REVISTA MEXICANA DE ANÁLISIS DE LA CONDUCTA 26, 129-134 NÚMERO 2 (SEPTIEMBRE)
MEXICAN JOURNAL OF BEHAVIOR ANALYSIS NUMBER 2 (SEPTEMBER)

EDITORIAL

NEURAL NETWORKS IN BEHAVIOR ANALYSIS: MODELS, RESULTS, AND ISSUES

JOSÉ E. BURGOS
UNIVERSITY OF GUADALAJARA

On July 25, 1989, the Senate and the House of Representatives of the United States of America approved a resolution that officially designated the 90s as the "Decade of the Brain". The main purpose of this resolution was to stimulate basic and applied neuroscientific research that eventually resulted in treatments to the various neurological disorders known to date. The extent to which such purpose has been achieved remains to be determined. But the resolution certainly stimulated an unprecedented amount of research on the structure and function of brains. This kind of research, of course, had been occurring well before the resolution in question. However, the number of techniques, methods, specialized journals, and publications in the neurosciences increased dramatically during the last 10 years. The resulting database is colossal, to say the least, and it keeps growing by the minute, not only in the United States of America, but also throughout the world. The spirit of the resolution thus has transcended its immediate cultural and political boundaries, to become a collective human effort towards an understanding of the brain.

The vast majority of neuroscientific knowledge has involved the application of a wide variety of methods and techniques to relatively simple laboratory preparations using nonhuman brains, vertebrate as well as invertebrate, from the molecular to the anatomical levels of organization. A lesson we have learned from this research is that brains are extraordinarily complex, far more than anyone expected. This lesson has become so patent that it could make us believe that brains are inherently complex, that complexity is a fundamental property of brains. Nonetheless, some efforts have been guided by a rejection of this idea, that is, by the notion that brain complexity can be viewed as emerging from multiple realizations of a few general principles that are sufficiently simple as to allow for the construction of *mathematical*

models. Although these efforts constitute a minority within the collective study of the brain, they have been substantial enough to be noticed.

They are known by several names, such as 'neurocomputing', 'computational neuroscience', 'connectionism', and 'neural-network modelling'. Like most labels, those ones admit different usages, so there is no universal agreement on their meanings. In certain uses, some of those labels roughly reflect a diversity of efforts that can be classified according to a number of different criteria. Under what seems to be the most used criterion, neural-network modeling efforts differ in how much they attempt to reflect or capture hard experimental data on specific parts of particular nervous systems. The term 'computational-neuroscience' is some times used to identify this kind of effort, which suggests that it could be viewed as a subfield within the neurosciences. Other efforts are more loosely guided (to different degrees) by those data, and usually come from computer/information science and psychology. But, again, the disciplinary boundaries among the different kinds of efforts are rather fuzzy.

The scientific interest for the brain thus has transcended the disciplinary boundaries of the neurosciences. Nevertheless, that has relaxed the criteria for evaluating neural-network models. Modeling work outside the neurosciences is usually regarded as being based on a philosophical position known as 'connectionism'. According to this position, 'intelligence', 'cognition', and 'minds' (however we choose to define these terms behaviorally) are to be understood as *emergent* (rather than fundamental) properties of brains. After all, whatever behavioral phenomena that are referred by those terms must relate in critical ways to the structure and function of nervous systems, although the exact nature of the relation remains elusive.

However, connectionists insist, we need not capture the biological (molecular, cellular, anatomical) details of brains in order to understand those phenomena. We just need to focus on the most general defining features of brains, namely nonsymbolic distributed representations and parallel processing. These features are in contrast with the centralized/symbolic kind of representation and the sequential kind of processing of Turing machines (the theoretical descriptions underlying today's personal computers). Connectionists thus tend to reject the Turing/von Neumann computer metaphor about the brain.

Computational neuroscientists, of course, will agree that intelligence, cognition, and minds should be viewed as emergent properties, and that nonsymbolic/distributed representations and parallel processing are indeed features of the brain. They will also reject the Turing/von Neurmann-computer metaphor about the brain. However, computational neuroscientists contend that those features are far too general to be related in specific ways to specific brains, for they are shared by virtually *all* brains. We need, computational

neuroscientists argue, to model the *biological details* of the structure and function of real brains. The construction of neural-network models thus must be tightly restricted by hard experimental evidence on real brains, at the molecular, cellular, and anatomical levels. Connectionists disagree, arguing that such models can (and should) be validated behaviorally, that is, on the basis of how well they simulate certain performances of interest.

The above distinction between computational neuroscientists and connectionists, of course, is largely idealized and crisp. In reality, these categories represent extremes of a continuum. To be sure, such extremes are nonempty. However, efforts in between them are also very real. In any case, the controversy on how neural-network models should be evaluated remains unresolved. Nonetheless, there is a general agreement on the idea that the structural and functional complexity of real brains is the result of many particular realizations and implementations of the same set of relatively simple principles. Whatever complex functions we wish to attribute to the brain (call them 'intelligence', 'cognitition', or 'behavior') are to be considered as emerging from collections of elements that are structurally and functionally far simpler than the brains they constitute. This crucial idea has been at the basis of neural-network modeling since its beginnings, in the seminal works by McCulloch and Pitts (1943) and Rashevsky (1938). It is often expressed by saying that this kind of modeling follows a *bottom-up* strategy. According to this strategy, simulations of behavioral phenomena of interest must emerge from systems whose constituting elements function like neurons. Again, how closely such components should emulate real neurons, or how closely such systems should emulate the anatomy of real neuronal circuits is still (and, I suspect, will remain for a long time) in discussion. But the notion of a bottom-up methodology represents a unifying theme that runs across neural-network modeling efforts, setting them apart from other efforts to understand behavior, like those found in nonconnectionist artificial intelligence and traditional cognitive psychology.

The "Decade of the Brain" is now over. However, it originated a momentum of research that, most likely, will remain for many more years to come, at least throughout the next decade. Neural-network modelling (within as well as without the neurosciences) shares part of this momentum. It still remains a minority effort in the neurosciences and psychology, although it is becoming increasingly popular in computer/information science. Nowadays, with the beginning of a new decade, this kind of effort is recognized as a firmly established and legitimate field of research.

In 1998, the American Psychological Association has proposed to designate this new decade as the "Decade of Behavior", a proposal whose resolution, as far as I know, awaits for official approval. If it is approved, it is

expected that it will encourage a great deal of research in behavioral science, at least as much as the amount of research that was encouraged in the neurosciences by the "Decade-of-the-Brain" resolution. That proposal, in combination with the momentum left in neural-network modeling by this resolution, provides us with an ideal *zeitgeist* for a more systematic and complete incorporation of behavioral principles into neural-network modeling. After all, behavior plays a central role in a phenomenalistic validation of neural-network models, for computational neuroscientists and connectionists alike. Again, researchers remain in disagreement on the extent to which behavioral validation is sufficient. But I believe that most would agree that such validation is at least necessary.

Acknowledging behavioral validation as a necessary criterion for evaluating neural-network models raises the issue of exactly what kinds of behavioral phenomena should be taken as empirical referents to apply such criterion. Not surprisingly, researchers also differ on this point. The range of behavioral phenomena of interest is so vast (from the simplest invertebrate behavior to the most complex human behavior) that it is futile to attempt to identify (much less impose) a single subject matter, or to dedicate a significant amount of time to the study of a significant portion of that range. And said range is expected to become even vaster, especially if this new decade is officially designated as the "Decade of Behavior".

So the best we can do is to be (very) selective. Several guidelines for selecting phenomena of interest are possible. According to one guideline, we should use *principles*, rather than phenomena, as validation criteria. Principles constitute powerful selection guidelines, for they are usually based on a few paradigmatic phenomena. Also, a principle is assumed to be valid for most (if not all) of those instances of such paradigmatic phenomena that have not been directly studied in the laboratory. On this basis, we can concentrate on the kinds of phenomena that are studied in the experimental analysis of behavior (broadly conceived, to include operant- as well as Pavlovian-conditioning research), as a discipline whose main objective has been and still is precisely to derive behavioral principles in the laboratory.

In view of all of the above, it is no coincidence that this Special Issue of the Year 2000 of the Mexican Journal of Behavior Analysis is on neural-network modeling. A feature that sets these efforts apart from others, however, is that they are being made by researchers that have been trained and work in a behavior-analytic tradition. This feature is remarkable, if we consider that behavior analysts have traditionally focused on the study of behavior as a subject matter in its own right, rejecting any description or explanation that appeals to nonbehavioral levels of analysis. Nonetheless, neural-network research within behavior analysis has become an undeniable reality that may

very well change the conceptual, theoretical, and methodological landscape of our discipline. The present Issue is a testimony to that.

The Issue contains seven contributions. In the first contribution, Potter and Wilson provide us with an introduction to neural-network modelling in behavior analysis, by reviewing a number of models. They also discuss specific ways in which behavior analysis and neural-network modeling can benefit from each other. In the second contribution, I offer an account of superstition, based on simulations of this phenomenon using selection neural networks. I also discuss some philosophical issues that underlie traditional criticisms towards neural-network modelling in behavior analysis. In the third contribution, Delepoulle, Preux, and Darcheville describe simulations of cooperation in a minimal social situation by artificial agents whose functioning is described by different reinforcement-learning models. In the fourth contribution, Jozefowiez, Darcheville, and Preux describe related work on an operant approach to the iterated Prisoners' Dilemma. Their simulations show the emergence of cooperation after indirect reinforcement of noncooperative behavior in artificial learning agents. In the fifth contribution, Kemp and Eckerman, guided by research on the effects of dopamine on hippocampal pyramidal cells, present simulation data of behavioral patterning in neural networks that do not receive antecedent exteroceptive stimulation. In the sixth contribution, Matt Morris describes simulations of various operant-conditioning phenomena, such as acquisition, extinction, reacquisition; variable-ratio, and variable-interval performance, using his "Artie" model. Finally, in the seventh contribution, Jackson Marr provides us with critical reflections on the nature of neural-network models, what they can offer to behavior analysis, and under what conditions.

I want to express my gratitude to Carlos Bruner for giving me this unique opportunity as a guest editor for this Issue, and to Laura Acuña for her invaluable assistance throughout the process of putting it together. I also want to extend my deepest appreciation to the other six authors for having accepted my invitation to participate in this collective effort, and for doing such a fine job in their respective contributions. Their patience and consideration certainly made my work as editor a true pleasure. Much work remains to be done. I hope this Issue serves at least to sensitize behavior analysts to the possibilities that neural-network models offer us for a more precise and comprehensive scientific understanding of behavior. Of course, if this Issue inspires more behavior analysts to actually do neural-network research, so much the better.

REFERENCES

McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115-133.

Rashevsky, N. (1938). *Mathematical biophysics: Physico-mathematical foundations of biology*. University of Chicago Press.