INTRODUCTION

Alberto Greco

University of Genoa, Italy

1. General Aspects of the Concept of Representation

If one has to judge the importance of a subject from the frequency of use of certain words in the literature, the concept of representation seems to have a special position in disciplines belonging to cognitive science. Indeed, the word "representation" is certainly one of the most used. Unfortunately, however, it is also one of the least clearly defined, perhaps because it comes from common sense, and almost everyone believes to know what it means. No single definition at present seems to capture all aspects considered relevant by someone. The aim of this special issue is to consider the present status of this concept, in a multidisciplinary perspective, and to discuss some of its most controversial aspects. It includes revised versions of papers that were presented at the special Workshop of the ESSCS (*Representation, a multi-disciplinary perspective* – Camogli, Italy, 12–14 April 1994), but there are other contributions as well. The aim of this introduction is to describe some of the problems in general, and how they are tackled in the papers.

The first aspect, of course, concerns the very definition of the concept. It constitutes a problem because perhaps the only thing that everyone agrees on is the fact that "representation" is *not* a completely clear and unambiguous term. The simplest and most common defining sense of the term is: "any thing that stands for something else". This sense is also the widest, so wide that it can include either internal processes of organisms, or concrete things that stand for other things like road maps, knots in handkerchiefs, ink marks on paper, or physical and electronic systems. This definition is functional, since "standing for" is intended as a *substitution function*, or – if one prefers – a reference function: no representation without reference. The differences between different uses of the concept of representation arise because – at least in these general terms – the reason *why there should be* such a substitution and how it works are unclear matters, left to different single interpretations. The multidisciplinary nature of the concept comes from this width of perspective. The substitution function is the concept uniting the different perspectives. For example, psychology and philosophy, in their models, often assume that concrete things or systems may stand for something else to which

Cognitive Systems 4–2, August 1995, 119–129

they have a similarity. This idea is extended to internal representations: in the head there should be "something" that is similar to what is to be represented, or something that could have the same function. Two other disciplines, artificial intelligence and connectionism, often assume that the most natural way for evaluating internal processes in an artificially intelligent system is to compare these processes with human representational functions, and then decide whether they are similar or not.

Ideally, psychological or philosophical approaches make reference to mental states, whereas other approaches refer to concrete things. But this is only an ideal sketch since humans can also be described as concrete things, as biological or even physical systems. In adopting the substitution or reference function, and assuming the existence of some *internal state* of the organism which is systematically related to objects or events in the world, there still is freedom to depict such internal state in many ways. Psychology and philosophy may speak of mental states, artificial intelligence of production rules, propositions, or other computational entities, and connectionism of neural processes.

In the different approaches to, and uses of the term 'representation', there are the same old different positions well–known from many psychological problems, extending from the more "mentalistic" to the more "physicalistic". One of the most common views about representations is that if they are internal states whose function, after all, is to refer, then the best thing to do is to turn to disciplines which are more professionally entitled to deal with problems of reference. Semiotics and linguistics have supplied a general framework for making handling reference. In the theoretical context of the so–called "cognitive revolution" in psychology and in the cognitive sciences, the theme of representation has acquired a crucial role because "knowledge" required for explanation of behaviour is expressed in terms of *symbols* manipulated according to formal rules. Also the description of mental processes as information–processing is nothing else than an effort to make explicit these processes in terms of formally defined procedures (programs). Since rules and programs concern the manipulation of symbols and they themselves can be expressed in symbolic terms, it became important to consider the problem of represention of these symbols. The hypothesis of the *language of thought* (Fodor, 1975), for example, tried to satisfy exactly this need.

This is why the concept of representation has become more and more linked with the adjective "symbolic", at least in that part of cognitive science that has inherited the role of classical cognitivism as far as it is committed to the computational approach. However, this connection is now questioned by many. As is well–known, this is one of the questions typical of connectionist views, but other perspectives also ask the question. There also exist the so–called *hybrid* approaches, which try to exploit the advantages of both computational and connectionist systems, and to avoid giving up the concept of symbolic

representation. As a consequence it runs into the problem of how to redefine the concept to accomodate the different aspects.

In the *system–dynamics* approach it is believed that the symbolic representation concept should be overcome. According to this perspective, the activity of a cognitive system is best described as a dynamic interaction with its environment, in a continuous flow of exchanges where situation and time are important and symbolic processing would be too abstract and limited to explain cognition.

Of course, there are many implications of these issues and many other points but, in sum, the current debate is focused on a few key points. More specifically, the main questions under discussion are:

1) does a correspondence between environmental inputs (or stimuli) and system's internal representations exist?

2) in the affirmative case, is it direct, in the sense that there is a direct mapping of environmental characteristics into meanings?

3) what role do biological processes play in the process of acquiring internal representations from stimuli? Is there more than neural (or sensor) states in sensory information?

4) what is the relationship between symbolic (or high–level) representations and sensory or neural processes? Are there low–level "representations"? Does symbolic emerge from non–symbolic? How? (This question – whether representation has symbolic or nonsymbolic nature – has reached a considerable importance, as witnessed by the fact that almost all papers in this issue raise, more or less, problems related to this challenge).

The first paper concerns the relationship between a representing system and represented things. No theory of representation can escape this problem. According to the classical view, implicitly based on insights from folk psychology, there should be a systematic, perhaps isomorphic, correspondence between such two domains. One possibility for explaining the path from stimuli to knowledge is that the input be *encoded* in some way. According to this perspective, the process of representing would be a *function* that performs such encoding. In the first paper, by **Gelepithis**, this view is criticised. He specifically directs his criticism against the version expressed, among others, by Newell (1990).

Newell was not interested in all types of knowledge, but especially in the knowledge that systems have about the external world. This knowledge is relevant in pursuing goals, or in producing intelligent behaviour. Such knowledge is used in problem solving, and mainly concerns external situations changing in time. The basic operation is, in Newell's view, the transformation of entities: entity X being transformed into another entity Y. Knowledge about this situation ideally requires that both the original entity X and the transformation T are

encoded in the internal medium of the system (X is encoded into x, T into t) in such a way that the internal transformation t in the medium produces a new internal entity y, corresponding to Y. In such a way, Newell introduced a particular notion of representation, called *representational law*. In order for such a schema to work, both the internal entity x and the internal transformation t should be analogues of the external counterparts. But actually this is not always possible, because external situations are much more complex than what can be captured by simple internal analogues. In practice, representation can only work if the internal medium is seen as a structure of combinatorially manipulable substates, in which complex states and transformations are obtained by composing internal entities (which can, obviously, be named "symbols").

Gelepithis claims that this view is too restricted, because it implies seeing representation as a function that directly puts in correspondence external states with internal ones, which does not always correspond to what happens in reality. In his opinion, this perspective is also biased by the wrong assumption that the typical way of conceiving human representation (as symbolic and composable) is the only perspective possible. The venue of parallel–distributed systems (PDP), more or less a synonym for connectionist systems, shows that there exist different kinds of representation. A connectionist representation is essentially a machine representation, not translatable into human primitives (not in symbolic terms, and not even in behavioural terms or brain terms).

The discourse then raises wide questions like the formalisability of knowledge, and the question how a system selects the appropriate representation in different situations. Gelepithis' proposal is to give up conceiving representation as a *function* that maps the external world onto an internal representation, and instead to view representation as an internal state in *relation* with the original. This relation has the standard features of models, i.e. simplification and preservation of essential characteristics. From the viewpoint of human cognition, this claim has two interesting implications: 1) the "linguistic" bias on representation is overcome, since encoding/decoding is not necessary anymore; 2) representation is conceived as a *new* phenomenon, *different* from the one that it is intended to represent, and as such it can be studied on his own.

The same questions stated above about the correspondence between the representing system and represented things are faced in the paper by **Peschl**. He proposes to give up the idea of a direct or isomorphic correspondence, since there is no simple correspondence between external features and representations (it is possible to have different representations for a single feature, and different features can be represented by a single representation). The paper's proposal is to look instead at the neural substratum, which has the explicit function to generate behaviour through a dynamic adaptation to its environment. Neural structures can be considered *recurrent neural networks*, which have the nice property – appealing since it looks natural – of interacting both with stimuli and with the internal states of the system. Peschl emphasizes

the absence of a stable correspondence between external features and internal states, basing his reasoning on an analogy between neural networks and finite automata. He concludes, then, that intelligent behaviour does not need representation nor semantic categories, but that it only is a product of mutual adaptation between two interacting systems: a cognitive system and its environment. This position is close to that of the system–dynamics theorists. In this context, Peschl argues, we can talk of "representation without representations", where representation seems to be knowledge belonging mainly to external observers, not the observed system where it is "embodied" (Note 1).

Also **Sommerhoff** has a stand, at least in principle, close to the neural point of view, where he repeatedly speaks of brain representations and states of the brain, but – in my opinion – his approach is closer to psychology than to neurology. Also, although he is specifically concerned with the problem whether representation is involved in vision, his treatment touches upon more general points. Sommerhoff's discussion is more psychological than neural because it goes beyond the problem of correspondence between input and representation, to face the hard question of how to go from inputs to meaning. Sommerhoff reminds us that perceiving is always *perceiving as* and that inputs must be interpreted. In perception, therefore, sensory information is not all that is important, but also a general model of the world is involved, which basically includes or gives rise to *expectations*. But this makes things complicated because having expectations implies comparing every single new input with some internal model. How does this process work? As we have previously mentioned (see **Gelepithis**' discussion) a popular view is that inputs are coded in symbolic form. This would make it easier to deal with the problem of comparison between new inputs and given memories, since it would be a standard computational problem of symbolic pattern–matching. Sommerhoff notes that Marr's theory of feature detectors encouraged an interpretation in symbolic terms of the output of cells in the visual cortex. The main shortcoming of symbolic theories is that symbols cannot refer only to other symbols, but sooner or later they must refer to something else (this is the so–called *symbol–grounding problem*). This something else, in Sommerhoff's view, must be primarily some response pattern (e.g. some neural or motor activity) that directs behaviour. Differently from others (e.g. Harnad, 1990), he does not consider symbol–grounding as a question of extracting invariants (categories) out of sensory patterns and connecting them with a system of labels. Rather, detecting new information is a sort of automatic reaction based on the outcome of previous experiences. Here Sommerhoff makes reference to his concept of *conditional expectancy*, rigorously defined elsewhere (e.g. in Sommerhoff, 1990). In essence, there is no need for an autonomous system of representations; the only internal thing we can find is that the brain can anticipate what will happen: in Sommerhoff's account the thing most similar to representations is the totality of expectations (simple or perhaps complex) elicited by a certain input (Note 2)

The next point to be examined concerns the role of biological processes like sensory activities or nervous processes in cognition. One trend in present cognitive science, compared to the past, is towards considering cognition as being more *embodied*. But what the concept of "embodiment" means is not so clear. It could be only a warning not to forget the situatedness of cognitive systems as physical systems plunged in an environment and having physico–chemical exchanges which logically come before informational ones. It is an open question if this is another form of reductionism of the mental (disembodied) to the physical (embodied). Computationalism and connectionism, including hybrid approaches, typically consider representations in *disembodied* systems. **Etxeberria** in her paper discusses some shortcomings of such views. She asks whether the dynamical–systems approach can be a real alternative. In her opinion, the dynamical–systems perspective is closer to *embodiment* than others, but the real question is which aspects of the *body* are implied when speaking of "embodiment". Biologically oriented cognitive science, in fact, is inclined to neglect the body in considering activity in the nervous system. The emphasis is solely on "informational activity" under the assumption that it can be studied independently of energetic processes. Etxeberria shows that in living organisms the relevance or irrelevance of input may depend considerably on internal states directed to maintain homeostasis, i.e. energetic stability. Perception also tends to be studied in the context of action, for which it often is difficult to separate sensory and motor aspects. Etxeberria cites cases of bacteria which react to chemical substances "with all the body" and of rats that learn to avoid certain types of food because they felt thirsty after having tried little quantities. In such cases it is difficult to say what the representation could be (taste?, thirst?), and where the sensory process is (perhaps the whole body acts as a sensor). Also for humans there are specific aspects of the environment which are meaningful – in the sense that the organism appropriately reacts to them, adapting itself – but for which we do not have a clear representation.

The correspondence between environmental stimuli and representations is no doubt mediated by sensors. What is not clear is how sensory information is used in order to achieve meaning or knowledge. One possibility is that sensors have a representational output. **Sommerhoff** criticised Marr's account of feature detectors having a symbolic output passed on to a computational device. **Etxeberria** broadens the field and claims that the question whether sensors have a symbolic output cannot be answered the same way for all sensors. For example, for vision more specialised and relatively autonomous routines exist than for olfaction. In any case, she argues, perhaps this is not the good question to ask, for it implies an unnecessary conception of sensors as simple "prostheses", separate from the rest of the system. Even vision is closely linked with the movement and expectations of the system concerned, as also **Sommerhoff** observes in speaking of *active* vision. A defence of

a non–symbolic approach to representation in complex and dynamical systems can be found also in the paper by **Patel**. Internal representations in complex systems should be flexible enough for coping with a continuously changing environment. Therefore they cannot be abstractly pre–set outside context as the symbolic account would imply, but constructed during the agent–environment interaction. Patel describes a simulation of a complex adaptive system which learns in a changing environment. The system is built as a neural network of which the architecture is not *a priori* determined but selected by means of a genetic algorithm: the most successful one in the actual task survives. In such a system there is no direct correspondence between specific network structures and specific behaviour of the network. The same structure may learn different things, and different structures may learn the same things. In Patel's view, the most "promising" way of conceiving representation is as a process in a dynamic system, not as a fixed structure. What this more concretely means, however, is left to further exploration.

If one wants to go deeper into the role of biological aspects in representation, one cannot limit himself to the analysis of the output of sensors or the issue of embodiment, but sooner or later one must tackle the issue of the physical basis of mental processes. **Dalenoort** in his paper considers many important problems concerning the difference between computational and physical representation. As is well–known, the computational metaphor (the human as an information processing system) has given cognitive science its vital nourishment, and has also been chosen as a means to overcome the mind–body dualism (see e.g. Pylyshyn, 1984). In that view mental activities are considered as software running on the body machine. Also the big discussion concerning "levels" of functioning (starting with Marr, 1982) has its roots in this metaphor. In a sense the terminology of computational thinking seems to provide shelter from the problems concerning physical processes, because one can always say that those problems concern others. In fact, Dalenoort argues, one can choose if he wants to *see* human representation as a computational *or* a physical process; one cannot see it both ways simultaneously. But this does not entitle us to say that one of them (not ours, obviously) is useless or even wrong. A complete knowledge can only come from both sides of the coin. This is the *raison d'être* of the multidisciplinary approach. Dalenoort's proposal looks like an implicit critique against those cognitive scientists who limit themselves to their narrow field, no matter whether at the top level or at the bottom level, and who think that knowing the details of research at the other level is a waste of time. There is a clear example: a programmer may ignore details of implementation, such as the speed of processors, or the amount of memory, but a *good* programmer must take these aspects into account as far as they are relevant. (Similarly a constructor of hardware has to know something about the software that will be run on his machine). Unfortunately, however, our knowledge of the implementation of symbolic

representations in our brains is far from adequate. When we ascribe new properties to matter at the low level, we usually start (at least at the very beginning) from observations and knowledge at the high level. After all, it is to explain macroscopic phenomena that theoretical entities concerning microscopic or hidden phenomena have been invented (in this sense, the starting point is always folk psychology). These new properties that have been "assigned" correspond to properties sometimes called "emergent". Dalenoort warns us that this does not mean that we *explain* high–level properties in terms of the low–level ones. Emergent properties are only new properties that can be seen after the observation of interactions at another level of the system. The operation of any system depends not only on its representations but also on its procedures, which in turn should be represented themselves in some way. Dalenoort deals also with this important point, showing the difference between computational rules and physical laws: only the former may lead to error and may be changed. Computational rules, or algorithms, are not always explicitly expressed or followed by a programmer, but they can be always extracted *a posteriori*. When there is no program, however, like in connectionist systems or in self–organising systems, this is not possible. This does not mean that in this case we have only physical laws: we can still speak of a rational organization, depending on the specific task at hand. We can achieve this result by adopting a constructive approach, i.e. putting together little modules that allow us to discover which architectures are best suited to provide understanding of the system.

As we have noted in the beginning, there is a discussion concerning the symbolic or nonsymbolic nature of representations. In general, nonsymbolic representations are gaining more and more consideration. **Gelepithis** criticises Newell's view of encoding just on the grounds that no account is taken of the existence of of nonsymbolic representation (he refers to "machine representation") which does not need any sort of encoding. **Sommerhoff** shows the shortcomings of Marr's theory of symbolic feature detectors, in relation to the symbol–grounding problem. **Peschl** asks for a replacement of symbolic structures of representation by some sort of system's dynamics, and also **Etxeberria** and **Patel** see cognition in the context of complex dynamic systems where there is little place for the symbolic account. The last two papers, by **De Vries** and **Greco**, face more directly the question of symbolic representation. In all the discussion the opposition between symbolic and nonsymbolic representation arises each time cognitive systems are considered at different levels: a low level – closer to neural or sensory processes – and a high level, closer to abstract, linguistic processes. The real problem is not whether the one or the other is the true concept of representation, but what the relationships between the two aspects are: are they different and incommensurable processes or do they constitute two parts of a larger process?

For example, it is tempting to consider high–level representations as *emerging* from low–level ones. Indeed, many theorists in the connectionist area have taken this position (Smolensky, 1988, perhaps has been the most influential). In general, emerging properties are seen as new features, which can be ascribed to a complex system at a high level of description, appearing as a product (or a by–product) of the low–level interaction of its components. But the concept of emergence is not so clear as it might seem at first sight (see also Greco, 1990): in particular, the problem is that high–level (meaningful and understandable) properties are explained on the basis of low–level activities which are meaningless and useless for someone interested in explaining the system's behaviour in psychological terms. **De Vries** in his paper shows that if one really wants to understand how a system self–organises and develops then he must take care of what happens at the low level (say, in the net) as a consequence of the behaviour of the system as a whole. According to this view, the converse of the standard interpretation of emergence may also be valid, hence the low–level activity can be affected by high–level properties (*downwards emergence*). The interesting idea is that we can start with an early state and then let the network develop by self–organisition. But we shall only be able to watch what is happening if we assign network elements a place relevant in the task: therefore a representation of a task is fundamental to understand the system's performance, and it is not only a simple matter of implementation.

De Vries describes a network, designed to perform a simple object identification task, where not only data (stimulus properties, etc.), but also procedures are represented (a *conceptual network*). Representations in this architecture are series of excitation loops, which are not symbolic but may be considered as equivalent to memories or concepts (one possible additional hypothesis, touched on by De Vries, is that when excitation is above a critical threshold then we are aware of the involved concept, otherwise it still works "below" our awareness). This system works because temporary connections are allowed between nodes, depending on context.

The important feature of De Vries' model is that elements have properties that depend on the whole organisation and that seem to be interpretable in psychologically significant terms. This could be an answer to the general question about what explanatory role representations have in high–level processes like learning. In my own paper I focus on psychological aspects of representation, raising problems of definition, as I make a general analysis of the concept's meaning, and attempt to tackle the problem from another angle, the symbolic–nonsymbolic issue. I claim that there is a different psychological significance between *representation* as a process of representing on one hand, and *representations* as the product of this process on the other hand. The concept of representation is appealing for folk psychology because there is a phenomenological evidence that internal states explain behaviour. Scientific

psychology translated this folk plea into the causal explanatory paradigm, postulating internal entities that act as causes. The causal power of these entities, however, in my opinion, may come from two different sources: either from the mere existence of internal entities, or because they are used by a different process of interpretation. In other words, internal representations may act either by virtue of a simple correspondence with external states of affairs, or by virtue of a more complex process of construction (as products of *representing*). My proposal is to distinguish between what, in fact, seem to be two different kinds of representations: 1) ones that *reflect* external reality, which exist passively, and which are meaningless on themselves, they are nonsymbolic, and 2) ones that *substitute* external reality, which are constructed, meaningful, symbolic. It seems difficult to reconcile these two aspects concepts in a single overall framework.

Notes

(As the editor, I take the liberty of adding some comments of mine, but in footnotes since authors did not have the opportunity of replying.)

1. It is easy to see that there are still representations and categories in **Peschl**'s account, namely in the observer. In this way the problem is only shifted. How these representations arise and are implemented in observer's neural structures (taken as a different cognitive system) is a question without answer. We don't know either how to handle the risk of having an infinite series of observers' observers. But the approach expressed in this paper is significant in other respects. Peschl's view is explicitly eliminative materialist. This conception puts again into the field old problems, attaining mainly to the question of what language we are to adopt in describing a working cognitive or psychological system. In my opinion, it seems perfectly legitimate to adopt either the objective, physicalist language of system dynamics, or the phenomenological language of system's self–description (when the system is a human). In different scientific contexts we can use either in order to explain different aspects of behaviour. The important thing, however, is not to deny that the phenomenological description might be relevant to explain behaviour in certain contexts. (Note to page 123).

2. My problem with conditional expectancies is that a "what–leads–to–what" expectancy can only work if these two "what"s are first recognised. And then the problem of how are they represented comes again. **Sommerhoff** says says (1990, p.101) that they are "an interpretation by the brain of current sensory inputs on the basis of past experiences". Is representation necessary to mentally "anticipate" events, then to have expectations? I think yes. Expectation and representation imply a certain degree of in tentionality, that we can attribute to things, animals, humans with different levels of confidence. Does a keyhole

"expect", is it "prepared", "ready" to recognise keys? we would not say so. Do bee dances imply expectations? We probably would say that bees are prepared to react to them, but in fact we would not say (if not metaphorically) that bees have dance expectations nor dance representations. (Note to pages 123–4).

References

Fodor, J.A. (1975), *The language of thought*, New York: Crowell.

Greco, A. (1990), Some remarks about connectionism in psychological simulation. *Cognitive Systems*, **2**, 359–372.

Harnad, S. (1990), The symbol grounding problem, *Physica D*, **42**, 335–346.

Marr, D. (1982), *Vision: a computational investigation into the human representation and processing of visual information*, San Francisco: Freeman.

Newell, A. (1990), *Unified theories of cognition*, Cambridge, Mass: Harvard University Press.

Pylyshyn, Z.W. (1984), *Computation and cognition, Toward a foundation for Cognitive Science*, Cambridge, Mass: MIT Press.

Smolensky, P. (1988), On the proper treatment of connectionism, *Behavioral and Brain Sciences*, **11**, 1–74.

Sommerhoff, G. (1990), *Life, brain and consciousness, New perceptions through targeted systems analysis*, Amsterdam: North–Holland.

Manuscript received in final form: 11–04–1995

Address author:

Dept. of Philosophy – Lab. of Psychology, University of Genoa,
Via Balbi, 4 ; 16126 Genova, Italy.

E–mail: greco@igecuniv.cisi.unige.it