Gerald Edelman: *Selection* versus *Instruction*

by *Astro Calisi*

Neuroscientist Gerald Edelman proposed a theory of mind – *the Theory of Selection of Neuronal Groups* (TSGN) – which assigns a primary role to the concept of *selection*. According to this theory, our brain would present at birth an overabundance of nerve connections, arranged more or less randomly. As a result of the activity of the subject, conducted in a given environment, the most stressed connections would be strengthened, while the unused ones would tend to atrophy until they disappear completely. In this way the specific attitudes and capacities of each individual would develop.

Edelman often points out the radical difference of the theory he proposed with respect to the different conceptions that try to explain mind by comparing it to a sophisticated computer. The operation of computers is in fact based on *instructions*, that is, on predefined algorithms that guide, step by step, the operation of the system. The TSGN is based instead on the *selection* of the models of representation corresponding to the behaviors that are successful.

Edelman has tried to demonstrate his thesis experimentally, building artificial systems capable of performing elementary tasks in circumscribed environments. The instructions initially given to these systems were quite generic and mainly concerned the types of stimuli to be considered significant. The specific behavioral strategies to adequately perform the assigned tasks were the result of a categorization developed progressively by the system through a trial and error process.

Edelman has attracted a lot of criticism because, in order to prove the validity of his theory, he resorted to models based, ultimately, on the very computation he intended to question. In fact, even the neural networks, which he used extensively in the realization of his experimental prototypes, are entirely attributable to ordinary computation. So much so that they are usually simulated through normal computers.

I believe, however, that the strongest objection that can be addressed to Edelman's experiments is another one. He wanted to demonstrate the fundamental difference in the way the brain operates with respect to traditional computational models, asserting the distinction, fundamental to him, between *selection* and *instruction*. However, this distinction only makes sense when we want to study the *genesis* of certain characteristics or properties of a system, that is, when we want to reconstruct the processes that have led the system to assume a certain internal conformation. It does not tell us anything, however, about the relationship that exists between this conformation and the behavioral responses exhibited by the system, that is, it does not tell us anything that helps us to understand the nature of this relationship.

If we consider the sophisticated automata conceived by Edelman when, through adequate interaction with the environment, they have developed cognitive and behavioral models that allow them to perform their assigned tasks satisfactorily, we must agree that they act exactly like their computational cousins, governed entirely by syntactical instructions. What do we find, in fact, by analyzing Edelman's automata at a given moment?

We find an existing hardware apparatus and instructions in *distributed form* in the different nodes that make up the neural networks. An experienced technician would be able, after a careful examination of the hardware parts and the state of each node, to reconstruct an exact copy of such an automaton, which could only behave in the same way as the original prototype. The second automaton, however, would have acquired its capabilities not through a selection process, but thanks to *instructions* inserted inside it.

What I am trying to highlight with this sort of imaginary experiment is that the distinction between *selection* and *instruction* is irrelevant for an explanation of the relationship between the internal organization of a system and the skills it exhibits. How irrelevant is to emphasize the importance of the environment when we have to the relationship that certain mental properties have with the nervous base of the brain. If we assume that the different manifestations of mind are merely a product of the electrochemical activity of the brain, calling into question the influence of the environment in the development of a given nervous organization does not contribute to the understanding of how the organization itself is able to give rise to phenomena such as conscious experiences, autonomy of will or creative abilities. It is clear that the specific *form* in which these phenomena present themselves in a given individual derives, more or less closely, from the characteristics of past interactions with the environment, but the interactions themselves do not explain how such phenomena *are possible*: in particular, they do not help us to understand the nature of their relationship with the material substratum. If we want to be rigorous and consequent to the end, remaining within current conceptions, we are in fact forced to recognize that the behaviors and properties of a given system (biological or artificial) *depend exclusively on its current organization*, i.e. the logic of functioning that distinguishes it, while the processes (or, if we want, the "historical vicissitudes") that have led the system to assume that particular conformation are completely irrelevant.

Faced with such attempts, the crucial question is: does the system's behavior depend or not on how it is organized at a given moment?

If the answer is affirmative, then, in the explanation of the causal relationship between the organization and the properties exhibited by it, we can completely ignore its history, i.e. the processes that have led the system to assume its current structure; if the answer is negative, then someone must explain to us what that *quid* is, that lies beyond the material organizzation, which can make a difference.

I hope, with this, that I have sufficiently clarified the meaning of my distinction between *ways of development* of certain characteristics of mind and *explanation* of the same characteristics with reference to the material substratum to which they are associated. I take this opportunity, indeed, to observe that - in my opinion - the claim of certain authors not to completely renounce to a historical approach in the construction of explanatory models of mind, while declaring their full adherence to scientific naturalism, is a clear indication of the "bad conscience" that lies at the basis of their proposals. Historical consideration in attempts to offer an explanation of the mind-brain relationship is a "historical remnant" completely incompatible with a materialistic conception of the mind, as is the case with the current models of scientific explanation.

[Astro Calisi, *Oltre gli orizzonti del conosciuto*..., pagg. 179-182 - English translation by the author]