If human cognition is adaptive, can human knowledge consist of encodings?

Robert L. Campbell[a] and Mark H. Bickhard[b]
[a]IBM T. J. Watson Research Center, Yorktown Heights, NY 10598; and
[b]Department of Psychology, Lehigh University, Bethlehem, PA 18015
Electronic mail: [a]rlc@ibm.com. or rlc@yktvmh.bitnet;
[b]mhb0@lehigh.bitnet

Have assumptions about mechanisms been avoided? Anderson asserts that a "rational" analysis of cognition can be separated from an "algorithmic" or "mechanistic" account (sect. 1.1), and that this amounts to "the framing of the information-processing problem . . . a nearly mechanism-free casting of a psychological theory." In his analyses of cognitive functions such as memory and categorization "the computational assumptions are indeed weak, involving claims that almost all information-processing theories would agree on" (Anderson 1990, p. 36). Weak though these assumptions may be, they still have consequences. Anderson's project is haunted by the ghosts of mechanisms that he has not yet exorcised from his "rational" level of analysis.

Is rational analysis committed to encodings? Anderson equates knowledge with *encodings.* By encodings, we mean objects, events, or structures of objects, in the mind that represent objects, or structures of objects, in the world and do so by correspondence (Bickhard & Campbell 1989; Bickhard & Richie 1983). In Anderson's (1983) ACT framework, declarative knowledge consists of hierarchical structures of encoding elements (the structures can be temporal, spatial, or propositional). Procedural knowledge consists of encoded production rules whose encoded conditions must be matched with symbols in working memory. Following Pylyshyn (1984) and Newell (1980), Anderson affirms that the algorithmic level is psychologically real and that what happens there is computations on symbols (Anderson 1990). Although Anderson does not mention it, he is also endorsing Fodor's (1975) "representational theory of mind," with consequences that we will explore below.

Anderson's specific rational analyses presume some obvious encoding atoms (such as "memory traces," sect. 2.2), or objects in the environment for such atoms to correspond to (such as discrete objects with discrete features, already clustered in predictively useful ways, sect. 3), so they too are hardly free of assumptions about mechanism.

In fact, Anderson takes encodingism for granted. Though he admits that "it has become apparent to me that this rational analysis has assumed the general ACT framework, if not the ACT* theory" (Anderson 1990, p. xi), the ACT assumptions, when finally enumerated, consist of such things as "a system in which memories are retrieved and tested for appropriateness" (p. 252). In such "weak" assumptions, encodingism is too deeply presupposed to be mentioned.

Is encodingism tenable? We have argued (Bickhard 1980; Bickhard & Campbell 1989; Bickhard & Richie 1983) that any framework that treats encodings as an irreducible, foundational form of representation is untenable. Encodings have to derive from some other form of representation, because they presuppose knowledge of what they are supposed to correspond to.

Foundational encodings are ubiquitous in the ACT framework. Two kinds of declarative encodings – temporal strings and spatial images – are hierarchical structures of elements that are held to encode the environment directly (preserving temporal or spatial sequence), whereas abstract propositions recode environmental structures by a process that has to be learned (Anderson 1983). Hence, for Anderson, temporal strings and spatial images appear to be foundational encodings, and their elements must be, along with the elements of abstract propositions.

Our argument against foundational encodings runs as follows: In the clear cases (Morse code, digital audio, etc.), encodings stand in for some other form of representation: X encodes Y means that X represents the same thing that Y represents. The encoding relationship presupposes that Y already represents something. Y might be an encoding itself, but if it is, it must stand in for another representation Z. The regress has to stop somewhere, and it cannot stop with an encoding. If Z is a foundational encoding, it must stand in for something already known, yet Z is supposed to be the means by which that thing is known. But "Z represents the same thing that Z represents" does not define an encoding. Hence encodings cannot be a foundational form of knowledge (Bickhard, in press a; in press b).

The incoherence of foundational encodings has been partially recognized by quite a few thinkers. For instance, Piaget (1970) argued that perception could not be a copy of structures in the world, but he did not extend the argument to concrete and formal operational structures, and Harnad (1990) has argued that digital encodings (symbols) cannot be a foundational form of knowledge, but he has not extended the argument to analog encodings (such as spatial images).

Can encodings develop? It follows directly from the incoherence argument that genuinely novel representations are impossible within an encoding framework. A fundamentally new kind of encoding cannot be acquired, because it would have to be defined in terms of the new kind of thing that it represents, yet that kind of thing supposedly cannot be known without the encoding (Bickhard, in press a; Campbell & Bickhard 1987).

The impossibility of novel encodings is best illustrated by the work of Fodor (1975; 1981). From Fodor's standpoint, all encoding approaches must posit an innate set of primitive encodings.

All that distinguishes such approaches is the extent to which "complex" encodings can be defined as simple combinations of the primitive encodings.

Fodor (1981) argues that, while "phrasal concepts" (e.g., sentences) may be built out of primitive encodings, "lexical concepts" cannot be, and must therefore be primitive encodings themselves. In Anderson's terms, this would be a claim that structures of encoding elements, such as phrase units, image units, or propositions, can be built of more basic encodings (it is not clear that they always are), but individual elements, – for example, words, basic subimages, and concepts – are not further reducible and are therefore primitive encodings.

Fodor's innate concepts cannot be learned; they must already be present to figure in any encoded hypothesis. Nor can they be products of any constructive developmental process (Fodor 1980). Fodor (1981) is forced to posit a process of "triggering," extrinsic to the passive built-in encodings, which elicits the activation of innate concepts through sensory conditions or the prior activation of other innate concepts.

Can encodings evolve? Positing innate encodings, however, just shifts the burden of their construction from the development of the individual to the evolution of the species. And Fodor's arguments imply that the acquisition of novel encodings through hypothesis testing is impossible in principle. Fodor's arguments lead to the conclusion that *if evolution is a variation and selection process, there is no way for encodings to evolve.* Though ambivalent about evolution, Anderson does treat it as a variation and selection process (sect. 1.1). He is occasionally willing to consider evolutionary constraints on the differentiation of "new representational types . . . there must reasonably have been time in our evolutionary history to create such a representation and an adaptive advantage to doing so" (1983, p. 46).

Given his commitment to encodingism and to evolution as variation and selection, Anderson is therefore obliged to (1) identify errors in Fodor's reasoning, (2) propose an alternative to encodingism, or (3) embrace Fodor's conclusions.

Preserving encodingism while refuting Fodor. To refute Fodor, Anderson would have to show that there is a process compatible with the rest of his theory that can generate emergent representation: representation constituted out of phenomena that are not themselves representational. This would be an uphill fight: A recent survey of production system models has concluded that the constructive processes invoked in such models are not even capable of generating new goals or radically reorganizing algorithms (Neches et al. 1987). Because Fodor's conclusions follow from the weakest assumptions of encodingism, Anderson would be hard pressed to avoid Fodor's *reductiones ad absurdum* while retaining anything like the ACT framework.

Replacing encodingism. Rejecting encodingism, in Anderson's case, would mean rejecting the physical symbol system hypothesis. Information-processing (IP) modelers would then have to venture into completely unexplored territory. There is not only the challenge of coming up with an alternative account of representation that avoids the stumbling blocks of encodingism and provides for the emergence of representation from something nonrepresentational. There is also the task of tracing its ramifications.

Convergently with some others (e.g., Brooks 1987), we have been engaged in this sort of effort for some time. An account of our alternative, *interactive representation,* would overflow this commentary. We would just like to point out that replacing encodings with interactive representation has forced changes throughout our conception of cognition, from learning to language to developmental stages to consciousness to psychopathology and beyond (Bickhard, in press b; Campbell & Bickhard 1986). For instance, traditional views of language as the recoding, transmission, and decoding of encoded messages cannot be maintained in an interactivist approach (Bickhard 1980a; 1987), hence language learning cannot be presumed to start with pairings of utterances and encoded meanings (as it is by Anderson 1983).

It would be most convenient if the changes that ensue from the adoption of nonencoded representations could be bottled up in a preprocessing stage, which converts everything into encodings, allowing computational business to go on as usual. Harnad (1990), for instance, proposes that perceptual categorization yields meaningful symbols, which can thenceforward be processed in the conventional fashion. But non-encoding-based sensory processing can't be divided this way (Bickhard & Richie 1983).

Embracing Fodor's nativism. The contemporary practice of IP modeling is already nativist, albeit unwittingly. Researchers simply introduce new elements of declarative representation whenever necessary to model a phenomenon, without considering their learnability. IP modelers could stipulate that any primitives introduced for modeling purposes are innately present and must be activated by triggering. Such a move, however, would wreak havoc on the empiricist allegiances usually professed by IP modelers. It would also restrict evolutionary constraints to operating on the generation and selection of combinations of encoding atoms, while leaving unsettled the question, ignored by Fodor, of the possible evolutionary origins of the primitive encodings.

Conclusion. It is certainly desirable to introduce evolutionary constraints, and constraints of optimization to the environment, into cognitive science, and we salute Anderson for doing so. But there is considerable irony in the introduction of such considerations within an encoding-based framework, which makes the evolution of mental representation impossible.

ACKNOWLEDGMENTS

Thanks to Jack Carroll, Steve Payne, and Kevin Singley for their comments.