

COMMENTARY ON "COGNITIVE MECHANISMS  
IN MINDREADING" (S. BARON-COHEN)

## The mindreading engine: Evaluating the evidence for modularity

DARE A. BALDWIN and LOUIS J. MOSES

University of Oregon

Simon Baron-Cohen has taken up the challenging task of describing the cognitive architecture underlying our earliest insights into mental life. His central thesis is that our everyday theory of mind develops out of a series of progressively more sophisticated "modules." The mindreading engine, as Baron-Cohen conceives it, consists of four separable, independent, evolutionarily preadapted, neural systems: an intentionality detector (ID), an eye-direction detector (EDD), a shared attention mechanism (SAM), and a mechanism for drawing inferences about others' mental states (TOMM). These four systems are argued to be modular on the basis of their ability to satisfy a set of criteria distilled from Fodor's (1983) original nine hallmarks of modularity. Of the six revised criteria that Baron-Cohen settles upon, three are taken to be especially critical for making a compelling case for "biological" modularity: characteristic ontogenetic course, dedicated neural architecture, and characteristic pattern of breakdown. He notes that the others – domain specificity, obligatory firing, and rapid speed – are also symptomatic of cognitive processes rendered automatic through practice, and as such are consistent with modules constructed over time through experience.

---

Authors' address: Department of Psychology, University of Oregon, Eugene,  
OR 97403-1227, U.S.A. (e-mail: baldwin @ darkwing.uoregon.edu; moses @  
darkwing.uoregon.edu).

Baron-Cohen's agenda is clearly an ambitious one: he attempts to specify the relations among a number of central aspects of early social cognition and, in so doing, he forges new and thought-provoking links among research findings previously scattered through several disparate literatures. In the end, however, we remain unconvinced that Baron-Cohen's modular architecture rests on a solid evidential foundation. We begin this commentary by first considering whether the evidence does indeed satisfy Baron-Cohen's "Big 3" biological criteria. We next give shape to what is perhaps the mainstream alternative to a modularity account: an alternative that centers on the construction and conceptual reorganization of knowledge. Finally, we turn to a more general consideration of the implications of Baron-Cohen's revised criteria for modularity.

### Evaluating the big 3 "biological" criteria

*Characteristic ontogenetic course.* An evolutionarily preadapted module should emerge according to a universal bioprogram in all normal members of the species. The evidence cited by Baron-Cohen in support of his hypothesis that ID, EDD, SAM, and ToMM fit such a developmental pattern is unfortunately far from abundant. With respect to ID, we are told only that linguistic reference to goals and desires typically appears early and that older children have a propensity to interpret even geometric shapes as having goals and desires. Regarding EDD, several studies are cited suggesting that humans are sensitive to eye contact and direction of eye gaze. The argument for characteristic ontogenesis in SAM rests on data indicating that towards the end of the first year infants begin to show a variety of joint attention behaviors in the visual modality (e.g., gaze monitoring and pointing). For ToMM, evidence is cited that pretense emerges towards the end of the second year and that epistemic mental states are first understood by 3 or 4 years of age.

Several problems undercut the value of much of this evidence. First, some recent findings are inconsistent with Baron-Cohen's evidence. For example, while the data cited by Baron-Cohen suggest that infants as young as 6 months of age will follow an adult's gaze toward the right or the left, findings from Corkum and Moore (in press) indicate that this ability does not begin to emerge until the end of the first year. Second, the existing data are, for the most part, not *normative*. Typically, the evidence arises from small-scale studies investigating highly specific

abilities. Such studies can only yield data concerning the statistically detectable *presence* of the abilities in question within a highly circumscribed population of children. Demonstrating characteristic ontogeny, however, requires the identification of a highly reliable and tightly predictable maturational timepoint at which the relevant abilities *emerge* in children from many different backgrounds and cultures. At this point the requisite cross-cultural studies are few and far between and, even within Western cultures, little information is available concerning the variability in the age at which children display basic mindreading abilities. Moreover, where such evidence is available it is not always supportive. With respect to ToMM, for example, the evidence indicates that even some young 3-year-olds pass false belief tasks with ease whereas some 5-year-olds continue to have great difficulty. Variability of this kind poses an immediate difficulty for Baron-Cohen's biologically-based modularity account. To salvage his proposal it would seem critical to show that variability in age of onset is somehow related to biological factors rather than to either experiential factors such as SES or parental input, or general cognitive factors such as IQ or verbal ability. Third, if Baron-Cohen is right, it ought to be the case that all the abilities linked to a given modular system are tightly correlated in terms of when they first emerge. The evidence cited, however, appears to suggest that different components of some modules develop at quite different times. With respect to EDD, for example, although neonates may be capable of detecting the presence of eyes, it is not until at least several months later that infants demonstrate any sensitivity to direction of gaze. Of course, what is to count as a tight correlation in the age of onset of such abilities is likely to be a matter of opinion. Still, one thing is clear: within-module variability should be smaller than between-module variability. That is, we should see tighter coalescence in the development of abilities controlled by a specific module than in the development of skills controlled by different modules. Much of the evidence presented by Baron-Cohen, however, is simply too crude to test this hypothesis because it is derived, not from infant studies, but from research with older children and even adults. In the absence of more fine-grained developmental data, claims about characteristic patterns in ontogeny, as well as their implications concerning modularity, are at best speculative. Finally, it is important to note that predictions about characteristic ontogenetic course are not unique to modular accounts. Certain abilities necessarily emerge in developmental sequence. For example, on any account, joint attention abilities would seem to be necessary for epistemic state attribution.

*Dedicated neural architecture.* Baron-Cohen readily acknowledges that, as yet, there is little or no evidence concerning the neurophysiological underpinnings of several of the proposed modules. Nevertheless, we would even take issue with the little evidence that is offered. For one thing, locating cells (in the monkey brain) that respond specifically to goal oriented actions or to direction of gaze is far from compelling evidence for dedicated neural architecture. At this time, we cannot rule out the possibility that, on further investigation, cells in other areas of the brain will also show specialized firing to such stimuli. Even more importantly, specialization of cells is not sufficient for inferring *a priori* dedication of those cells for the detection of their eliciting stimuli. Those cells may have become specialized through experience and, in the absence of stimulus information of one kind, they might well become specialized for other kinds of stimuli. For all we know, cells in the STS that might be specialized for gaze direction in sighted people could turn out to be specialized for quite different stimuli in blind individuals lacking input about gaze direction. In other words, it may be that certain regions of the brain are well-suited to carrying out computations important for certain tasks and, through general processes like competition, such regions over time come to monopolize the execution of those computations. If a given task, such as eye-direction detection, is for some reason not performed, these areas would simply become dedicated to a different task that also required the same or similar computations. Interestingly, Jacobs, Jordan, and Barto (1991) have demonstrated just such a process of progressively acquired modular dedication in a connectionist simulation. Consequently, we remain skeptical that dedicated neural architecture is a clear criterion for biological modularity. It seems to us that, even if clear-cut evidence for specialization of cells to domain specific stimuli were to be found, a case for a dedicated, *biologically-based*, modular system would still not have been established. Such evidence would be equally consistent with a module constructed through general learning mechanisms.

*Characteristic pattern of breakdown.* Double dissociations – complementary patterns of neural breakdown in two different abilities across at least two different individuals – are recognized as the only compelling clinical evidence for independent neural systems (e.g., Farah, 1992). A single dissociation, in which one patient displays breakdown in one area but remains intact in another, is most parsimoniously explained by reference to just one damaged system that is heavily utilized by the impaired ability and less heavily utilized by the intact skill. Ideally,

evidence for double dissociations should come from individual data rather than the averaged group data that Baron-Cohen presents. That aside, the patterns of breakdown he describes generally do not fit the double dissociation pattern even for grouped data, and therefore they provide little basis for claims about separate modular processes. With regard to ID and EDD, for example, one clinical group is discussed that shows intact intentionality detection but impaired eye-direction detection (AB type prosopagnosics), but no group is described that displays the required complementary pattern – intact eye-direction detection and impaired intentionality detection. The same is true for SAM and ToMM: One group – Autism (B) – is described as exhibiting impairments in ToMM while maintaining capability in shared attention, but no group is reported as showing the opposite pattern. Of course, the hierarchical relation between these two modules would render such a pattern improbable but, unless a double dissociation of this kind can be demonstrated, characteristic breakdown loses its force as a strong criterion of modularity.

Patterns of breakdown that Baron-Cohen links to EDD and SAM come closest to satisfying the double dissociation requirement. Those with Autism (A) display intact eye-direction detection but are impaired in the negotiation of joint attention, while AB type prosopagnosics show the opposite pattern: impairment in EDD without deficits in shared attention.<sup>1</sup> Even in this case, however, the evidence for a double dissociation is incomplete. In particular, two important points are not made clear: a) whether this is the sole deficit that AB type prosopagnosics display, and b) whether they display deficits across the full range of abilities that EDD is thought to control. That is, a convincing case for modularity can only be made if the deficit is highly specific: everything in that module, and that module alone, should be impaired.

In sum, the evidence regarding characteristic patterns of breakdown does not strongly indicate the existence of Baron-Cohen's four independent modular systems. Instead, most of it is equally consistent with breakdowns of increasing severity in just one system or process – a

1. Although in Table 1 blind individuals are listed as displaying a pattern similar to AB type prosopagnosics (i.e., intact SAM but impaired EDD), we do not regard their blindness as evidence for the breakdown of EDD as an independent neural system. In the blind, it is typically aspects of the peripheral visual system that are impaired. In these cases, the neural system which in sighted people subserves EDD may well remain intact.

process that is most heavily taxed by inferences about others' epistemic states, next most heavily taxed by negotiation of shared attention, and so on. Moreover, even if convincing evidence for independent neural systems were available, it would not amount to evidence for *biologically pre-specified* systems. Such evidence would be equally consistent with modules constructed through general learning processes.

#### Knowledge reorganization as an alternative model

If, as we suspect, the development of mindreading skills does not consist in the emergence of a series of ever more "intelligent" biological modules, how is a theory of mind acquired? A clear alternative to modular proposals like Baron-Cohen's is a model in which processes contributing to mindreading (e.g., intentionality detection, eye-direction detection, etc.) are viewed as skills constructed through world experience and developmentally reorganized as knowledge and experience accumulate. We do not have the space here to provide a detailed account of such a model. What is most distinctive about it, however, is that the end-state system might well function via the contribution of semi-independent component processes, and yet have achieved this high level of functioning through an *acquired* appreciation of different factors such as eye-gaze, goals, intentions, and beliefs, and their interrelations. To make this more concrete, abilities such as eye-direction detection might emerge as infants come to recognize the practical importance of gaze-direction and its role in predicting others' actions. Inferences about intentionality might emerge as different kinds of actions begin to be distinguished – those that display a predictable coalescence of action patterns with cues to emotion versus those that do not. Shared attention abilities might arise as the link between eye-direction and intentionality comes to be appreciated, while inferences about others' epistemic mental states might develop as children come to realize that inferences about intentionality alone cannot provide a fully coherent account of the confluence of action, outcome, and emotion observed in others' behavior.

#### "Modularity-lite" and falsifiability

We end this commentary by raising some larger concerns with respect to the falsifiability of Baron-Cohen's particular version of modularity. As we have already mentioned, Baron-Cohen dismisses

three of the criteria that Fodor originally proposed as central to modularity. Arguing that modules might usefully interact, informational encapsulation is rejected, while the shallow output and inaccessibility to consciousness criteria are reformulated as empirical questions. Baron-Cohen's reduced-fat version of modularity, however, like most "lite" products, is not entirely palatable. In particular, the slimming process renders the modularity claim increasingly difficult to discriminate from its competitors and, hence, increasingly difficult to falsify.

For one thing, as we have already argued, an initially *non-modular* system could well display precisely the six characteristics that Baron-Cohen considers criterial for attributing modularity. This clarifies an important asymmetry in Baron-Cohen's criteria: Although it would seem critical that the criteria be satisfied to bolster any claim about biologically-based modules, the satisfaction of these six criteria is not, in and of itself, sufficient to guarantee the validity of such a claim. As far as we can tell, the only characteristic of Fodor's original set that is truly incompatible with a model in which processing is achieved by all-purpose mechanisms is the discarded informational encapsulation criterion, and that is so only if informational encapsulation can be shown to hold during the acquisition phase. We say this because, as previously discussed, general-purpose learning mechanisms might play a role in the construction of automatized components that, in the end-state system, function largely as informationally encapsulated modules. Clearly, for such a general-purpose system to function, it would be necessary that, during acquisition, higher-level processes initiate and influence the construction of what would ultimately become independently functioning modularized components. Hence, if it could be shown that the system is informationally encapsulated from the beginning, a strong case for modularity would be established. Interestingly, however, recent evidence indicates that at least one of Baron-Cohen's proposed modules – EDD – does not meet the informational encapsulation criterion during acquisition (Veebra & Johnson, in press).

Finally, we are not entirely certain whether Baron-Cohen views the satisfaction of his criteria as *necessary* for establishing modularity. At times, he notes that a particular module does not meet all the criteria, yet he concludes that the evidence remains consistent with his proposal. For example, although he believes that ID does not satisfy the domain specificity criterion, he nevertheless continues to think of it as a module. Parenthetically, it also seems that SAM would not meet this criterion because it can apparently take more than one kind of input (from EDD or ID). In any case, if Baron-Cohen does have something

weaker in mind than a set of necessary criteria, then modularity-lite includes a hidden fudge-factor, in which case the proposal becomes virtually impossible to falsify.

#### REFERENCES

- Corkum, V., & Moore, C. (in press). Development of joint visual attention in infants. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development*. Hillsdale, NJ: Lawrence Erlbaum.
- Farah, M. (1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain specificity in visual object recognition. *Current Directions in Psychological Science, 1*, 164-169.
- Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science, 15*, 219-250.
- Vecera, S. P., & Johnson, M. H. (in press). Eye gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition*.

