Brain Complexity: Analysis, Models and Limits of Understanding

Andreas Schierwagen

Institute for Computer Science, Intelligent Systems Department
University of Leipzig
Leipzig, Germany
schierwa@informatik.uni-leipzig.de
http://www.informatik.uni-leipzig.de/~schierwa

Abstract. Manifold initiatives try to utilize the operational principles of organisms and brains to develop alternative, biologically inspired computing paradigms. This paper reviews key features of the standard method applied to complexity in the cognitive and brain sciences, i.e. decompositional analysis. Projects investigating the nature of computations by cortical columns are discussed which exemplify the application of this standard method. New findings are mentioned indicating that the concept of the basic uniformity of the cortex is untenable. The claim is discussed that non-decomposability is not an intrinsic property of complex, integrated systems but is only in our eyes, due to insufficient mathematical techniques. Using Rosen's modeling relation, the scientific analysis method itself is made a subject of discussion. It is concluded that the fundamental assumption of cognitive science, i.e., cognitive and other complex systems are decomposable, must be abandoned.

1 Introduction

During the last decade, the idea has gained popularity that time is ripe to build new computing systems based on information processing principles derived from the working of the brain. Thus, corresponding research programs have been initiated by leading research organizations (see [1], and references therein).

Obviously, these research initiatives take for granted that the operational principles of the brain as a complexly organized system are sufficiently known to us, and that at least a qualitative concept is available which only needs to be implemented into an operational, quantitative model. Tuning the model then could be achieved since lots of empirical data are available, due to the ever-improving experimental techniques of neuroscience.

Trying to put this idea into practice, however, has generally produced disenchantment after high initial hopes and hype. If one rhetorically ask "What is going wrong?", possible answers are: (1) The parameters of our models are wrong; (2) We are below some complexity threshold; (3) We lack computing

J. Mira et al. (Eds.): IWINAC 2009, Part I, LNCS 5601, pp. 195–204, 2009.

power; (4) We are missing something fundamental and unimagined (see [2] for related problems in robotics). In most cases, only answers (1)-(3) are considered by computer engineers and allied neuroscientists, and appropriate conclusions are drawn. If answer (1) is considered true, still better experimental methodologies are demanded to gather the right data, preferably at the molecular genetic level [3]. Answers (2) and (3) often induce claims for concerted, intensified efforts relating phenomena and data at many levels of brain organization [4].

Together, any of answers (1)-(3) would mean that there is nothing in principle that we do not understand about brain organization. All the concepts and components are present, and need only to be put into the model. This view is widely taken; it represents the belief in the efficiency of the *scientific method*, and it leads one to assume that our understanding of the brain will major advance as soon as the 'obstacles' are cleared away.

As I will show in this paper, there is, however, substantial evidence in favour of answer (4). I will argue that, by following the standard scientific method, we are in fact ignoring something fundamental, namely that biological and engineered systems are basically different in nature.

2 The Standard Approach to Brain Complexity

There is general agreement that brains, even those of simple animals, are enormously complex structures. At the first moment, it seems almost impossible to cope with this complexity. Which methods and approaches should be used? Brains are said to have fortunately - miraculously? - a property that allows us to study them scientifically: they are organized in such a way that the specific tasks they perform are largely constrained to different sub-regions. These regions can be further subdivided in areas that perform sub-tasks [4,5,6,7].

A well-known exponent of this concept is Marr [8] who formulated much of these ideas. In order to explain the human capacity of vision, he discussed detection of contours, edges, surface textures and contrasts as sub-tasks. Their results, he suggested, are combined to synthesize images, 2 1/2-D sketches and the representation of form. This kind of approach has been also employed in other areas of cognition such as language and motor control. Common assumption is that human behavior and cognition can be partitioned into different functions, each of which can be understood independently and with algorithms specific to the area of study. Obviously, this strategy illustrates the standard method used in science for explaining the properties and capacities of complex systems. It consists in applying a decompositional analysis, i.e. an analysis of the system in terms of its components or subsystems.

Since Simon's *The Sciences of the Artificial* [9], decomposability of cognitive and other complex systems has been accepted as fundamental for the Cognitive and Computational Neuroscience (CCN). We call this the fundamental assumption for CCN, for short: FACC. Simon [9], Wimsatt [10] and Bechtel and Richardson [11] have spent much work to elaborate this concept. They consider decomposability a continously varying system property, and state, roughly, that systems fall on a continuum from aggregate (full decomposable) to integrated (non-decomposable). The FACCN states that real systems are non-ideal aggregate systems; the capacities of the components are internally realized (strong intra-component interactions), and interactions between components do not appreciably contribute to the capacities; they are much weaker than the intra-component interactions. Hence, the description of the complex system as a set of weakly interacting components seems to be a good approximation. This property of complex systems, which should have evolved through natural selection, was called *near-decomposability* [9]. Simon characterizes near-decomposability as follows: "(1) In a nearly decomposable system, the short-run behaviour of each of the component subsystems is approximately independent of the short-run behaviour of the other components; (2) in the long run the behaviour of any one of the components depends in only an aggregate way on the behaviour of the other components" [9, p.100].

Thus, if the capacities of a near-decomposable system are to be explained, to some approximation its components can be studied in isolation, and based on their known interactions, their capacities eventually combined to generate the system's behavior. In CCN (and in other areas of science), the components of near-decomposable systems are called *modules*. This term originates from Engineering; it points at the assembly of a product from a set of building blocks with standardized interfaces. Thus, modularization denotes the process of decomposing a product into building blocks (modules) with specified interfaces, driven by the designers interests and intended functions of the product. Modularized systems are linear in the sense that they obey an analog of the *superposition principle* of Linear System Theory in Engineering [13]. This principle represents the counterpart of the decomposition analysis method which therefore is often denoted as *reverse engineering method*. A corresponding class of systems is characterized in Mathematics by a theorem stating that for homogeneous linear differential equations, the sum of any two solutions is itself a solution. The terms *linear* and *nonlinear* are often used in this sense: linear systems are decomposable into independent modules with negligible interactions while nonlinear systems are not [13,14].

Applying this concept to the systems at the other end of the complexity scale, the integrated systems are basically non-decomposable, due to the nonlinear interactions involved. Thus, past or present states or actions of any or most subsystems always affect the state or action of any or most other subsystems. In practice, analyses of integrated systems nevertheless try to apply the methodology for decomposable systems, in particular if there is some hope that the interactions can be linearized. Such linearizable systems were denoted above as nearly decomposable. However, in the case of strong nonlinear interactions, we must accept that decompositional analysis is not applicable to integrated systems. Their capacities depend in non-negligible way on the interaction between components, and it is not possible to identify component functions contributing to the system capacity under study. The question then arises, should we care about integrated systems, given the FACCN that all relevant systems are

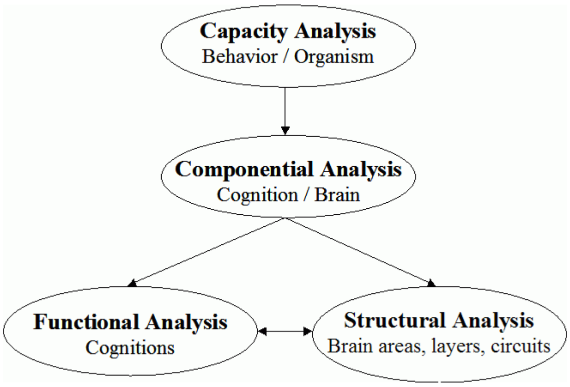

Fig. 1. View on decompositional analysis of brain and behavior. See text for details.

nearly decomposable? Non-decomposability then would be only in our eyes, and not an intrinsic property of strongly nonlinear systems, and - as many cognitive and computer scientists believe - scientific progress will provide us with the new mathematical techniques required to deal with nonlinear systems. We will return to this problem in Section 4.

In CCN, two types of componential analysis must be differentiated, i.e. functional and structural decomposition (see [12] for a clear, intelligible exposition of these matters). If one attempts to identify a set of functions performed by some (as yet unspecified) structural components of the system, a functional analysis is undertaken. Structural analysis involves to attempt to identify the structural, material components of the system. Functional analysis and structural analysis must be clearly differentiated, although in practice, there is a close interplay between them (as indicated by the arrows in Figure 1). Functional analysis should also be differentiated from capacity analysis. The former is concerned with the functions performed by components of the whole system which enable this whole system to have certain capacities and properties. The latter is concerned with the dispositions or abilities of the whole system, whereas functional and structural analysis is concerned with the functional and structural bases of those dispositions or abilities.

Especially important in the present context is this caveat: There is no reason to assume that functional and structural components match up one-to-one! Of course, it might be the case that some functional components map properly onto individual structural components - the dream of any cognitive scientist working as *reverse engineer*. It is rather probable, however, for a certain functional component to be implemented by non-localized, spatially distributed material components. Conversely, a given structural component may implement more than one distinct function. According to Dennett [15, p. 273]: “In a system as complex as the brain, there is likely to be much ‘multiple, superimposed functionality’ ”. With other words, we cannot expect specific functions to be mapped to structurally bounded neuronal structures, and vice versa.

3 Decompositional Brain Analysis

A guiding idea about the organization of the brain is the hypothesis of the columnar organization of the cerebral cortex. It was developed mainly by Mountcastle, Hubel and Wiesel, and Szenthágothai (e.g. [16,17,18]), in the spirit of the highly influential paper " The basic uniformity in structure of the neocortex" published in 1980 by Rockel, Hiorns, and Powell [19]. According to this hypothesis (which has been taken more or less as fact by many experimental as well as theoretical neuroscientists), the neocortex is composed of *building blocks* of repetitive structures, the *columns* or *neocortical microcircuits*, and it is characterized by a basic canonical pattern of connectivity. In this scheme all areas of neocortex would perform identical or similar computational operations with their inputs.

Referring to and based on these works, several projects started recently, among them the *Blue Brain Project.* It is considered to be "the first comprehensive attempt to reverse-engineer the mammalian brain, in order to understand brain function and dysfunction through detailed simulations" [20]. The central role in this project play cortical microcircuits. As Maas and Markram [21] formulate, it is a "tempting hypothesis regarding the computational role of cortical microcircuits ... that there exist genetically programmed stereotypical microcircuits that compute certain basis functions." Their paper well illustrates the modular approach fostered, e.g. by [4,22,23]. The tenet is that there exist fundamental correspondences among the anatomical structure of neuronal networks, their functions, and the dynamic patterning of their active states.

Starting point is the 'uniform cortex' with the cortical microcircuit or column as the structural component. The question for the functional component is answered by assuming that there exists a one-to-one relationship between the structural and the functional component (see Section 2). Experimental results confirming these assumptions are cited, but also some with contrary evidence. Altogether the modularity hypothesis of the brain is considered to be both structurally and functionally well justified. As quoted above, the goal is to substantiate the hypothesis "that there exist genetically programmed stereotypical microcircuits that compute certain basis functions".

Let us consider the general structure of the decompositional analysis of the cortex performed from computational point of view. In the modular approach, the problem of the capacity to be analyzed often is not discussed explicitly. Founding assumption of Cognitive Science is that "cognition is computation", i.e. the brain produces the cognitive capacities by computing functions. We know from mathematical analysis and approximation theory that a continuous function $f : R \rightarrow R$ can be expressed by composition or superposition of basis functions. This leads to the functional decomposition as follows: The basic functions are computed by the structural components (cortical microcircuits), and the composition rules are contained implicitly in the interconnection pattern of the circuits.

Obviously, this type of approach simplifies the analysis very much. The question is, however, Are the assumptions and hypotheses made appropriate, or must they considered as too unrealistic?

In fact, most of the underlying hypotheses have been questioned only recently. To start with the assumptions about the structural and functional components of the cortex, the notion of a basic uniformity in the cortex, with respect to the density and types of neurons per column for all species, turned out to be untenable (e.g. [24,25,26]). It has been impossible to find the cortical microcircuit that computes specific basis functions. No genetic mechanism has been deciphered that designates how to construct a column. It seems that the column structures encountered in many species (but not in all) represent spandrels (structures that arise non adaptively, i.e. as an epiphenomenon) in various stages of evolution.

If we evaluate the modular approach as discussed above, it is obvious that the caveat expressed in Section 2 has been largely ignored. There is evidence, however, for a certain functional component to be implemented by spatially distributed networks and, vice versa, for a given structural component to implement more than one distinct function. With other words, it is not feasible for specific functions to be mapped to structurally bounded neuronal structures [24,25,26]. This means, although the column is an attractive idea both from neurobiological and computational point of view, it cannot be used as an unifying principle for understanding cortical function. Thus, it has been concluded that the concept of the cortex as a large network of identical units should be replaced with the idea that the cortex consists of large networks of diverse elements whose cellular and synaptic diversity are important for computation [27,28,29]. It is worth to notice that the reported claims for changes of the research concept belong to the category of answers (1)-(3) to the question “What is going wrong?” (Section 1). A more fundamental point of criticism is formulated in the spirit of answer (4); it concerns the method of decompositional analysis itself and will be discussed in the next section.

4 Salient Features of Complex, Integrated Systems

In Section 2, we concluded that integrated systems are basically not decomposable, thus resisting the standard analysis method. We raised the question, Should we at all care about integrated systems, given the FACCN that all relevant systems are nearly decomposable? According to the prevalent viewpoint in CCN, non-decomposability is not an intrinsic property of complex, integrated systems but is only in our eyes, due to insufficient mathematical techniques (e.g. [5,30]). Bechtel and Richardson, instead, warn that the assumption according to which nature is decomposable and hierarchical might be false: “There are clearly risks in assuming complex natural systems are hierarchical and decomposable” [11, p. 27].

Rosen [31,32] has argued that understanding complex, integrated systems requires that the scientific analysis method itself is made a subject of discussion. A powerful method of understanding and exploring the nature of the scientific method provides Rosen’s modeling relation. It is this relation by which scientists bring “entailment structures into congruence” [31, p. 152]. What does this mean?

The modeling relation is the set of mappings shown in Figure 2 [33,34]. It relates two systems, a natural system N and a formal system F, by a set of

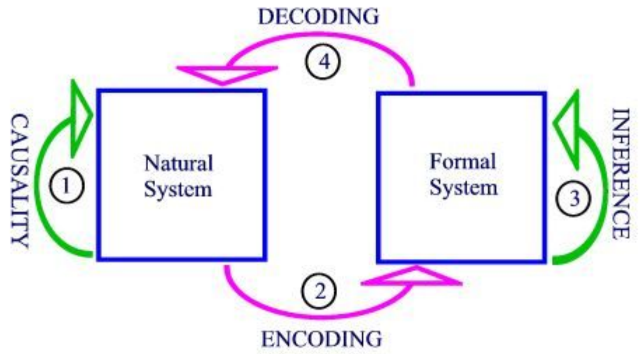

Fig. 2. Rosens Modeling Relation. A natural system N is modeled by a formal system F. Each system has its own internal entailment structures (arrows 1 and 3), and the two systems are connected by the encoding and decoding processes (arrows 2 and 4). From .

arrows depicting processes and/or mappings. The assumption is that this diagram represents the various processes which we are carrying out when we perceive the world. N is a part of the physical world that we wish to understand (in our case: organism, brain), in which things happen according to rules of causality (arrow 1). On the right, F represents symbolically the parts of the natural system (observables) which we are interested in, along with formal rules of inference (arrow 3) that essentially constitute our working hypotheses about the way things work in N, i.e. the way in which we manipulate the formal system to try to mimic causal events observed or hypothesized in the natural system on the left. Arrow 2 represents the encoding of the parts of N under study into the formal system F, i.e. a mapping that establishes the correspondence between observables of N and symbols defined in F. Predictions about the behavior in F, according to Fs rules of inference, are compared to observables in N through a decoding represented by arrow 4. When the predictions match the observations on N, we say that F is a successful model for N.

It is important to note that the encoding and decoding mappings are independent of the formal and/or natural systems. In other words, there is no way to arrive at them from within the formal system or natural system. That is, the act of modeling is really the act of relating two systems in a subjective way. That relation is at the level of observables; specifically, observables which are selected by the modeler as worthy of study or interest. Given the modeling relation and the detailed structural correspondence between our percepts and the formal systems into which we encode them, it is possible to make a dichotomous classification of systems into those that are *simple* or *predicative* and those that are *complex* or *impredicative.* This classification can refer to formal inferential systems such as mathematics or logic, as well as to physical systems. As Rosen showed [33], a simple system is one that is definable completely by algorithmic method: all the models of such a system are Turing-computable or simulable. When a single dynamical description is capable of successfully modeling a system, then the behaviors of that system will, by definition, always be correctly predicted. Hence, such a system will be *predicative* in the sense, that there will exist no unexpected or unanticipated behavior. Simple systems are decomposable sensu Simon [9](Section 2). This is the basis for the classical scientific method, the compositional analysis.

A complex system is thus by exclusion not a member of the syntactic, algorithmic class of systems. Its main characteristics are as follows. A complex system possesses non-computable models; it has inherent impredicative loops in it. This means, it requires multiple partial dynamical descriptions - no one of which, or combination of which, suffices to successfully describe the system. It is not a purely syntactic system, it necessarily includes semantic elements, and is not formalizable. Complex systems also differ from simple ones in that complex systems are not simply summations of parts - they are non-decomposable. This means, when a complex system is decomposed, its essential nature is broken by breaking its impredicative loops. This has several effects. Decompositional analysis is inherently destructive to what makes the system complex - such a system is not decomposable without losing the essential nature of the complexity of the original system! In addition, by being not decomposable, complex systems no longer have analysis and synthesis as simple inverses of each other. How you build a complex system is not simply the inverse of any analytic process of decomposition into parts. In other words, reverse engineering a cognitive system (which is a complex, integrative and thus non-decomposable system) will not enable its full undertanding!

5 Conclusions

Given the characteristics of complex systems - being non-decomposable, non-formalizable, non-computable - can such systems be studied by the scientific method at all? Indeed, they can, provided we acknowledge the inherent limitations of the compositional analysis if questions on the scale of the complex whole are to be answered. In the present context, this means that the fundamental assumption for CCN (cognitive and other complex systems are decomposable) must be abandoned.

Instead, we must consider the set of simple, predicative models of the organism, its behavior and brain *in the limit*, i.e. the infinite set of models, each providing partial dynamical descriptions. Thus, we cannot expect any ultimate model but a multitude of models corresponding to the infinite possible aspects of analysis.

In the case of complex biological systems, Rosen argued in favor of an approach oriented to study them at the level of the organizational structure of the system. This approach of *Relational Biology* - originally created by Rashevsky - involves composing descriptions of organisms at the functional level, thereby retaining the impredicative complexity. Of course, this approach (as well as, e.g., the related concept of autopoietic systems [35]) is not compatible with the standard engineering approach which is oriented to gain control over systems, be it natural or artificial. We must learn, however, to take into account the impredicativities as essential characteristics of complex, integrative systems. This will avoid exaggerated expectations und pitfalls in projects investigating the brain in order to derive operational principles which can be used for unconventional computing models.

References

1. Schierwagen, A.: Brain Organization and Computation. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4527, pp. 31–40. Springer, Heidelberg (2007)
2. Brooks, R.: The relationship between matter and life. Nature 409, 409–410 (2001)
3. Le Novere, N.: The long journey to a Systems Biology of neuronal function. BMC Syst. Biol., 1–28 (2007)
4. Grillner, S., Markram, H., De Schutter, E., Silberberg, G., LeBeau, F.E.N.: Microcircuits in action - from CPGs to neocortex. Trends in Neurosciences 28, 525–533 (2005)
5. van Vreeswijk, C.: What is the neural code? In: van Hemmen, J.L., Sejnowski Jr., T. (eds.) 23 Problems in System neuroscience, pp. 143–159. Oxford University Press, Oxford (2006)
6. Furber, S., Temple, S.: Neural systems engineering. J. Roy. Soc. Interface 4, 193–206 (2007)
7. Anderson, J.A.: A brain-like computer for cognitive software applications: the Ersatz Brain project. In: Fourth IEEE International Conference on Cognitive Informatics, pp. 27–36 (2005)
8. Marr, D.: Vision. W. H. Freeman & Co., New York (1982)
9. Simon, H.: The Sciences of the Artificial. MIT Press, Cambridge (1969)
10. Wimsatt, W.: Forms of aggregativity. In: Donagan, A., Perovich, A.N., Wedin, M.V. (eds.) Human Nature and Natural Knowledge, pp. 259–291. D. Reidel, Dordrecht (1986)
11. Bechtel, W., Richardson, R.C.: Discovering complexity: Decomposition and localization as strategies in scientific research. Princeton University Press, Princeton (1993)
12. Atkinson, A.P.: Wholes and their parts in cognitive psychology: Systems, subsystems, and persons (1998), `http://www.soc.unitn.it/dsrs/IMC/IMC.htm`
13. Schierwagen, A.: Real neurons and their circuitry: Implications for brain theory. iir-reporte, Akademie der Wissenschaften der DDR, Institut für Informatik und Rechentechnik, Seminar “Neuroinformatik”, Eberswalde, 17–20 (1989)
14. Forrest, S.: Emergent Computation: Self-Organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks. Physica D 42, 1–11 (1990)
15. Dennett, D.C.: Consciousness explained, p. 273. Little, Brown & Co., Boston (1991)
16. Hubel, D.H., Wiesel, T.N.: Shape and arrangement of columns in cat’s striate cortex. J. Physiol. 165, 559–568 (1963)
17. Mountcastle, V.B.: The columnar organization of the neocortex. Brain 120, 701–722 (1997)
18. Szentágothai, J.: The modular architectonic principle of neural centers. Rev. Physiol. Biochem. Pharmacol. 98, 11–61 (1983)
19. Rockel, A.J., Hiorns, R.W., Powell, T.P.S.: The basic uniformity in structure of the neocortex. Brain 103, 221–244 (1980)

20. Markram, H.: The Blue Brain Project. Nature Rev. Neurosci. 7, 153–160 (2006)
21. Maass, W., Markram, H.: Theory of the computational function of microcircuit dynamics. In: Grillner, S., Graybiel, A.M. (eds.) The Interface between Neurons and Global Brain Function, Dahlem Workshop Report 93, pp. 371–390. MIT Press, Cambridge (2006)
22. Arbib, M., Érdi, P., Szentágothai, J.: Neural Organization: Structure, Function and Dynamics. MIT Press, Cambridge (1997)
23. Bressler, S.L., Tognoli, E.: Operational principles of neurocognitive networks. Intern. J. Psychophysiol. 60, 139–148 (2006)
24. Rakic, P.: Confusing cortical columns. PNAS 105, 12099–12100 (2008)
25. Horton, J.C., Adams, D.L.: The cortical column: a structure without a function. Phil. Trans. R. Soc. B 360, 386–462 (2005)
26. Herculano-Housel, S., Collins, C.E., Wang, P., Kaas, J.: The basic nonuniformity of the cerebral cortex. Proc. Natl. Acad. Sci. USA 105, 12593–12598 (2008)
27. Destexhe, A., Marder, E.: Plasticity in single neuron and circuit computations. Nature 431, 789–795 (2004)
28. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Rev. Neurosci. 10, 186–198 (2009)
29. Frégnac, Y., et al.: Ups and downs in the genesis of cortical computation. In: Grillner, S., Graybiel, A.M. (eds.) Microcircuits: The Interface between Neurons and Global Brain Function, Dahlem Workshop Report 93, pp. 397–437. MIT Press, Cambridge (2006)
30. Poirier, P.: Be There, or Be Square! On the importance of being there. Semiotica 130, 151–176 (2000)
31. Rosen, R.: Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life. Columbia University Press, New York (1991)
32. Rosen, R.: Essays on Life Itself. Columbia University Press, New York (2000)
33. Rosen, R.: Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations. Pergamon Press, Oxford (1985)
34. Mikulecky, D.C.: Robert Rosen: the well posed question and its answer–Why are organisms different from machines? Syst. Res. 17, 419–432 (2000)
35. Maturana, H.R., Varela, F.J.: Autopoieses and Cognition: The Realization of the Living. D. Reidel, Dordrecht (1980); 49, pp. 27–29 (2006)