represent its normative assumptions about the essentially teleological nature of actions in terms of the very same inferential principle of rational action that forms the central component of the mentalistic stance. The organism could then evaluate the efficiency of an observed action as a means to a goal by applying the rationality principle to the relevant aspects of reality states, as they are specified by the organism's own representations formed while perceiving the action unfold. Note that this teleological evaluation of the efficiency of means should provide the same results as the application of the mentalistic stance as long as the actor's action is driven by true beliefs. The teleological interpretation would break down, however, if the interpreted action were based on pretence or false beliefs. This would be so because our 'mindblind' interpreter could not represent the mental states of the actor that could specify the relevant fictional or counterfactual situations within which the agent's action should be evaluated. For the same reason, in teleological interpretations, judgments about the rationality of means always translate into judgments about 'efficacy': to do something rational that is nevertheless not efficient in actual reality, one needs to act in a fictional or counterfactual world (see the first example in Box 1).

We assume that young infants below one year of age still lack (or because of performance limitations cannot yet use) the complex metarepresentational structures needed to represent intentional mind states [1,24]. We hypothesize, however, that they already possess a non-mentalistic teleological interpretative stance to explain and predict goal-directed actions. (In fact, it seems clearly possible that other similarly – but possibly more permanently – 'mindblind' creatures, like children with autism or non-human primates also possess a non-mentalistic teleological stance.) This non-mentalistic action interpretational system can then explain the empirical findings indicating a precocious understanding of goal-directed actions by the end of the first year.

In fact, the non-mentalistic teleological stance might turn out to be independent both in its functioning and possibly even in its evolutionary origins from theory-of-mind [14]. However, the hypothesized presence and identical role that the rationality principle plays in both the teleological and the mentalistic stance may represent an important structural linkage that could suggest how the former is developmentally related to the latter. When the ability to represent intentional mental states (including pretence and false beliefs) becomes available in the young child, the domain of applicability of the rationality principle (that was restricted to actual and future reality states in the teleological stance) could become enriched by also including fictional and counterfactual world states, as they have now become representable. This would be an example of theory change where the core principle of the earlier (teleological) theory would become applied to an enriched ontological domain [27] thus forming a qualitatively different, mentalistic theory of actions.

Productive teleological inferences about goal-directed actions in one-year-olds

An important property of the rationality assumption is its systematic inferential and predictive generativity: given information about the specific contents represented by any two of the three mental states (desire, beliefs and intention) involved in a mentalistic action representation, one can infer what the content represented by the third mental state *ought* to be (see the examples in Box 1). Therefore, to demonstrate convincingly our central thesis that the rationality principle is also the central inferential component of the non-mentalistic teleological stance we should be able to show that one-year-olds can draw each of the three possible types of inference that adults can in their practical reasoning about intentional actions.

To demonstrate this, we habituated infants to computer-animated goal-directed actions in three types of situations [7,10,15] (Fig. 1). The different event displays were designed so that in each case one of the three basic elements necessary for a well-formed teleological action interpretation was made visually inaccessible. To interpret the action as an efficient and justifiable goal-approach, the infant had to use the rationality principle to infer and 'fill in' the relevant missing element.

Figure 1a exemplifies the first type of teleological inference where infants had to infer the particular 'means action' that is congruent with (i.e. can be seen as an efficient goal-approach in relation to) the visually specified goal state and situational constraints. As described in the introduction, the finding that infants looked significantly longer at the incongruent test display (old jumping approach) than at the congruent one (novel straight-line goal-approach) is evidence that they could draw the inference in question.

Figure 1b illustrates the second type of teleological inference where the infants had to infer a goal state to rationalize the incomplete action whose end state was occluded from them, as an efficient 'chasing' action. During habituation a large ball was approaching a moving small ball until the latter passed through a small aperture between two obstacles and left the screen. The large ball, being too big to get through the aperture, had to make a detour around the obstacles before it also disappeared from view. In the two test events the upper part of the screen was opened up revealing one of two different end states: one congruent with the inferred goal state of an efficient 'chasing' action (the small circle stopped, at which point the large circle changed its course so that it 'caught up with' the small circle and contacted it), and one that was incongruent with the inferred goal (when the small circle stopped, the large one, without modifying its direction, passed by it leaving the screen without ever 'catching' the small circle). Twelve-month-olds looked significantly longer at the incongruent than at the congruent test display, suggesting that the incongruent outcome violated

Questions for Future Research

- Is the teleological stance an independent evolutionary adaptation to the wide-ranging presence of rational goal-directed organization of behavior among animal species in our evolutionary environment providing a specialized mechanism to discriminate and predict goal-directed actions of other agents [14]?
- Is teleological reasoning present in organisms that lack, or possess a deficient, mentalistic stance, like non-human primates or people with autism [28]?
- Is teleological understanding of actions related to teleological reasoning about functions of artifacts and biological properties of living things [29]?
- The inferential role of the rationality assumption in action interpretation resembles that of the principle of relevance [30] that governs our interpretation of communicative acts. Is there a deeper connection between these principles and can, in certain circumstances, the relevance principle be also applied without making reference to intentional mental states?
- Current attempts to design robots capable of imitative learning face the difficulty of how to figure out what aspects of observed behavior should be imitated [31]. Would equipping robots with a teleological stance (and the rationality principle) help to solve this problem?

their expectation about the goal state that they had inferred to rationalize the incomplete action as an efficient 'chasing' event.

Finally, Figure 1c provides an example of the third kind of teleological inference to specify the particular situational constraints (occluded from view by a screen) to rationalize the small circle's visible action (jumping approach) as an efficient means to realize the visible goal state (contacting the large circle). In the two test displays the screen was lifted either revealing an obstacle whose presence justified the jumping approach (congruent display) or revealing no such obstacle (incongruent display). Twelve-month-olds again looked significantly longer at the incongruent than at the congruent display, indicating that they inferred the presence of the occluded obstacle to justify the jumping approach as an efficient means to the goal.

Conclusions

Overall, these results indicate that by the age of 12 months, infants can take the teleological stance to interpret actions as means to goals, can evaluate the relative efficiency of means by applying the principle of rational action, and can generate systematic inferences to identify relevant aspects of the situation to justify the action as an efficient means even when these aspects are not directly visible to them.

Recent studies also indicate that the evaluation of rationality of actions is not restricted to computer animations in infancy. Woodward and Sommerville [12] have shown that one-year-olds expect a hand to perform the most direct means action available to grab a target object, and that the target object is attributed as a goal of the action only if the hand acted efficiently to obtain the object. Furthermore, evaluating the rationality of actions is not restricted to passive observational contexts either. Gergely *et al.* [18] demonstrated that infants modulate their imitative behavior according to the justifiability of the goal-directed actions performed by a model. Briefly, they demonstrated that if the action could be rationalized by the constraints of the model's situation, but the infant's own situation had different constraints, 14-month-olds did not imitate the observed means act: rather, they tried to achieve the same goal by the most rational action available within their own situational constraints.

In summary, we have argued that one-year-old infants possess a naïve theory of rational action that allows them to interpret and predict other agents' goal-directed actions in a variety of different contexts. We have summarized converging evidence demonstrating that when taking the teleological stance one-year-olds apply the same inferential principle of rational action that drives everyday mentalistic reasoning about intentional actions in adults, even though they may not yet be able to represent and attribute intentional mental states to other minds.

References

1 Leslie, A.M. (1987) Pretence and representation in infancy: the origins of 'theory of mind'. *Psychol. Rev.* 94, 84–106
2 Fodor, J.A. (1992) A theory of the child's theory of mind. *Cognition* 44, 283–296
3 Dennett, D.C. (1987) *The Intentional Stance*, MIT Press
4 Perner, J. (1991) *Understanding the Representational Mind*, MIT Press, Cambridge
5 Bartsch, K. and Wellman, H.M. (1995) *Children Talk About the Mind*, Oxford University Press
6 Wellman, H.M. (2002) Understanding the psychological world: developing a theory of mind. In *Blackwell's Handbook of Childhood Cognitive Development* (Goswami, U., ed.), pp. 167–187, Blackwell
7 Gergely, G. *et al.* (1995) Taking the intentional stance at 12 months of age. *Cognition* 56, 165–193
8 Carpenter, M. *et al.* (1998) Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63, 176
9 Woodward, A.L. (1998) Infants selectively encode the goal object of an actor's reach. *Cognition* 69, 1–34
10 Csibra, G. *et al.* (1999) Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition* 72, 237–267
11 Tomasello, M. (1999) *The Cultural Origins of Human Cognition*, Harvard University Press
12 Woodward, A.L. and Sommerville, J.A. (2000) Twelve-month-old infants interpret action in context. *Psychol. Sci.* 11, 73–77
13 Csibra, G. *et al.* (2003) One-year-old infants use teleological representations of actions productively. *Cogn. Sci.* 27, 111–133
14 Gergely, G. (2002) The development of understanding self and agency. In *Blackwell's Handbook of Childhood Cognitive Development* (Goswami, U., ed.), pp. 26–46, Blackwell
15 Csibra, G. (2003) Teleological and referential understanding of action in infancy. *Philos. Trans. R Soc. B Biol. Sci.* 29, 447–458
16 Gergely, G. and 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
17 Csibra, G. and Gergely, G. (1998) The teleological origins of mentalistic action explanations: a developmental hypothesis. *Dev. Sci.* 1, 255–259
18 Gergely, G. *et al.* (2002) Rational imitation in preverbal infants. *Nature* 415, 755
19 Meltzoff, A.N. (1988) Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli. *Dev. Psychol.* 24, 470–476
20 Meltzoff, A.N. (1995) Understanding the intentions of others:

re-enactment of intended acts by 18-month-old children. *Dev. Psychol.* 31, 838–850
21 Kelemen, D. (1999) Function, goals, and intention: children's teleological reasoning about objects. *Trends Cogn. Sci.* 3, 461–468
22 Premack, D. (1990) The infant's theory of self-propelled objects. *Cognition* 36, 1–16
23 Baron-Cohen, S. (1994) How to build a baby that can read minds: cognitive mechanisms in mindreading. *Curr. Psychol. Cogn.* 13, 1–40
24 Leslie, A.M. (1994) ToMM, ToBy, and agency: core architecture and domain specificity. In *Mapping the Mind: Domain Specificity in Cognition and Culture* (Hirschfeld, L. and Gelman, S., eds.), pp. 119–148, Cambridge University Press
25 Premack, D. and Premack, A.J. (1995b) Intention as psychological cause. In *Causal Cognition: A Multidisciplinary Debate* (Sperber, D. *et al.*, eds.), pp. 185–199, Clarendon Press
26 Meltzoff, A.N. (2002) Imitation as a mechanism of social cognition: origins of empathy, theory of mind, and representation of action. In *Blackwell's Handbook of Childhood Cognitive Development* (Goswami, U., ed.), pp. 6–25, Blackwell
27 Carey, S. and Spelke, E. (1994) Domain-specific knowledge and conceptual change. In *Mapping the Mind: Domain Specificity in Cognition and Culture* (Hirschfeld, L. and Gelman, S., eds.), pp. 169–200, Cambridge University Press
28 Abell, F. *et al.* (2000) Do triangles play trick? Attribution of mental states to animated shapes in normal and abnormal development. *Cogn. Dev.* 15, 1–16
29 Keil, F.C. (1994) The birth and nurturance of concepts by domains: the origins of concepts of living things. In *Mapping the Mind: Domain Specificity in Cognition and Culture* (Hirschfeld, L. and Gelman, S., eds.), pp. 234–254, Cambridge University Press
30 Sperber, D. and Wilson, D. (1986) *Relevance Communication & Cognition*, Blackwell
31 Breazeal, C. and Scassellati, B. (2002) Robots that imitate humans. *Trends Cogn. Sci.* 6, 481–487