Intellectica, 2003, 36-37, pp. 381-384

John Stewart*

External vs. Mental Representations

The article by Gennaro Auletta[1] gives an interesting new twist to a long-standing debate in cognitive science as to the nature and status of representations. His arguments appear to me clear and convincing in the case of "external representations", i.e. the case when both the representation and the referent are external to an intentional subject (or subjects) - for example, the countryside and a map, or Chirac and a photo of Chirac. In this case, the intentional agent (for example, the map-maker) is indeed able to set up the appropriate correspondence relations, by procedures that are neither magical nor particularly mysterious. In addition, dependent on the context of use, the relation between representation and referent (which is which) can indeed be reversed; and this reversibility indicates that the relationship cannot be intrinsically ("ontologically") causal. So far, so good.

My problem comes when we seek to apply this to the case of *mental* representations and *external* referents. The question, raised also by Palma in his commentary, is whether it is tenable to consider that external representations on the one hand, mental states on the other, are entities of the same essential class. This point is not answered by appealing to a "General semiotics", i.e. a theory of signs which treats linguistic signs - words and phrases - as a special case of a more general class of signs which includes animal signals. Human language did not arise "out of the blue" (nor by a genetic mutation giving rise to "innate grammar" as Chomsky implicitly suggests); it arose, surely, on the basis of pre-existing systems of animal communication in which the "signs" were calls, bodily gestures, insect pheromones, etc. However, the point is that in a "General semiotics" of this sort, the signs *are always external*: in the case of language, the signs are words which clearly have a material external instantiation, be it phonological sounds in the case of oral language or written traces in the case of written language; and of course non-linguistics signs (or signals) are also external. Thus, the validity of a "General semiotics" (which I think can be argued for) in no way justifies treating external signs (or "representations") as being the same type of entity as internal mental states.

In order to pursue the discussion, there are two possibilities: either i) one rejects the classical computational theory of mind (CTM) in favour of a radical alternative (e.g. constructivist enaction);

---

* COSTECH, Université de Technologie de Compiègne.

[1] G. Auletta (2002). Is representation characterized by intrinsity and causality? *Intellectica*, 35, 83-113.

or ii) one adopts CTM (or some reformist variant). In my view, it is vital to make a clear and coherent choice; navigating vaguely between the two can only lead to confusion. Since my own preference is for (i), I will start by considering this option.

i) **Constructivist enaction**. On this view, it is the whole concept of "mental representations" which is simply thrown out. It may be objected that "mental representations" are necessary, on any view, in order to account for the possibility of anticipation. I agree that even on a constructivist view, relatively sophisticated cerebral/mental processes are necessary in order to anticipate the sensory consequences of motor actions: but it is misleading at best to call these processes "representations". They are certainly *not* of the same nature as external semiotic signs.

ii) **CTM**. In the classical computational theory of mind, the mental items are primarily well-formed formula of purely formal symbols, and cognition consists of operations on these symbols according to purely syntactical rules. This is precisely what guarantees that the operations can be carried out *mechanically* (by an appropriate formal equivalent of a Turing machine), and hence provides a solution to the mind-body problem. It is then necessary to provide these intrinsically meaningless symbols with semantic content by setting up appropriate correspondence relations with external referents: by this means, the symbolic formula become indeed "representations". The difficulties in doing this constitute the well-known "symbol-grounding problem". Now it seems to me that what Auletta proposes for the relationship between *external* representations and *external* referents cannot work here: there *is* no "intentional agent" able to set up the appropriate relations between an *internal* "mental representation" and an *external* "referent". The "intentional agent" certainly cannot be the cognitive subject himself, because this subject has access to only one of the terms; he has no possible access to the "thing in itself", *other* than his own *representation of* the referent. To put it another way, more germane to Auletta's argument, the relationship between "representation" and "referent" is *not* reversible and ontologically symmetrical. Who are the intentional agents who reverse the relationship by taking an external entity as a "representation" of the "mental state" of a cognitive subject? Note that for such intentional agents, the "mental state" of another subject is intrinsically unobservable (cf. Fodor's "methodological solipsism"); they are thus also in the situation of having access to only one of the terms of the relation.

The rules of the game here are to accept as a basic paradigmatic postulate that "cognition" will be *defined* as syntactically regulated operations on formal symbols. The exercise then consists of adjusting all the "auxiliary" hypotheses so as to make this postulate tenable. According to the CTM, cognition takes place in the domain

of a "language of thought" (Fodor), which is akin to the formal languages of the propositional calculus, the predicate calculus, etc. developed by the formalist school of mathematics following Hilbert. And, again according to the CTM, natural human language is akin to these formal languages (cf. Chomsky). Then, if one were to admit a "General semiotics", linguistic signs would be akin to signs in general - including photos, maps etc. where (as Auletta argues) the representation-referent relationship is reversible (and hence cannot be causal). Coherent computationalists have to break the chain somewhere, since they cannot admit that the relationship between the symbols of the computational "language of thought" can stand in a reversible relationship to their referents. Where could they best make the break?

As I understand the CTM, the central point is the relationship between the formal symbols (and well-formed formula) of the formal language on one hand, and the "referents" (the "model" in the Model Theory of formal semantics) on the other. This relationship is not symmetrical or reversible. The formal symbols have very special properties: they are completely devoid of any intrinsic significance, so that they can be freely manipulated in complete accordance with the rules of the formal syntax (the theory of Universal Turing Machines guarantees that these syntactical operations can be instantiated mechanically). The "referents" do not have these special properties: they are intrinsically the source of "meaning", and correlatively cannot be freely manipulated by the rules of a formal syntax. Hence, the relationship cannot be "reversed".

If this is correct, where could a computationalist make the break? A constructivist would deny that natural languages are akin to formal languages; more precisely, he would say that the formal symbols involved in computational operations are entities that are simply not of the same type as external representations, so that arguments about the reversibility (or not) of the relation between external representations and their referents (the photo of Chirac and Chirac, the sign "Mont Blanc" and the mountain) are simply irrelevant. However, this option is hardly open to proponents of the CTM. Thus it may well be that the best break-point from the point of view of the CTM is between linguistic signs and "signs in general" - and this does indeed seem to be Palma's option. In other words: it is by virtue of the fact of being amenable to syntactical manipulations, that linguistic signs necessarily lose their relationship of "reversibility" with their referents. Let us come back to the example of "Mont Blanc" and the mountain. If "Mont Blanc" is painted on a road-sign, Auletta's point about the reversibility with the mountain (which is representation and which is referent) works; but the *symbol-string* "Mont Blanc" can enter into syntactical operations (to take Palma's example of negation: "Mont Blanc is *not* in the Himalayas") whereas this is not the case for the mountain. Hence a basic asymmetry.

To sum up: constructivists will predictably welcome Auletta's arguments, although they will have some reservations about considering that external semiotic signs are entities of the same sort as internal mental states or processes. Computationalists, on the other hand, will predictably find Auletta's arguments unacceptable. In either case, however, the debate seems interesting.