Contributi di Ricerca in Psicologia e Pedagogia
Sezione di Psicologia e Pedagogia del Dipartimento di Filosofia
Università di Genova (1990), pp. 9-17

Alberto Greco

What kind of psychological processes can be modelled by a connectionist system? *

1. The concept of model

The question addressed in the title calls directly in consideration how the concept of "modeling" should be understood.

In general, to model an event[1] or an object of the world is to have another event or another object which shares some similarities with it, that is *analogous* to it.[2] This analogy can be realized in two ways: by *constructing* a model or by *using* something *as a model.*

To *construct* a model means to originate artificially a different event or object on the grounds of some analogy with the original, the natural one. In this case, before constructing the model one should be able to specify *what it is a model of*, and where the analogy lies. This is the ideal concept of model, in the original sense of "giving a shape" to something shapeless.

To *use* something *as a model*, instead, is to recognize a posteriori that some object or event already realized (perhaps for different purposes) shares some analogies with a natural object, and then consider the artifact as a model of this object.

There could be some differences between these two cases, since in *constructing* a model the analogy is the starting point; instead, when *using* something as a model the analogy comes subsequently, from an interpretation of the artifact. In the latter case, the analogies which give the artifact the nature of a model, not being intentionally set up, can be less systematic and often contingent; in this case, then, it could be difficult these analogies be stated very clearly, as required.

Traditional psychological models made using computer programs are usually "constructed". One has a symbolic description of a psychological event (for example, a

---

* Paper presented at the Workshop "Connectionism: bridge between mind and brain?", ZIF (Center for Interdisciplinary Research), University of Bielefeld, W.Germany, April 1990.

[1] To clarify my terminology: by world, I mean the environment of a system or some part of the system itself. I call what we find in the world an "event" if we are to stress its dynamic character (if we want to refer to something that "happens" in the world); I call it an "object" to refer to its static aspects. Sometimes - to be brief - I use only one of these terms, "event" or "object", but almost always referring to one implies referring to both. The difference between these two terms is often only a question of emphasis.

[2] Obviously, in a psychological model the events or objects of the world should have a psychological nature. From the psychological point of view, a mental process is an event, and its product (a concept, a schema, etc.) may be called an object.

theory about some operation or about a psychic process) and writes a program trying to realize a correspondence between this description and the program. Similarly, one has a symbolic description of a psychological object (for example, some concept, a representational structure like a schema, etc.) and tries to realize a correspondence between this object and contents of memories or, more generally, stored data. In this kind of model the interpretation given to the symbols in a program is crucial in determining the actual operation of the program. Different symbols, in fact, are different programs or different data. This way, saying what these systems do or what they represent is a simple business, since this has to be known in advance.

2. The place of a connectionist system as a psychological model

Some of the most important topics which are of interest from the psychological point of view are: how does a person select, perceive, learn, react, etc. when faced to a natural object or event in the world? Most of these topics are explained by current psychological theories (notably, by the human information processing approach) appealing to representations in which symbols are processed by a person (subject). Let's call this Subject's Symbolic Representation SSR (figure 1).

Connectionist Systems (CS), on their own, can be described as ordered, structured patterns of activities of large networks of units. There can be a functional correspondence between some event in the world and these systems: in fact, their initial activities may be random, but after a certain time they become ordered, and their order clearly reflects functionally the order of the world that the system is given as input.

We have seen that, in order to consider such systems as psychological models, we should be able to specify the analogy with psychological processes. How this could be done, however, is not clear.

As is well known, the operation of a connectionist system does not depend on the

Figure 1

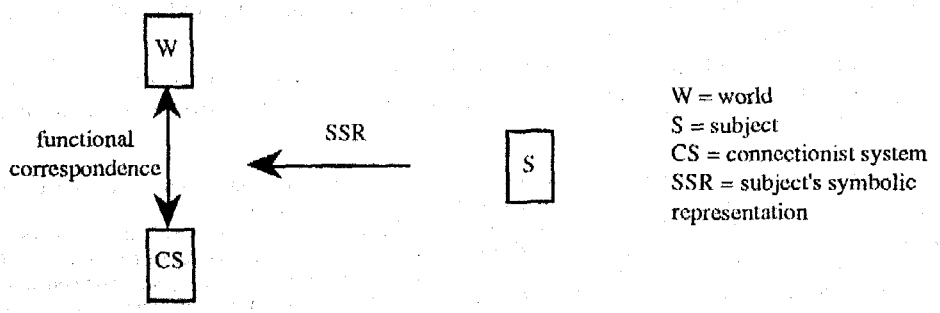
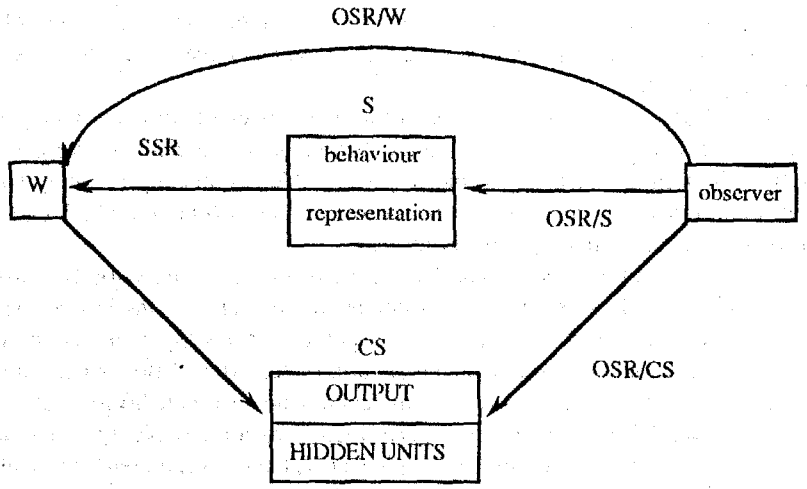


Figure 2



W, S, CS, SSR: see fig. 1
OSR = observer's symbolic representations

interpretation of its states (that is, on the state of its units in different moments), since a network can behave the same way irrespectively of what it is interpreted as doing. This has often been considered a limit of connectionist systems as psychological models. A critique of connectionism which has received wide acceptance, that of Fodor & Pylyshyn (1988), in fact, is based on similar arguments. We will return later to this topic (see section 5). For the moment, the fact that the behaviour of connectionist networks does not depend on their interpretation leads us to understand why these systems cannot *be constructed* as models but are only *used* as such.

If this is the case, then - as mentioned above - the analogy must come from an interpretation of the system. The problem, however, is that an interpretation of what a connectionist system does (or, if possible, of what it represents) is not straightforward. To ascribe functions or representative structures to such a system is not a simple matter. In particular, in order to be able to use a network as a *psychological* model, we should be able to give it a *psychological* interpretation. In this case, the object or event to be modelled is something that happens in the subject and the interpretation we have to give the system should consist of ascribing psychological functions or contents to it.

Usually this is done by using psychological labels to give an overall description or interpretation of the system states. The network builder or user (called *observer* in fig. 2) can interpret the states of the network by means of a symbolic representation. The system, then, exactly as the events found in the world, can be symbolically described (call this symbolization OSR/CS). So the observer can "read" words, letters, phonemic features, sides of a cube, or anything he likes into the system, because he knows the meaning of the input (via the OSR/W) and then he can label consistently what the network does or represents.

We can say, then, that the observer recognizes the analogy by interpreting or labeling *something* in the state of system units as analogous to psychological events or objects. But what is this *something*? Or, to see it from the other side of the coin, what kind of psychological processes can share some analogy with a connectionist network? These are the crucial problems.

In fact, usually this *something* turns out to be the state of some particular units in the connectionist system, the "output" units, from which we read the behaviour of the system. This because we are not able to give "hidden" units any meaning (or, at least, any meaning relevant to the process performed by the system and which can be seen as corresponding to some psychological process).

Thus, in choosing the appropriate labels, the observer is substantially constrained by the recognition (made a posteriori) of some analogies between the system performance and, to be exact, the *behaviour* of a subject which performes psychological processes (for example, recognition, learning, conceptualization, etc.). But, if this is true and we are to recognize only behavioural analogies, this cannot be sufficient to give the system the character of a model since analogies in performance do not necessarily imply analogies in structure or function. Indeed, this labeling has no value as psychological modelling since it is still attributing psychological features to the final product (the performance of the system), not to the process (the "development" of the system) (Greco, 1989).

3. Non-symbolic representations

In actual fact, there is a way to give hidden units a psychologically acceptable meaning: that is to consider the states of these units as *representations*.

But there are some problems. Fodor & Pylyshyn (1988) have pointed out that connectionist systems *cannot* exhibit such analogies, for the simple reason that they do not work by manipulating symbolic representations. This claim is plainly grounded on the presupposition that the only kind of mental representation is *symbolic*. But this presupposition is now challenged by numerous psychologists and philosophers, also outside the connectionist camp (for example: Dennett, 1983; Millikan, 1984; Hatfield & Kosslyn, 1984; Hatfield, 1989).

The concept of representation without symbols certainly is not new. Without going too far, any analogic device, like a clock or a voltmeter, can "represent", in a sense, some physical phenomenon (time, electric potential,etc.) without using symbols. Perhaps likewise affective-emotional states can reflect physiological events in a systematic and ordered fashion without symbols (as is evident from the difficulty of expressing them in linguistic form).

Although this concept is not new, the connectionist way to representation without symbols seems to have a special interest from the psychological point of view. The reason will be clearer below (see section 5).

However, may be the use of the term "representation" is not of much help in order to clarify this issue because it is not a fully understood concept.[3] In fact, a definition of this concept is not that simple, and it would require much more space than available here. But since for the purposes of the present discussion I have to assume one definition, provisionally I will propose the following: *representation is any state of a system which is internal to it (that is, not observable), but is in a functional and systematic correspondence with some state of the environment or with some other state of the system itself.* May be more is needed, but this should be the least. Anyhow, this definition is sufficiently wide to allow the possibility of representing without symbols.

4. The concept of *differentiation*

It is difficult to say that representing - in this sense, which is limited to a simple correspondence between a system and the world - is sufficient to fully model and explain complex psychological processes such as conceptualization or perception. For example, it is hard to say that a network which discriminates, say, dogs from nondogs has a category or a concept of "dog". Having a concept is not only a matter of reacting differently to certain objects: otherwise, we should easily admit that a key which has a different behaviour towards keyholes and non-keyholes has a concept of "keyhole". Simple discrimination, simple correspondence is not enough. What is needed is to know something about this discrimination, that is being able to represent representation. In other words, to have "self-awareness".[4]

At this point it seems necessary to examine more deeply how the natural symbolization (SSR) works. We have seen that current psychological theories appeal to mental representations and to symbol processing. But how do these symbols arise? How are they constructed? These questions are simply not posed by such theories.

In fact, in the classical perspective, some object *is* symbolic if it refers to some other object different from itself or, anyway, if it fulfils whatever condition is required to be a symbol. If this is the case, then we have a symbol; otherwise, we do not have a symbol. A third possibility is not given.

We must acknowledge that in current psychology, and more in current cognitive science, the fact that *human ideas develop in time* is still neglected. Nevertheless, there are clues which make us suspect that consciousness is not "all or none" but there are qualitatively different degrees in it. Much of the debate on perception between the "ecological" and the "constructive" perspectives (Gibson, 1979; Fodor & Pylyshyn, 1981; Turvey et al. 1981) focuses on the problem of whether perception is or is not mediated by symbols; however, neither of the two camps has considered symbols as not

[3] Even in broad discussions about the problem of representing, little attempt is made to define the concept of representation, and this is puzzling; in fact, everyone behaves as if the term were commonly understood clearly, or perhaps following their own understanding of it.

[4] Here one could go into the question of whether it is more appropriate to speak of "intentionality" or of "consciousness" instead of "self-awareness", but I don't want to follow this path now. Labels are not important, the important thing here is that when speaking of non-symbolic representation the problem of the transcendence of this representation comes to be posed almost automatically.

given but as the result of a process of symbolization. To my knowledge, only few people have focused on this problem; among them, Werner and Kaplan (1963), although not necessarily in terms to be shared completely.

This process of construction of symbols from non-symbols may be called "differentiation", which is a term coined by some Gestalt psychologists (in particular by Heinz Werner, 1957, even if he was accounting only for perceptual phenomena). The concepts of differentiation and of "microgenesis" (or *actual genesis*) can be usefully extended to account for the development of ideas. In particular, using these concepts one could describe how intellectual activity gradually develops from early stages to final shapes, through a differentiation. What happens to mental states through this differentiation is similar to what happens to an embryo, in which there are at first undifferentiated cells and subsequently different tissues develop from these (connective, nervous, etc.): some elements, which initially were all alike, come to specialize themselves as to their function.

Differentiation, then, consists of giving some interpretation to representative states of the system at their initial stage, when they are not yet expressible although they are causally and functionally related to what happens in the world. In this process, symbols are used to give these states a particular set-up. The result of this process might well be conceived as directed firstly to themselves, characterizing what Vygotsky (1962) called "inner speech". And later more differentiated ideas might reach a sufficient explicitness to be expressed in a more structured language, suited for communication with other people.

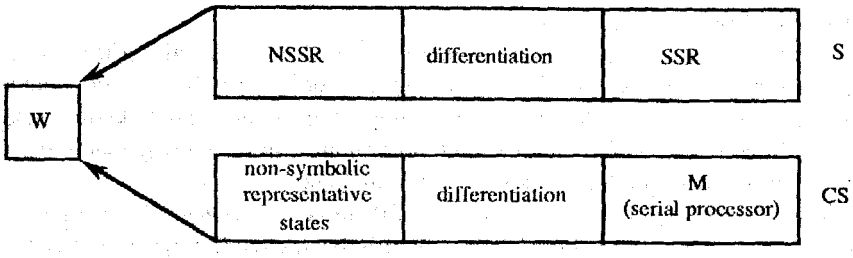
One particular symbolization is simply one of the many possible "readings" of the pre-symbolic state; different symbols could be used to express the same state. (This can be easily conceived in terms very similar to what happens in some perceptual situations such as in the celebrated Necker's cube: same state, different description). May be the same thing happens at the microscopic level as happens at a macroscopic level with distributed knowledge of hypertexts: different ways may be chosen to build a particular text, a particular formulation of knowledge.

5. A possible connectionist implementation of symbolic systems

Now we can see why connectionist claims about representation without symbols seem to have a special interest from the psychological point of view. What gives them this special interest is the fact that from the connectionist perspective it is possible to *analyze separately representation and interpretation*. And this is possible for the very reason that, as we have seen, the way representation is accomplished is not affected by its actual interpretation or meaning. So the fact, mentioned above, that the system states are the same independently of the meaning we are to give them, which seemed a limit, now may in fact become an advantage.

Why should we be interested in a representation disjoined from interpretation? The answer is: because considering non-symbolic representations can help us study (and perhaps simulate) how some object *becomes* symbolic, how symbols are grounded in representational states, and how they emerge differentiating themselves from these states.

Figure 3



W, S, CS, SSR, OSR: see previous figures.
NSSR = *not necessarily symbolic* representations; M = module

Let's consider how this can be realized using a connectionist system (figure 3). The interpretation of the CS could be done having some part of the system which can recognize this discrimination and can use symbols to label it consistently (exactly as the network builder or user does when he himself discriminates). In this case we could say that it would have modelled a psychological process, namely the process of differentiation of symbols from pre-symbolic representative states.

A connectionist model, then, could be implemented having a part or a module (M) of the system, a sort of "self-awareness" module. The job of this module would be to achieve an interpretation of what the main system (CS) carries out. If this interpretation worked like the symbolization SSR, then we would consider it as a true model of it.

One could conceive this module as a different connectionist network, whose input would be the states of the main network, and which achieves a representation of it (or, as we could say, a meta-representation of the whole system) on the basis of a functional correspondence (for example, detecting regularities, etc.).

But things are not so simple. This module, indeed, would still not accomplish a symbolization and the states of this new network would need a further interpretation, exactly like the states of the main network. This could not be done to infinity. At some point one should stop and appeal to a symbolic device in order to have a definite (stable, and consistent) meaning. This process of drawing out symbols from non-symbolic states is just what above has been called "differentiation".

This process, in fact, is a true development, and it cannot be a simple product of random interaction. There is a true *organization* here, and there could not be an organization without a plan, without rules, without (why not?) a program. Here there is a parallel aspect and at the same time a serial aspect. From the parallel side, in different moments in time you can find different "states" of the system; and you may well conceive them as resulting from a parallel interaction among elementary units. From the serial side, you are also faced with the system's behaviour in a period of time, which can be called a "process".

In this view, to build knowledge from the mingling of experiences, symbolic processes put some order, serially constructing different conscious states, identifying one

thing at a time even when everything seems to be connected with everything. The result of this process is true information, in the original sense of structure emerging from the indistinct, of order versus entropy.

Now we can realize that this module should have a serial nature, since symbol construction can only be done in successive times (necessarily one symbol comes *before*, others follow and so on). A natural way to implement it, then, could be to conceive of it as a traditional computer processing unit, programmed in such a way that it examines the states of a network and draws out symbols from them.

However, many problems remain. One of them is the problem of writing the program for this unit: who could do it? how can this be realized? One of the reasons why connectionist systems are so attractive is just the fact that they need not be programmed. In principle, the same job could be done using different layers of networks, hierarchically arranged in such a way that some of them "control" others. But, as I have said, how a definite interpretation could come out from such systems is difficult to see.

In any case, how this module is physically implemented is not so important. More importantly, the fact of drawing out certain symbols does not end the story. This very event becomes part of the actual world (or situation) in which the system lies. The module M, whether implemented as a network or as a program, must be therefore a part of the whole system, not outside it, and must interact with the main part of the system. Indeed, I think that the most interesting psychological phenomena arise from this interaction and that how this interaction works might be a subject for future investigations.

References

Dennett D.C. (1983). Styles of mental representation. *Proceedings of the Aristotelian Society*, 83.

Fodor J.A., Pylyshyn Z.W. (1981). How direct is visual perception?: Some reflections on Gibson's "Ecological approach". *Cognition*, 9, 139-196.

Fodor J.A., Pylyshyn Z.W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28, 3-71.

Gibson J. (1979). *The ecological approach to visual perception.*

Greco A. (1989). Some remarks about connectionism in psychological simulation. *Cognitive Systems*, 2-4, 359-372.

Hatfield G. (1989). Computation, representation, and content in noncognitive theories of perception. In: Silvers S. (Ed.), *Rerepresentation.* Kluwer, Dordrecht.

Hatfield G., Kosslyn S. (1984). Representation without symbol systems. *Social Research*, 51.

Millikan R.G. (1984). Language, thought, and other biological categories. MIT Press, Cambridge.

Turvey M.T., Shaw R.E., Reed E.S., Mace W.M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn. *Cognition*, 9.

Vygotsky L.S. (1962). *Thought and language.* MIT Press, Wiley, New York.

Werner H. (1957). The concept of development from a comparative and organismic view. In: Harris D.B. (Ed.) *The concept of development*. University of Minnesota Press, Minneapolis.

Werner H., Kaplan B. (1963). *Symbol formation*. Wiley, New York.