Review of

Neural Organization

Arbib. M.A., Érdi, P. and Szentágothai, J., MIT Press, Cambridge, Mass., 1998., pp. xii+407.

George Kampis

Department of History and Philosophy of Science

Roland Eötvös University, Budapest

How does the brain work? Indeed, what is the brain? Ostensively the brain can be defined as that greyish stuff residing in the scull; from the point of view of science it is an agglomerate of anatomical, physiological and behavioral findings. The anatomy of the brain includes knowledge about nerve cells and their different types, together with an architecture of the intricate and as yet mostly unknown wiring that connects them. From the physiological point of view, the brain is, above all, a complex of electric and chemical phenomena, combined at the surface of cell membranes and synapses; functionally, it is a conglomerate of more or less localized units for performing various subtasks of sensory and motor processing. Behaviorally, the study of the brain involves an explanation of the sensory and autonomous guidance of locomotion and other low-level behaviors as well as an attempt to account for what can be called mental events. It is presently not clear how aspects of the mind can be included in the proper study of the brain, or whether, from the often eliminative point of view of science, there can be a mind at all; the so-called identity thesis, which claims that all mental events are ultimately brain events, is but an expression of a humble desire for a respectable research program - *any* research program that attempts to build a bridge.

The authors of the monumental book under review have exactly the just outlined ambitious breadth of topics, that range from neuroanatomy to the organization of higher brain functions (or "mind"), in their view. They are builders of the bridge.

One could have expressed *prima facie* doubt that such a book can be written at all. The chief question of every such enterprise is, whether the various parts connect. Take our case. The late Szentágothai, an anatomist, discusses diagrams of the cortex and experiments on implanted limbs of frogs; Arbib, who in the book plays a role which I suspect many prospective readers think should have been played by a physiologist, reaches from Immanuel Kant to the logic of feedback circuits; Érdi, a dynamical modeler, writes about themes of nonlinear maps and self-organization; together they hope to get to the point where a first picture of the mind emerges. This latter needs some additional explanation, as this is clearly not a book on cognitive science, despite the sporadic references to Rumelhardt, connectionism, the computational hypothesis of the mind, Johnson-Laird, or even autopoiesis, that is, items normally not found in "brain books". Yet a desire for an over-encompassing perspective up to the mind is clearly there. They write (on

p. 3.): "the true fascination of the brain is the many functions that it serves, perception and the control of action, memory and thought".

How do they fare? There are so many books about the brain, and most of them are not nearly as ambitious as this one. So, here is the question again: Is it possible to see this book as a coherent effort towards something significant?

My answer is strongly affirmative. It is true that no one with even the least knowledge about the matter may believe that it should be possible, at present, to write a book that "solves" the problem of the brain, especially when so many of even the most elementary facts are missing (think of the recent debate and the new results concerning the growth of new nerve cells in adult humans). So, what we can expect as a maximum achievement is a cautious and progressive approximation of various fields which when studied separately will probably never provide a full picture of the brain. And this is exactly what N*eural Organization* is about.

The authors argue that a multidisciplinary, integrative approach that involves several perspectives in dialogue is necessary to make breakthrough in the study of the brain. How often have we heard these words; but in most of the cases, they meant nothing, or at best very little. But now here we have the result of a sincere attempt in our hand. In judging this effort, I think the correct attitude is (if I may speak figuratively) not a top-down, but a bottom-up one: the effort is not to be judged from the imaginary perspective of where the authors should have arrived but from the real perspective of where they started out. And the readily available platform from which an integrative approach to the brain (or anything) can be launched is very frustrating indeed. Somehow the times do not favor truly unified or integrated science. There is inter- (cross-, trans-, or whatever) disciplinarity instead. A typical example is evolution theory. Perhaps nothing is as influential today as the idea of evolution, which after a long pause invades sociology, psychology, and even philosophy again, and rightly so. However, instead of a mutual enriching of perspectives the new aproaches typically deal with caricatures of each other's fields; the result is that in, say, evolutionary psychology the only aspect of evolution worth mentioning is that of natural selection and the resulting differential survival - if that was so simple, evolution theorists in biology would be without jobs. Want it or not, the parallel fact is that anatomists and physiologists who deal with the "hard facts" of the brain seldom talk to the "purely theoretical" system theorists or dynamical modelers; perhaps because of the same caricature effect. Typically, the latter's people's work is developed in isolation and under mutual suspicion, a situation which only changed a little since the relatively recent establishment of computational neuroscience as a respectable scientific field.

Let me bring up the topic of Artificial Life as another, not-so-distant illustration for a field where unnecessary isolation causes severe delays in development. The result is that many ideas that earlier theoretical biology and biological cybernetics have developed must be, to say the least, reinvented from scratch in lack of communication between the fields. A concrete example is the bunch of locomotion control problems of situated or autonomous robotics. Even a diagonal study of the literature shows that researchers today not even secretly read the old cybernetics folks who offered solutions, at least some kind of solutions, to all these problems decades ago, when studying the neuronal logic of insect motor control. It is difficult to see why these traditions, which although never made it to mainstream science were very fruitful in their own days, should be so much despised from another science's present perspective. McCulloch and Pitts may get their due citations for neural networks, but few people remember McCulloch as a pioneer, or indeed a visionary of mathematical brain research, something that does not really exist even today, and which both Artificial Life and laboratory neuroscience could directly benefit from. If there is a message of *Neural Organization*, it is the need for substantial co-operation of this kind of which it is a conscious example.

To return to the case of the brain: from the point of view of philosophy of science it is clear that it's impossible to make any *experiment* without a previous *theoretical* commitment; in this sense, the "mere" theoretician is in no worse situation than the "mere" experimentalist. They in fact need each other. This sounds rather trivial: is that all?, you may ask. Well, that is actually not all, but that would already be sufficient for rehabilitating "mere" theoretical work. To show how theoretical brain research can in fact help traditional approaches is one of the main concern of the authors of Neural Organization.

Let me illustrate this, by analogy, on the problem of the identification of unknown electronic devices, an operation popularized by the philosopher D.C. Dennett under the heading "reverse engineering". An electronic device is a fairly dumb piece of matter until we have a strong hypothesis about what it is *for*. Based on this hypothesis, or theory, of the device, we can make predictions of what *should* happen in the device if we do this-and-this, or alternatively, we can tell what potential data about the construction of the device are missing in order to calculate a prediction. Now of course there can be various levels on which to formulate our hypothesis: some people would prefer electronic circuits, others logical gates, and still others formal automata theory as the final units of their analysis, but perhaps nobody would opt for charge topologies of the spatial region in which the device is located. Anyway, the interesting point is that the levels are intertwined, and that is why none of them can be successfully studied without the others. The purpose, or *meaning* of the device, be it an electronic one or the brain, can only be inferred using several levels together.

Take a concrete electronic device, the familiar computer to see how this works. It will be perhaps agreed upon that the most obvious levels of analysis are these: the hardware, the operating system, and the programs or applications that run on the machine. These levels correspond to the *anatomy*, the *functional description*, and the *behavioral analysis* of the computer, respectively.

The first few steps are obvious, and so we realize that the computer works by means of burst electric signals that are directed along a variety of pathways. However, it is utterly impossible to get much further than this, without invoking the hypothesis that the machine is actually a computer, on which an operating system is found; the seemingly random switching behavior of the circuits makes no sense by itself. On the other hand, with the aid of the operating system postulate we will discover that there are global rules to determine what happens after what; in other words, by this hypothesis we get instructions like (say) "after having found this impulse from the CPU to the RAM look for another one that goes from the RAM to the hard disk because this is what any operating system should do next". Such instructions may relate distant regions and provoke us to perform complex measurements which would never be performed otherwise. So perhaps a book on computer identification cannot be written without including chapters on operating systems and, for that matter, on typical applications that actualize the behavioral patterns made possible by the operating system. One of the important benefits of *Neural Organization* is to make it clear that the brain is no different. So any future brain book should discuss function and behavior besides (and in unity with) anatomy.

At this point the reader will notice: the brain has a physiology as well. What happened to the physiology of the brain? Under the present perspective it becomes part of the anatomy, which invalidates another old distinction. Similarly, the anatomy of computers includes the hardware parts *and* their properties. To conclude our engineering detour: *Neural Organization* adopts the same perspective for the brain as we discussed for the computer (although perhaps less clearly; some parts of physiology are subsumed under "the empirical basis of dynamic models"). That is, I think, why we find no separate place for physiology in the book, and why the mediator between the anatomist and the dynamical modeler is a system theorist of the brain.

And now we come to the authors and the topical organization of the book. Szentágothai was a legend of brain research, of Eccles-Ito-Szentágothai fame (*The cerebellum as a neuronal machine,* Springer, 1967) and many other fames. His complex interests in the brain resulted in a co-operation with theoretical chemist and dynamical theorist Érdi and with Arbib, who by the start of the joint work on this book has had a carreer in mathematics and another one in the systems science of brain modeling; Szentágothai's personality framed the entire project. It is a tragic expression of this fact that Professor Szentágothai died one morning while working on the last touches of the book. The underlying identity that connected these people in their writing of the book is a concern with determinism and randomness, most explicitly expressed in the developmental problems of the brain. During his long life Szentágothai was among the very first, back in the good old Waddingtonian days of morphogenesis in the 50's, to stress the impossibility of coding the brain determininstically by means of positional information of the neurones,via a genetic machinery. It was clear from the outset that the brain cannot be completely random either; he supported the concept of self-organization that can aviod both extremes by providing a deterministic framework for the origin of well-defined dynamic structures which take final shape in the presence of noise. The book contains nicely written historical parts that describe the early struggles and the attempted solutions to these developmental jigsaw puzzles, as well as a technical discussion of more recent anatomical findings and mathematical models (e.g. of the columnar architecture or of ocular dominance) that show the same ideas at work today.

The always present topic of determinism versus randomness, or of pre-fixed structure and dynamic behavior, is presented via a three-steps approach. The first step is that of anatomy, itself viewed through a developmental lens. One of the most important units of anatomy under discussion are the *columns* discovered in the sixties by Szentágothai; the columnar organization of the brain means that locally interconnected regions exist that contain frequent within-column and sparce between-column axonal connections. The second step involves what are called *schemas* for the basic units of functional treatment. A schema is understood as a logical decomposition of global behavior coordination (such as, for istance, of viewing and grasping, detour behavior, or predator avoidance) into functional modules that have, or can have, anatomical counterparts. The third item is dynamics in the sense of the *patterns* of rhythms, oscillatory and chaotic behaviors associated with both developmental and memory functions. The basic pattern-generating device is that of a differential equation based on physiological variables such as membrane potentials. Owing to the inevitable intermixing of the three steps in real phenomena, spatial and time scales play an important organizing role in separating the anatomical, functional, and dynamic apects and their respective approaches in complex neural phenomena such as plasticity.

Biological cybernetics gets its day in court in *Neural Organization* at long last. The schema theory of Arbib has strong ties to both the perceptual and motor control problems of classical cybernetics, in the sense of, let us say, the Tübingen school of Reichard, Braitenberg and Varjú, and the psychological concepts of Piaget and Bartlett on representation. Well before Fodor and the concept of modularity for mental processing Arbib has introduced schemas as composable programs that consist of modules. His schemas involve block diagram specifications of systems that as a rule may encompass several brain regions while based on modules which are local to one such region. A central notion is that of the difference between a schema as a composable program and its instance as an activated wiring circuit over the functional modules (often the computational vision reseracher D. Marr is quoted for a similar distinction, and there is a close connection between Arbib's earlier works and Marr's three level analysis into concepts, algorithms, and realizations, of which Arbib's just mentioned notions would cover the last two). Schema theory has provided numerous results on models of various aspects of perception and behavior control in insects or, as in Arbib's own case, frogs. It is interesting and motivating to separate these traditions in the book from what they have given rise to since their conception, namely, a new wave of experimental biophysics where control concepts are used naturally as part of the applied mathematical framework.

Dynamical systems are the most diffucult part of the book. Dynamical systems are notoriously difficult to present; to put it simply, either you are a mathematician or you are not, in which (i.e. the default) case you are lost. The book does a good deal of effort to find a compromise between technical details and too much generality that excludes a meaningful treatment. I find it especially useful that many classic concepts and results are presented for the nonmathematical reader in mind. Then there comes the inevitable quantum jump, in both Arbib and Érdi's text, to the technical words; nobody knows what to do about this old problem. However, dynamical systems theory is a natural player in brain theory because it offers a toolbox of methods for dealing with temporal phenomena of electrochemical brain states. Neurodynamics, the branch of dynamical systems expecially adapted to dealing with biological neural networks, is presented in the book as another rich field with an unduly supressed tradition that goes back to the sixties. Érdi himself has taken part of the development of this tradition since the seventies. Besides several successful models detailed in the case studies to be discussed below, his own contribution to the field involves the establishment of a clear and coherent hierarchical view of the various dynamical scales involved, ranging from single cell modells to group and compartmental models, up to statistical and populational models of many-neuron systems. The complexity of neurodynamical problems is truly captivating. From the temporal point of view, there are the time scales of the membrane processes, of synaptic modification, and of the self-organization of entire networks. On each level there can be fixed point, cyclic and, due to the nonlinearlity of the components, chaotic behaviors. The rapidly growing literature of this plethora is presented in the book from a unified perspective that encompasses even artificial neural networks (e.g. the Hopfield model) and the methodological problems of computing with attractors.

The book offers six consecutive case studies in order to present the interplay of anatomical, functional, and dynamic aspects. The six interacting systems of the brain discussed in detail are: the olfactory system, the hyppocampus, the thalamus, the cerebral cortex, the cerebellum and the basal ganglia. Together these discussions form almost exactly half of the book: about 200 pages of altogether 400-odd. Not always do the three approaches converge, unfortunately. In the spirit of my earlier remarks, this complaint goes to the state of the art of the whole field and not to the individual authors. My favorites and at the same time the most positive examples for co-operation among the authors are the discussions of the olfactory system, the hyppocampus and the ocular dominance parts of the cerebral coretex; of course I am not an expert on all matters covered in the book and these may be simply the ones that I personally understand best.

Be it as it may, the olfactory bulb is a fruitful theme due to the recent interest in the interplay between chaos theory, mental modeling, and computational neuroscience as raised in the works of Walter Freeman whose principal target of investigation is olfactory discrimination and memory. After presenting the anatomy of the olfactory system the authors turn to possible schemes for olfaction and a critical analyis of its dynamic behavior. The price for this completeness is that under schemes we should understand wiring diagrams and not schemas in the sense of algorithms that reach over various brain regions; as a reconciliation we get an almost complete and balanced treatment of the Freeman works and their criticisms, as well as serveral mathematical models for the 40 Hz-oscillations (an understandably fashionable topic that pops up at various parts of the book).

The hippocampal part is distinguished by the multitude of problems raised in this subfield since the seminal work of O'Keefe and Conway in 1980. The popular understanding since then is that the hippocampus is a cognitive map, i.e. an analogous allocentric representation of a navigated territory in Eucidean space. In short, the hyppocampus is a site for spatial memory which is spatial itself. Here the authors argue convincingly that the hippocampus does in fact not provide a cognitive map but is rather part of a larger system that may provide one; in passing, they offer critical discussions of many underlying assumptions of the received views on spatial memory such as the homogeneity of spatial representation and the role of food and other cues or markers in forming a representation; a review of the neurophysiology of representation and human spatial memory; a treatment of associative matrix models that occur in many further discussions (cf. e.g. the famous Kohonen-type self-organizing maps or models of the cerebellum).

The problem of ocular dominance is simply the question, why the innervation of the eyes is not homogeneus over the visual cortex; instead, there are alternating local structures corresponding to the left and the right eye, respectively. Since Hubel and Wiesel the role of development and external stimuli in this innervation is taken as granted; here it is shown by using neural competitive models of the Changeux-type how ocular dominance columns can emerge in a family of dynamic models which is compatible with anatomical data.

Besides its brave perspective and delicious details the book has a few less attractive features as well. There is a certain degree of redundancy in it, largely owing to the fact that the authors included three separate introductory chapters with the intent of making the reader familiar with the basics of each involved specialty. The result is less than optimal, some concepts are re-explained and even some figures are re-used later (e.g. figure 2.16. is identical with 8.16., and 2.14. with 9.2.). And despite the multiple introductions, many concepts remain unexplained or will be introduced and discussed later. This is perhaps inevitable in a book which is already *large* anyway, so adding more introduction would make it unreadable, but all the more, one gets the impression that a linear development of the text, or using the familiar "boxing" method would have been better than this particular simulation of hypertext technique. These and other minor faults (for instance, the occasional lack of clarity about what is important and what deserves only mention; too many details about matters of purely technical importance now and then; an incomplete subject index and a missing name index) do not detract significantly from the overall very positive impression, a judgment which is topped by the very high level physical appearance of the book.

In summary, I believe this book will turn out to be an important companion for several years for the various students of the brain, and will be read not only as a complex sourcebook of both anatomy and the various facets of modelling but, as I was trying to imply in the detailed appraisal, a path-breaking example of methodology for future studies to come.