Representational redescription and cognitive architectures

Antonella Carassa[1] and Maurizio Tirassa[2,3]

1. Dipartimento di Psicologia Generale • Università di Padova
2. Centro di Scienza Cognitiva • Università di Torino
3. Istituto di Psicologia della Facoltà Medica • Università di Milano

Abstract. We focus on Karmiloff-Smith's Representational redescription model, arguing that it poses some problems concerning the architecture of a redescribing system. To discuss the topic, we consider the implicit/explicit dichotomy and the relations between natural language and the language of thought. We argue that the model regards how knowledge is employed rather than how it is represented in the system.

Karmiloff-Smith's Representational Redescription (RR) model (1992) accounts for changes across learning and development in terms of knowledge translation into increasingly higher level formats.

A basic feature of the model is the distinction between implicit, procedural and explicit, declarative knowledge (respectively, formats I and E1/E3, in terms of RR). Learning of procedures might well be the first to take place when an individual starts to explore a new domain; but this implies nothing about how the relevant knowledge is actually represented. Procedural representations should be distinguished from (possibly declarative) representations of procedures (Rumelhart & Norman 1985). The former are actually encapsulated, similarly to what RR posits; on the other hand, it is hard to see how they might be decomposed into basic steps for redescription, if not through an "external" observation of one's own actions (which would open a sort of regression, because it requires to procedurally represent one's own procedures, and so on). Representations of procedures, on the contrary, if expressed declaratively, are more easily decomposable, the only problem, so to say, being the removal of "brackets" around them. Thus, decomposition requires no translation into different codes. Another advantage of using a single declarative code is that it eliminates the necessity that any learning starts from a procedural phase.

A further problem with declarative/explicit formats is their relation with natural language. Whatever their nature, mental representations have little to do with it. Language-of-thought theories (e.g., Fodor 1975) simply assert that thought shares a number of properties with natural languages, such as constituency, compositionality and so on (Fodor & Pylyshyn 1988). The distinction between natural languages and the language of thought (LOT) is mandatory for many obvious reasons: e.g., because it is difficult to imagine Italians, French etc., each having their own LOT; and there is no reason to postulate that thought requires anything like an "inner talk" beyond the simple use of representations. But then, what does it mean to say that formats E2 and E3 differ in that the latter is "close enough to natural language for easy translation into statable, communicable form" (p. 23)? If it just means that E3 is mentalese, what differentiates the three explicit codes, conscious access apart? Is it the case that there are two (or three) different

kinds of mentalese? This is not a theoretical impossibility, but needs a careful consideration: if a single LOT poses a great deal of problems, what will it be like with a number of LOTs? Another critical point regards the formation of inter-domain links which, according to Karmiloff-Smith, depends on the usage of a pseudolinguistic (E3) format; we see no necessity to postulate this, both because such links only require a LOT with constituency, and because there is no reason to exclude the existence of cross-domain links in higher animals.

As regards the latter point, we would like to argue against the idea that the chimpanzee has to stop at level I (i.e., that "it does not go beyond behavioral mastery", p. 192). This seems too strong a statement, at least because at level I "procedural-like representations are simply used in response to external stimuli" (p. 22), and such procedures are "independently stored" (p. 20), i.e., no generalization is possible out of them, nor is any cross-domain link possible. There is a great deal of discussion of these topics; our position is that it is premature to exclude the possibility of higher level representations in animals (see Prato Previde *et al.*, 1992, for a discussion).

All of our considerations bring us to the underlying topic of the architecture hypothesized of the system. The RR model posits three or four different kinds of codes, and a corresponding number of translating devices. Thus, there must be a great deal of machinery (which must be built-in, since the system cannot be expected to learn new codes); furthermore, a huge control structure is needed, in charge of deciding what to redescribe, and when. The general architecture of the system is not described in detail, so it is difficult to assess its plausibility. An obvious problem regards the overall efficiency of the system; though we agree on not being obsessed with parsimony, is it necessary to postulate all this stuff?

Our suggestion is that the content of representations matters more than their format in determining their implicitness or explicitness, which might be viewed as just the presence or the absence of "brackets" around the content itself. The interpreter of representations must be very complicated anyway, so why believe that the representations it operates on are complicated as well?

Analogous considerations hold for conscious accessibility: why should it depend on the use of a particular code, rather than on properties of the consciousness itself? When we first learn procedures in a new domain, we are deeply (and painfully) conscious of the difficulties of the task, of our need for better knowledge, and so on. Thus, unavailability to conscious access only regards the content of the representation, and not the representation itself; i.e., we have metaknowledge of these representations, though we are unable to consciously decompose ("unbracket") them. In the end, the content of any representation, once reduced to its basic components, is no longer accessible to consciousness. So, one might think that the only difference between inaccessible and accessible representations (I/E1 and E2/E3, respectively) consists in the possibility to further decompose their contents.

To conclude, the RR model seems to deal with how knowledge contents are used, rather than with how they are actually represented in the system: in Newell's terms (1982), it lies on the knowledge level, rather than on the symbol level; when one goes down to the latter, some problems start to show. This does not mean that the RR model should be discarded: the psychological evidence it brings is impressive, and the psychological plausibility of the model is not impaired by our considerations; but much work will be needed to answer questions about the architecture of a RR system.

References

Fodor, J.A. (1975) *The language of thought*. Crowell.
Fodor, J.A. & Pylyshyn, Z.W. (1988) Connectionism and cognitive architecture. *Cognition* 28: 3-71.
Karmiloff-Smith, A. (1992) *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.
Newell, A. (1982) The knowledge level. *Artificial Intelligence* 18:87-127.
Prato Previde, E., Colombetti, M., Poli, M. & Cenami Spada, E. (1992) The mind of organisms: some issues about animal cognition. *International Journal of Comparative Psychology* 6: 79-100 and 116-119.
Rumelhart, D.E. & Norman D.A. (1985) Representation of knowledge. In: *Issues in cognitive modeling*, eds. A.M. Aitkenhead & J.M. Slack. Lawrence Erlbaum Associates.