:PC, 1994, 13 (5), 607-614

:OMMENTARY ON "COGNITIVE MECHANISMS N MINDREADING" (S. BARON-COHEN)

Eye to eye, but not a meeting of minds

CECILIA M. HEYES[1] AND TIM P. GERMAN[2]

University College London and MRC

Two of the most distinctive and impressive features of Baron-Cohen's model of a "mindreading system" are the degree to which it 1) makes explicit use of Fodor's (1983) criteria for assessing modularity, and 2) emphasises the role of eye detection in the development of mind-reading. We will comment on each of these, urging, in the case of modularity, further exploitation of Fodor's approach, and, in the case of eye detection, even higher standards of evidence in the attribution of concepts such as "seeing" to infants. The latter point is motivated by another of the unusual and commendable features of Baron-Cohen's approach: extensive use of research on nonhuman animals.

Modularity

Reference to "modules" is common in contemporary cognitive science, but Baron-Cohen's use of the term is rare in its clarity. He recognizes both that modularity is a matter of degree, and that the questions raised by Fodor's modularity criteria are empirical questions: they set an agenda for research. In this he follows Fodor, who stated

---

1. Department of Psychology, University College London, Gower Street, London WC1E 6BT, U.K. (e-mail: ucjtsch @ ucl.ac.uk).
2. MRC Cognitive Development Unit, 4 Taviton Street, London WC1H OBT, U.K.

plainly that "each question is susceptible to a 'more or less' sort of answer... When I speak of modular, I shall always therefore mean 'to some interesting extent'" (1983, p. 37). While Fodor clearly meant this to apply to all of his nine criteria, Baron-Cohen excludes three from consideration: information encapsulation, shallow outputs and accessibility to consciousness.

The exclusion of informational encapsulation from the analysis of modules is perhaps most surprising, because it was judged by Fodor to be "the 'essence of their modularity'" (ibid., p. 71). Baron-Cohen decides that information encapsulation is not needed on the grounds that it prevents the "possibility that modules interact with one another...". However, if, as Baron-Cohen's overall approach implies, the organization of the mindreading system is an empirical issue, and its modularity is to be judged in Fodorian terms, then surely the question of informational encapsulation should be addressed rather than put to one side. The question as to what interaction the systems responsible for eye-direction detection have with other components of the mindreading system is dealt with in Baron-Cohen's sections 3 and 5. In the second half of this commentary we look at the EDD module and its interactions more closely, but for the moment, let us suppose that some interaction between modules does in fact happen. Several positions would then be possible: One position assumes that because the EDD and SAM modules interact they therefore cannot be highly encapsulated. Another might instead hold on to the encapsulation notion; taking the penetration of SAM by eye direction information to be an instance of penetration by information within the same module, which preserves encapsulation (see Fodor, 1983, p. 80 ff.). A third possibility is that EDD and SAM are encapsulated, and only interact with one another in the sense that one provides inputs to the other, but not vice versa. This uni-directional interaction would not suggest non-encapsulation. The point is that the organization of these processes is going to turn out to be highly complex. None of these questions can be prejudged, and the evidence that might be brought to bear is exactly the type that Fodor (1983) discusses in his section on information encapsulation.

More generally, application of all nine of Fodor's criteria is desirable because it provides maximum opportunity to discover the extent to which modular properties co-occur in various systems, and thereby to find out about their causal and functional relationships. For example, speed may be achieved by virtue of informational encapsulation and at the cost of yielding shallow inputs and/or obligatory firing. More interesting still, systems that have the properties traditionally identified with

innateness – characteristic ontogenetic course, dedicated neural architecture, characteristic pattern of breakdown – may be more likely than systems that fail to meet these "biological" criteria to have the on-line operating characteristics specified in the other six modularity criteria. Surely one of the main reasons why Fodor's views on modularity have received so much attention is because they embody this bold hypothesis about the relationship between the evolutionary origins and operating characteristics of psychological systems, or, as it has been put elsewhere, the view that "core architecture" is the basis for development rather than its outcome (Leslie, 1988, 1994). Embracing modularity is viewed as taking one closer to the "abyss of nativism" (Wellman, 1990), and Baron-Cohen is undoubtedly aware of this issue. He cites both Bates' (1993) argument that some modularity criteria are also "overlearning" criteria, and Karmiloff-Smith's (1992) characterization of the structure of the mind that posits gradual "modularization" during development. It would be regrettable, therefore, if by treating Fodor's criteria as a menu rather than a set, we lost the opportunity to test his central, nativist hypothesis.

Turning to the criteria that Baron-Cohen includes in his analysis, we note that Fodor (1983) offers guidelines on their application that could be used profitably in the investigation of mindreading. Thus, Fodor suggests that the firing of modules may be regarded as obligatory if it cannot be prevented by instruction or preference. For example, you cannot hear speech as noise *even if you would prefer to*; even by attempting not to hear speech in the first place, a great deal of information is processed (e.g., Lackner & Garrett, 1973). Does this type of operation hold for detection of eye-direction or computation of goal? Evidence that you-couldn't-avoid-doing-it-even-if-you-tried is required for this issue. On this point, Baron-Cohen concedes that one can suppress eye-gaze monitoring if one is told to, but more evidence of this type would seem to be required.

"Rapid speed" is another criterion for which Fodor provides useful guidelines. Speed is a characteristic that Baron-Cohen deals with for all his putative modules. The treatment that he gives it, for example in section 1.3, stresses how fast the attribution of goal "seems". While we agree that the "How fast is fast...?" question is not an easy one, it would seem appropriate at least to offer some "relative-to" type evidence. Fodor (1983, p. 62) notes that the problems of quantification are severe, but suggests a metric in terms of hard vs easy problems. Although his "months vs milliseconds" rule might be a little extreme, it at least provides us with an idea of how speed might be assessed, in the

absence of reaction time data. Computing "instantly" that a bee wants to go to the flower may be a result of the simplicity of this computation. Are all computations of goal and desire performed as quickly as this, given some behaviour to operate on, or can one find examples of goal computations that make much longer?

With respect to domain specificity, Baron-Cohen addresses the "well-definedness" of the domains in question, and the employment of unique representations by the modules that are proposed to deal with them, while Fodor deals with domain specificity in terms of the eccentricity of the stimulus domain. The more eccentric the stimulus domain, the more plausible the claim that psychological processes defined over that domain are carried out by special purpose processors. He suggests that the eccentricity of language as a stimulus domain makes it a candidate for such processing. The notion that specialised representational systems might be employed in language comes from Chomsky (1976, 1986) and Pinker (1989).

The representations that Baron-Cohen proposes do not turn out to be straightforward. Firstly, he defines dyadic representations as specifying two entities in a relationship. These dyadic representations are held to operate in both EDD *and* ID. As well as this strain on their uniqueness, it would seem that representations of this general type might be very useful in other domains, to specify for example mechanical relations (e.g., cup-IS SUPPORTED BY-table), or transitive operatives (e.g., green strick-IS BIGGER THAN-red stick). It is not clear whether EDD and ID are the only modules that use dyadic representations, or whether the representations computed by EDD and ID are just a unique class of dyadic representations. If the latter, what is it that makes them unique?

Secondly, Baron-Cohen identifies four forms of dyadic representation. This seems complex in that it assumes not only that infants have a concept of self that is distinct from another person (Baron-Cohen, footnote 2), but also that this self-concept results in representations of a different form rather than a different content. It seems to us that the assumption of a self concept only warrants the inclusion of "self" as a specific content within agent-agent or agent-object representation forms.

Finally, one of the attractions of Baron-Cohen's position is the apparently fluid progression in representational terms from dyadic to triadic representations. Triadic representations are held to be required for shared attention. They are formed by including an embedded element (presumably in a dyadic representation) that specifies that an agent and another agent are attending to the same object. However, it is not clear in the examples that Baron-Cohen gives (section 3.1) that the transition

from dyadic to triadic representations is as natural as it first appears. The triadic representation said to underlie shared attention does seem to involve the embedding of one dyadic representation in another, e.g., [I-SEE-(you-SEE-the-bus)], but the knowledge that two agents represent the same object, that you and I see the *same* bus, would necessitate the combination of this triadic representation with another, dyadic one (e.g., I-SEE-the bus) in an inference. Thus, the simple building that appears to mark the transition between dyadic and triadic representations is not strong enough to represent shared attention.

Eyes and minds

Having considered, in general terms, Baron-Cohen's use of the modularity thesis, our remaining comments relate to two of his putative modules: the eye-direction detector (EDD) and shared attention mechanism (SAM). In both cases, a capacity to represent the relation "sees" is inferred from sensitivity to gaze direction, and in both cases we doubt the validity of the inference. Much of the behaviour that is cited as evidence of sensitivity to gaze direction could in fact reflect no more than sensitivity to the presence vs absence of eye-like stimuli, and even where there is good reason to believe that individuals (infants or non-human animals) are sensitive to gaze direction, it would seem hasty to attribute the concept of "seeing" on this basis.

Consider, first, the phylogenetic evidence for sensitivity to gaze direction reviewed in section 2.2. In Ristau's (1990, 1991) studies of plovers, eye direction and head direction were confounded, and therefore we cannot be confident that gaze direction was the cue. In each of the remaining examples, concerning a range of species from hog-nosed snakes to macaques, animals were found not to respond in different ways to different *directions* of gaze, but to respond less vigorously to averted than to direct gaze (e.g., Burghardt, 1990; Mendelson, Haith, & Goldman-Rakic, 1982). While topographically or directionally distinct responses to different directions of gaze would indicate sensitivity to eye direction, the cited effects are likely to indicate only that the animals concerned have a simple eye detector mechanism which fires maximally in response to the stimulus configuration that we humans would describe as direct gaze (a certain d:w ratio in Baron-Cohen's notation), and at a proportionally lower rate in response to stimuli that differ from this configuration, regardless of the "direction" of the difference. If, as we have suggested, the comparative data fail to

provide evidence that nonhuman animals are sensitive to gaze direction, then those data could not, as Baron-Cohen implies, indicate that a human EDD fires obligatorily.

Turning to the cited evidence of sensitivity to eye direction in infants, and applying the same standards of evidence as for nonhuman animals, we do at least find one experiment in which the phenomenon was demonstrated: Butterworth (1991) showed that 6 month olds distinguish eyes that are looking to the left from those that are looking to the right. Should this be taken to indicate that infants at 6 months can represent the relation "sees"? We think not, for two reasons. First, there is nothing in Butterworth's study to suggest that the infants' represented an object in addition to the agent's eyes, and therefore no reason to suppose that the eyes were represented in any kind of relation. Second, had there been evidence of relational encoding, identification of the represented relation with the concept "sees" would imply, *inter alia*, that infants at 6 months appreciate the relationship between visual access and knowledge, and, as far as we are aware, the youngest infants to provide evidence of such an appreciation were about 3 years old (see Perner, 1991, for a review). In view of this, at minimum it would be preferable to attribute the concept "looking", an observable activity, rather than "seeing", an intentional state, to infants on the basis of existing evidence of their sensitivity to direction of gaze.

A similar gap between theory and evidence opens up when one looks carefully at SAM. In this case, joint visual attention behaviour – looking in the same direction, or at the same object, as another agent – is explained in terms of the formation of triadic representations in which an agent (self or other) sees that a second agent sees the same object as the first. Triadic representations of this sort, not only contain a hidden inference, as noted above, but they are much more complex than is necessary to explain the observed behaviour. An infant that consistently turns in the direction of another agent's gaze need only be assumed to expect that such behaviour will result in them (the infant) seeing something interesting (Moore & Corkum, in press). Thus, it is not clear that the infant need be assumed to represent what the agent sees in any way, let alone as the "same thing" that they see themselves.

Corkum and Moore (in press) have provided evidence in support of a simpler interpretation of joint visual attention; one postulating that infants expect to see something interesting, but do not represent what another individual sees. This interpretation implies that joint visual attention behaviour may be acquired through associative learning, and Corkum and Moore (in press) rendered this plausible by showing that,

while infants do not exhibit spontaneous joint visual attention until 10-12 months, they can learn to align their gaze with that of an adult in an instrumental conditioning procedure at 8-9 months or earlier.

Mendelsohn, M., Haith, M., & Goldman-Rakic, P. (1982). Face scanning and responsiveness to social cues in infant monkeys. *Developmental Psychology, 18,* 222-228.

Moore, C., & Corkum, V. (in press). Social understanding at the end of the first year of life. *Developmental Review.*

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure.* Cambridge, MA: MIT Press.

Ristau, C. (1990): Aspects of the cognitive ethology of an injury feigning plover. In C. Ristau (Ed.), *Cognitive ethology: The minds of other animals.* Hillsdale, NJ: Erlbaum.

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

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