Cognition, 27 (1987) 285–290

Discussion

On not having a theory of mind*

BEATRICE DE GELDER
Tilburg University

In a recent paper Baron-Cohen et al. (1985) claim to have demonstrated that autistic children "suffer from a specific deficit that is largely independent of general intellectual level" e.g., "not having a theory of mind", "failing to employ a theory of mind" or "being unable to appreciate the difference between their own and the doll's knowledge". The crucial question is whether this conclusion makes sense beyond the bounds of the experimental paradigm. If so, this study confirms the applicability of a computational theory of cognition to the study of development and of developmental deficiencies. It would also show how to build a bridge between theories of intra-individual cognition and the inter-individual processes of belief attribution. The credit for these successes would go to the internal common language basis of all the abilities involved.

The following research paradigm was adopted. The child, sitting across the table from the experimenter is watching two puppets. One, Sally, puts a marble in a basket and leaves. The other, Ann, takes the marble out and puts it in a box. Sally then returns to the scene and the child is asked "Where will Sally look for the marble?". The majority of autistic children answer by telling where the marble is now. On that basis the authors conclude that autistic children do not have a theory of mind because they are not able to attribute beliefs to other people.

Is this conclusion warranted? Our discussion will focus on three specific aspects: (a) the play situation; (b) the linguistic format of the interaction situation and (c) the focus on false belief attribution. Our conclusion suggests a *deconstruction* of the ability to have a theory of mind into heterogeneous abilities as opposed to the '*homunculisation*' of the ability by Baron-Cohen et al.

The expression "having a theory of mind" is borrowed from Premack and Woodruff (1978). Their research illustrates a general perspective on the study

*Reprint requests should be addressed to B. de Gelder, Tilburg University, Postbus 90153, 5000 LE Tilburg, The Netherlands.

0010-0277/87/$2.30

of mentality in complex organisms which Dennett defends as the "intentional stance" (Dennett, 1971, 1981). The hallmark of the intentional strategy is its agnosticism in matters of mind. Concepts like intentionality, belief, desire or purpose are exclusively treated as instruments of behavioral prediction. For example, it does not consider beliefs as intrinsic mental states of organisms. It follows that the intentional strategy is just as well worth adopting in the case of animals or artifacts as in that of humans. It is neither recommended nor rejected by a particular epistemological or metaphysical perspective on mental states nor by theories of intrinsic cognitive abilities. This agnosticism is a major asset in circumstances were one is likely to be mislead by folk typologies.

Baron-Cohen et al. combine the intentional stance with a fodorean perspective on propositional attitudes, taking the objects of propositional attitudes to be mentalese representations or expressions in the language of thought (Fodor, 1975, 1981). Fodor's thesis is very complex, ambitious and still in its programmatic stages. A rough characterization will illustrate the use these authors make of it.[1] Crudely speaking, a mentalese representation is an entry in the mentalese lexicon containing all the information the subject needs for dealing with the real world represented by the mental representation. For Baron-Cohen et al. (and Leslie, in preparation; Morton 1986) these mentalese representations underlie straightforward, realistic or 'first degree' behaviour, i.e., behaviour where things are simply taken for what they are. But they go further than Fodor and postulate a metacognitive operator whose function it is to 'mentalize' (Morton, 1986) the mentalese representations, to generate second order representations, supposedly explaining secondary behaviour like pretending. For example, a two-year-old child is able to treat a telephone as a telephone because he has the entry underlying that behaviour in his mental lexicon. A slightly older child can also treat a banana as if it was a telephone. The representation underlying this pretend play would be the product of the metacognitive operator maintaining temporarily the primary representation in a state of suspended reference.

It is well-known, as the authors note, that autistic childern cannot pretend play. The theoretical apparatus just sketched is designed to explain that inability. But it is equally well-known that autism manifests itself as an inability to interact with other people, to attribute them intentions, beliefs, desires etc. The theory tested in Baron-Cohen et al. is that the social interaction deficiency has exactly the same cause as the inability to pretend play. Both would be consequences of the fact that autistic children do not have the metacognitive operator. For that reason they would be unable to mentalize and would be restricted to realistic interactions with objects as well as with people, based on representing objects with true properties and people with true beliefs. In other words, if inability to pretend play and inability to attribute beliefs have a common cause then we should observe that autistic children turn out to be unable to attribute beliefs to other people. This is the actual hypothesis of the experiment. On the strength of their general theoretical framework Baron-Cohen et al. assume they can study the ability to mentalize in relation to people in the experimental situation described above. This combination of an intentionalist methodology with a computationalist theory of mental states seems to allow a remarkable streamlining of the methods for studying mentation. Combining the theoretical assumption about the language of thought with the evident fact that autistic children have natural language reduces painstaking observation of autistic children to asking questions. Will this work?

(a) The conditions for successful study of belief attribution may not be present here because belief attribution is embedded in a play situation which may mask the belief attribution task. In general, normal children enter willingly a play which involves pretending that puppets (or animals) are people, have beliefs, intentions, desires etc., can talk and act. But this is pretend play, a complex achievement that begs an explanation. Indeed, we know that autistic children cannot pretend play and we can therefore expect that they will fail a belief attribution task when embedded in a pretence situation. We are, of course, not arguing that a play situation is inappropriate per se, only that the task should be validated for the study of autistic children. Before trusting the puppet paradigm one needs to study natural interactions in situations that the experiment intends to reproduce. For example, it is crucial to know whether autistic children in daily interaction with people manifest the types of behaviours that beg intentional language, like, for example hiding objects from others, not acknowledging in their actions the possibility that others may or may not have the same belief about a situation etc.

(b) The second problem is that in this experiment the belief attribution situation is embedded in language and created through verbal requests and questions. The difficulty here is that autistic children have impaired discourse abilities. Most typically, autistic children have problems with topic maintenance and with the illocutionary force of utterances, for example with understanding questions. Their verbal interactions often reduce to requests for information (e.g., Johnson, 1985). Hence, one can doubt whether the autistic children do in fact understand the experimental question as a question about what Sally knows, as opposed to a question on where the marble is. It is then difficult to tell what kind of failure the experimenter is witnessing.

[1] It is hard to tell whether the use Baron-Cohen et al. make of the mentalese thesis is appropriate. We

(c) If we grant that the autistic children answer the question but answer it wrongly, the conclusion that they do not have a theory of mind is nevertheless still difficult to support. Indeed, when treating the responses autistic children give as answers to the question, whether good or bad, one grants them the ability to understand questions. To assume that the autistic children have understood the question means that one is presupposing that the autistic children do have a theory of mind which enables them to attribute beliefs, intentions etc. to the experimenter asking a question: without a theory of mind one cannot participate in a conversation. It might thus be argued that the results indirectly do show that autistic children have a theory of mind. Only, their theory of mind is less complex than that of normal children. It might for instance be limited to the sort of same belief attribution which Premack and Premack (1983) have reported for chimpanzees. This positive conclusion does not square easily with the theory behind the Baron-Cohen et al. experiment. Unlike for example, Premack and Woodruff, they do not study interactions consisting of simple belief attribution. Their theory seems to reduce the explanation of attribution of a true belief to the ability to have a realistic representation of a person. In conflating attribution and representation they exclude one likely explanation of inability of false belief attribution, e.g., the need for cooperation between belief attribution ability and conceptual representation capacities. Thus they exclude that having a theory of mind admits of degrees and that it is an ability that as Premack (1986) notes, can be educated.

The foregoing remarks suggest a deconstruction of the notion of having a theory of mind into heterogeneous abilities and project a picture of having a theory of mind which is antithetic to the homogeneous ability for which Baron-Cohen postulate a single metacognitive operator. The research on autism (and more generally, on development in relation to deficiency of abilities) might illustrate the limits of an integrative theory and accompanying methodology. The crucial question is whether it makes sense to reduce all the known autistic deficiencies to dysfunction of the same underlying computational mechanism. If not, the suggestion reduces to an 'homunculisation' of the ability instead of a functional analysis.

Our comments stress three different components of the ability to have a theory of mind: the biological function of interaction, the linguistic and conversational skills and the conceptual ability of having a theory of mind. The biological perspective is particularly relevant since autism appears very early on in development (Rutter, 1983), at the time communication gets organized, before the appearance of speech (e.g., Bates et al., 1979), long before the mastery of verbs of propositional attitude (e.g., Wellman, 1985) and long before the child manifests explicit knowledge of his own and others' mental states. This suggests the existence of a biological proper function (in the sense of e.g., Millikan, 1984) of social interaction. Autism might reflect a partial breakdown of that biological function and it might follow from the influence of that breakdown on the development of conceptual and linguistic abilities. Thus deficiencies and breakdowns of normal functioning confront us with the breakdown of concepts (like theory of mind) that are designed for understanding the conceptual coherence of an ability, but not for the empirical study of its functional basis in organisms. The ability to attribute beliefs observed in normal adults is likely to be the result of a progressive integration of the biological, conceptual and linguistic components and of their reciprocal interactions (de Gelder, 1987). Each component has its own structure and its own conditions of satisfaction. At the same time, the componential view guarantees that breakdowns and developmental deficiencies are often only partial. If autistic children would not have a theory of mind in the absolute sense suggested by Baron-Cohen et al. there would be no point to this kind of research.

References

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

Bates, E., Benigi, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Communication and cognition in infancy*. New York: Academic Press.

Dennett, D. (1971). Intentional systems. *Journal of philosophy, 68*, 87–106.

Dennett, D. (1981). Three kinds of intentional psychology. In R. Healey (Ed.), *Reduction, time and reality*. Cambridge: Cambridge University Press.

Fodor, J.A. (1975). *The language of thought*. New York: Cromwell.

Fodor, J.A. (1981). Propositional attitudes. In *Representations*. Cambridge, MA: MIT Press/Bradford Books.

de Gelder, B. (1987). Beyond Suspicion. Cognitive and intentional basis of the ability to lie. *Argumentation, 3*.

Johnson, J.R. (1985). The discourse symptoms of developmental disorders. In T.A. van Dijk, *Handbook of discourse analysis (vol. 3)*. London: Academic Press.

Leslie, A. (in preparation). Pretense and representation in infancy.

Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge, MA: MIT Press/Bradford Books.

Morton, J. (1986). Developmental contingency modelling. A framework for discussing the processes of change and the consequence of deficiency. In P. van Geert (Ed.), *Theory building in developmental psychology*. Amsterdam: North-Holland.

Premack, D., & Premack, A.J. (1983). *The mind of an ape*. New York: Norton.

Premack, D. (1985). 'Gavagai' or the future history of the animal language controversy. *Cognition, 19*, 207–296.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515–526.

Rutter, (1983). Cognitive deficits in the pathogenesis of autism. *Journal of Child Psychology and Psychiatry, 24*, 513–531.

Wellman, H.M. (1985). The child's theory of mind: The development of conceptions of cognition. In S.R. Yussen (Ed.), *The growth of reflection in children*. San Diego: Academic Press.

Woodruff, G., & Premack, D. (1977). Intentional communication in the chimpanzee: The development of deception. *Cognition*, *7*, 333–362.