Developmental Science 1:2 pp 255–259

REPORT

The teleological origins of mentalistic action explanations: A developmental hypothesis

Gergely Csibra[1] and György Gergely[2]

1. *MRC Cognitive Development Unit, London, UK*
2. *University College London, UK*

Abstract

In this paper we shall argue that mentalistic action explanations, which form an essential component of a mature theory of mind, are conceptually and developmentally derived from an earlier and purely teleological interpretational system present in infancy. First we summarize our evidence demonstrating teleological action explanations in one-year-olds. Then we shall briefly contrast the structure of teleological vs. causal mentalistic action explanations and outline four logical possibilities concerning the nature of the developmental relationship between them. We shall argue for the view that causal mentalistic action explanations are constructed as useful theoretical extensions of the earlier, purely teleological, nonmentalistic interpretational stance.

Q: *Why did the chicken cross the road?*
A1: *To get to the other side.*
A2: *It wanted to be on the other side.*

1 The infant's teleological stance

In previous work (Gergely, Nádasdy, Csibra, & Bíró, 1995; Gergely & Csibra, 1996, 1997; Csibra, Gergely, Bíró, & Koós, submitted) we provided evidence that 9-month-olds can interpret the behaviour of an abstract computer-animated object as being goal-directed and can infer its novel action in a changed situation. Infants were habituated to an event in which a small circle repeatedly approached a large circle by 'jumping over' a rectangle. During the test phase, when the rectangle is removed, infants look longer if the small circle repeats its familiar jumping approach than when it takes a novel but shorter (straight line) approach route.

We argued (Gergely & Csibra, 1996; 1997) that to interpret such an event as a goal-directed action infants must establish a specific explanatory relation among three elements: the action, the goal state, and the constraints of physical reality. Such a representational structure constitutes a well-formed teleological interpretation, however, only if it satisfies the principle of rational action which states that an action can be explained by a goal state if, and only if, it is seen as the most justifiable action towards that goal state that is available within the constraints of reality. Thus, the behaviour of the jumping circle can be explained by reference to its final state (contacting the large circle) if the rectangle is interpreted as an impenetrable obstacle and so the jumping behaviour can be considered as a rational action leading through the shortest available path to the large circle.

The principle of rational action also generates an action prediction for the new situation in which the obstacle is removed: the small circle ought to approach its goal through the shortest straight-line path that has now become available. In fact, 9- and 12-month-olds seem to generate such an expectation as they look longer (experiencing incongruence) when the circle's behaviour remains unchanged after the removal of the obstacle.

Adults tend to describe the jumping event in mentalistic terms such as 'it *wants* to go to the other circle and *thinks* the obstacle is impenetrable' (cf. Heider & Simmer, 1944). Note, however, that such mentalistic extensions are not necessary for a viable teleological

Address for correspondence: Gergely Csibra, Department of Psychology, Birkbeck College, Malet Street, London WC1E 7HX, UK.; e-mail: g.csibra@cdu.ucl.ac.uk.

interpretation. The interpretation works even if it makes reference only to the relevant states of current reality (the presence or absence of obstacle) and future reality (the goal state) as represented by the infant herself. Thus, even without attributing these representational elements to the actor's mind as causal intentional states (beliefs and desires) present prior to the action, infants could construct a viable teleological action interpretation or prediction. We hypothesize, therefore, that in its initial form the infant's 'teleological stance' generates reality-based explanations for actions that are neither mentalistic nor causal.

2 Teleological versus causal action explanations

Teleological explanations differ from causal ones in at least two important respects. First, the explanatory element referred to is in a different *temporal* relation to the to-be-explained action: teleological interpretations make reference to the outcome that follows the action, while causal explanations point at some necessary condition that is prior to the event. Second, they use different *criteria* of acceptance: causal explanations single out a prior condition that *necessitates* the action providing its generative source, while reference to a future state is accepted as a teleological explanation (reason) for a behaviour in case it *justifies* it, i.e., when, given the constraints of reality, the behaviour can be seen as a rational way to bring about the goal state.

As a motto to this article, we cited two paradigmatic answers to the old query about the chicken's behaviour. Although they sound similar or even interchangeable, the answers represent two different kinds of explanation for the same action. Let us consider them from the point of view of the two differentiating features of causal vs. teleological explanations. In terms of temporal relations, *A1* is a classic example of a teleological explanation because the outcome of the action ('being on the other side') is cited to account for the event. *A1* also counts as teleological in so far as the future state referred to ('being on the other side') can, indeed, be seen as justifying the chicken's road-crossing behaviour. The situation is not so straightforward, however, in the case of the mentalistic explanation *A2*. In terms of temporal relations, it is like a causal explanation since a prior state of affairs ('wanting to be on the other side') is brought up to explain the action. However, when evaluating *A2* as an explanation, we do not concern ourselves with whether 'wanting to be on the other side' is a necessary precondition that generates road-crossing behaviour, rather, we appeal to the teleological justificatory criterion to see whether 'wanting to be on the other side' justifies the action in the circumstances given.

This 'double nature' of mentalistic explanations is related to the philosophical question whether reasons can be considered as causes in action explanations (Davidson, 1980; Tanney, 1995). Note that causal mentalistic explanations of actions that make reference to desires are always 'teleologically contaminated' in the above fashion. In fact, the generation of desire-based mentalistic action explanations can be seen as a two-stage process involving a *teleological inferential component* (Stage 1) and a *mental attributional component* (Stage 2). During Stage 1, a teleological explanatory relation is established among three elements: a) a behaviour as means action, b) a future state of reality as goal in relation to which the behaviour is rational, and c) relevant aspects of reality which form background conditions for judging the behaviour as a justifiable means to bring about the end state. Stage 2 involves the further operation of attributing the representation of the (teleologically inferred) future state to the mind of the agent as desire, and, similarly, of attributing the representation of the relevant states of reality as beliefs. In short, even though mentalistic belief-desire explanations of actions are formulated as causal relations between intentional mental states and actions, their inferential structure always involves a teleological (reason-giving) element.

3 The relation between the infant's teleological stance and mentalistic action explanations

What is the nature of the relation between infants' early ability to interpret behaviour as goal-directed action and the more mature mentalistic action explanations of young children? We see four logical possibilities: (i) they are unrelated, (ii) they are manifestations of the same capacity, (iii) teleological explanations are derived from a causal theory of mind, or (iv) causal mentalistic explanations are extensions of earlier, purely teleological interpretations. We consider these options in turn.

3.1. *Teleological and mentalistic action explanations are unrelated.*

One could argue as follows: Similarities between interpretations for diverse phenomena do not necessarily imply some inherent link between the representational and interpretational systems involved. For example, the fact that we apply causal explanations to both physical and mental events (say, to colliding balls vs. people's thoughts) does not imply that our understanding of these two domains must be conceptually or developmentally related. Similarly, the fact that both the interpretation of computer-animated events and of people's intentional actions have a teleological aspect does not necessarily suggest any deeper link between them. This account suggests that teleological explanations such as *A1* or infants' interpretations of moving objects, have nothing in common with mentalistic explanations such as *A2*.

We find this view implausible, however, on a number of grounds. First, note that in generating causal intentional explanations the mental states referred to are often attributed via teleological reasoning (e.g., attributing desire from observing or inferring outcome). Second, *A1* and *A2* seem semantically related: If we accept one as an answer, we are likely to accept the other as well. Similarly, adults readily use mentalistic counterparts of infants' teleological interpretations of computer-animated events. These suggestive phenomena receive no principled explanation on the above view, however, which must consider them as accidental.

3.2. Teleological explanations are abbreviated forms of mentalistic explanations.

This account assumes that teleological action explanations are simplified versions of mental explanations in that they refer directly to the content of a causal mental state without explicitly mentioning the intentional attitude itself. Thus, *A1* would be a linguistic shorthand for *A2* implying something like 'Driven by a desire with the content "being on the other side"'. In this view, infants in our studies do not engage in 'purely' teleological reasoning but impute mentally represented goal states (desires) to the agent. Clearly, our evidence is not incompatible with this hypothesis as we have no way of demonstrating that no mental attribution has occurred.

A weaker form of this hypothesis holds that infants' nonmentalistic 'purely' teleological reasoning is but a performance heuristic (see Fodor, 1992) that, nevertheless, already presupposes a theory of mind. The heuristic would be based on the statistical realization that since beliefs tend to correspond to reality, actions can in general be usefully predicted from the interpreter's own representations of the relevant aspects of (current and future) reality, without having to infer and attribute corresponding causal mind states to the actor. Thus, the teleological stance would be seen as a simpler but generally useful interpretational strategy that can economize on computational resources.

While we are certainly not unsympathetic to this position, it should be made clear that it represents an empirically distinct alternative to the view that we shall cautiously advance: namely, that while the nonmentalistic teleological reasoning system will eventually form a proper (inferential) subcomponent of the child's theory of mind (see Stage 1 above), it also functions as an independent interpretational stance which is developmentally prior to mentalistic action explanations. Note in particular that the representational requirements for a 'purely' teleological interpretational system are less severe than the ones presupposed by a theory of mind as the former does not require representing propositional attitude relations(Leslie, 1987; Fodor, 1992) or understanding the representational nature of intentional mind states (Perner, 1991). The fact, that teleological explanations need to make reference only to actual and future states of reality as represented by the infant herself, may help us explain the remarkably early appearance of such interpretations during the first year. We know of no evidence at this early age, however, that would indicate the functioning of other (mentalistic) aspects of theory of mind involving *representational* understanding of intentional mind states.

One can derive differential empirical predictions from the two views above. If the teleological stance doesn't exist independently of theory of mind, then they must be present or absent concurrently in any organism. Children with autism show a specific deficit in tests requiring mental attribution (Baron-Cohen, Leslie, & Frith, 1985; Leslie & Thaiss, 1992). Therefore, if teleological interpretations are but abbreviated or heuristic versions of theory of mind explanations, autistic children should be impaired in tests that require teleological but not necessarily mental reasoning, like the ones we used with infants (Gergely *et al.*, 1995). These studies have not been conducted yet. However, we know of no evidence showing individuals with autism to have difficulties in understanding goal-directed actions, and so we expect them to pass such a test. Such an outcome would undermine the view that teleological interpretations are parasitic on mentalistic explanations, while providing support for our alternative hypothesis.

3.3. Teleological explanations are derived from causal mentalistic ones

According to the third alternative, teleological explanations are derived from a conceptually and ontogenetically earlier causal mentalistic stance. If the infant interprets other people's actions as driven by causal intentions, she may notice that their actions usually end in accord with the content of the preceding intentions. This may encourage the child to explain actions by referring to their end state even in situations when she has no access to the agent's prior causal intentions. A similar notion of derived teleology is advanced by

Kelemen (in press) in the domain of understanding artifacts and biological objects. She explains the child's tendency to interpret man-made objects in functional (teleological) terms as a result of understanding the causal intentions that produced those objects.

Several problems emerge, however, when we apply this hypothesis to the domain of action explanations. First, in this view early mentalistic understanding of intentionality should be demonstrable well before teleological interpretations. Kelemen's (in press) evidence for a teleological construal of artifacts and living things comes from preschoolers and so she can point to the possibly earlier stage of mentalistic action interpretations of two to three year olds (Bartsch & Wellman, 1989; Clements & Perner, 1994) as the derivational basis of the preschooler's teleology. This strategy does not work, however, when applying the model to the domain of action interpretation. Our evidence (Gergely *et al.*, 1995; Csibra *et al.*, submitted) demonstrates teleological interpretation already in 9-month-olds, significantly earlier than the first indications of causal mentalistic interpretations of actions are reported. This time-table is more compatible with our suggestion that purely teleological interpretations precede and provide a developmental basis for causal mentalistic action explanations rather than *vice versa*.

Secondly, as noted earlier, teleological reasoning often plays a crucial role in identifying the content of a causal intention. This should be even more so in the preverbal stage where direct verbal communication of intentional content is ruled out. Behavioural cues such as gaze or movement direction (Baron-Cohen, 1994; Premack & Premack, 1995) are in themselves not sufficient to specify intentional content (see Gergely & Csibra, 1994; 1997) and so, while they may provide useful supporting information, they cannot substitute for teleological reasoning.

Thirdly, the above view, which derives teleological interpretations from a mentalistic understanding of human intentions, should be able to explain how infants generalize their teleological stance to the behaviour of computer-animated spots. One can, of course, refer to the fact that the goal-approaching object in our 'jumping circle' study (Gergely *et al.*, 1995) exhibited movement cues (such as self-propulsion) that may indicate animacy (Mandler, 1992) or even intentional agency (Premack, 1990; Baron-Cohen, 1994). In an other study (Gergely & Csibra, 1996; Csibra *et al.*, submitted), however, we demonstrated similar teleological action interpretation even when all agency or animacy cues were removed. This finding suggests that the infant's teleological interpretation of action is is not mediated by the perception of movement cues of animacy or human agency.

3.4. Causal mentalistic action explanations are extensions of teleological interpretations

We propose that causal mentalistic action explanations are theoretical extensions of teleological explanations. This involves (a) the inclusion of fictional[1] as well as real states of affairs in the explanation, which are (b) 'localized' within the acting agent, and which (c) exist prior to her actions. The fictionalized and mentalized goal-states become desires, while the fictionalized and mentalized reality constraints become beliefs.

One can elucidate the nature of this extension of the teleological stance by drawing an analogy with the extension of real numbers to complex numbers in mathematics. In that case imaginary numbers are introduced to make a useful operation (i.e., square root) applicable to every real number. Similarly, we suggest that the introduction of fictional worlds (whose representations are attributed as mental states) is a theoretical invention which enables the young child to extend the scope of her successful teleological interpretational system. Another implication of this analogy is that just like the introduction of complex numbers does't mean that one always has to treat real numbers as complex-numbers whose imaginary part is zero, we do not have to continuously infer mental states that represent reality (i.e., true beliefs) when explaining action. In other words, in most cases when beliefs correspond to reality, the nonmentalistic teleological stance continues to be sufficient for interpreting action even after the mentalistic stance, which includes fictional states in its ontology, has become available. Thus, theory of mind development would consist not only in mentalizing the teleological stance, but also of learning the conditions under which it is a good strategy to switch to the (computationally more costly) mentalistic action construal.

We can identify at least two good reasons that justifies the transformation of teleological interpretations into causal mentalistic ones. First, when someone acts on false beliefs or engages in pretence, the reality-based teleological stance breaks down as it cannot rationalize the behaviour as a justifiable action. By moving to the level of mentalistic explanations, however, the child can hang on to the rationality assumption through projecting a fictional world in relation to which the observed behaviour can be justified. As a result, the predictive and explanatory scope of the interpretational system becomes significantly enlarged. Second, the teleological stance is more restricted than theory of mind

[1] The term 'fictional' is used here to refer to representations of possible states of affairs which do not map onto the infant's representations of reality.

in the adaptive means it provides for modifying the other's anticipated actions. Since no mechanism is available to influence future goal states, it can modify action only through physical intervention by changing the states of reality that constrain the action (e.g., by obstruction). The momentous gain provided by mentalizing the teleological stance is that one becomes able to influence the other's actions by changing the mental causes that generate it: i.e., by modifying the other's inferred beliefs and desires through communicative interventions such as informing, promising, or pleading.

In sum: our two-stage model for theory of mind development in the domain of action explanations proposes that (a) the initial theory is not causal (source-based) but teleological (outcome-based), (b) the earlier stage does not involve mental state attributions, and (c) the theory of mind stage retains the very same core principle of rational action that governs the reasoning in the earlier, teleological stage. The theoretical transformation proposed here can be seen as an instance of the more general idea that conceptual change in a domain may involve the establishment of new ontological types (in our case, fictional mental states over and above reality states) while leaving the core inferential principle of the domain unchanged (cf. Carey & Spelke, 1994). Note, furthermore, that our developmental hypothesis also accounts for the peculiar 'double nature' or characteristic 'teleological contamination' of causal mentalistic action explanations described earlier.

Acknowledgements

We thank Paul Bloom, Mark Johnson, Jean Mandler, Csaba Pléh, and John Watson for their valuable comments on an earlier version of this paper.

References

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**, 1–40.

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

Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development*, **60**, 946–964.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the Mind. Domain Specificity in Cognition and Culture* (pp. 169–200). New York: Cambridge University Press.

Clements, W.A., & Perner, J. (1994). Implicit understanding of beliefs. *Cognitive Development*, **9**, 377–395.

Csibra, G., Gergely, G., Bíró, S., & Koós, O. (submitted). Goal attribution without agency cues: The perception of pure reason in infancy.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.

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

Gergely, G., & Csibra, G. (1994). On the ascription of intentional content. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, **13**, 584–589.

Gergely, G., & Csibra, G. (1996). Understanding rational action in infancy: Teleological interpretations without mental attribution. Paper presented at the 10th Biennal Conference on Infant Studies, April, 1996, Providence.

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

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

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

Kelemen, D. (in press). The origins of teleological thought. In M. Corballis & S. Lea (Eds.), *Evolution of the Hominid Mind*. Oxford: Oxford University Press.

Leslie, A.M. (1987). Pretence and representation in infancy: The origins of 'theory of mind'. *Psychological Review*, **94**, 84–106.

Leslie, A.M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, **43**, 225–251.

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

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

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

Premack, D., & Premack, A.J. (1995). Intention as psychological cause. In: D. Sperber, D. Premack, & A.J. Premack (Eds.) *Causal Cognition: A Multidisciplinary Debate* (pp. 185–199). Oxford: Clarendon Press.

Tanney, J. (1995). Why reasons may not be causes. *Mind and Language*, **10**, 105–128.

Received: 18 September 1997
Accepted: 8 December 1997