Science in Context **20**(1), 97–124 (2007). Copyright © Cambridge University Press
doi:10.1017/S0269889706001153 Printed in the United Kingdom

Getting Abstract Mathematical Models in Touch with Nature

Andrea Loettgers

California Institute of Technology

Argument

This article focuses on the scientific practice of a specific class of models in neuroscience and biology that approach biological systems as computational or information processing systems. This specific approach to biological systems has a long tradition that started with the information discourse in the 1940s. Borrowing concepts, methods, and techniques from cybernetics, information theory, and physics, these models are situated at the interface of different scientific disciplines. This article examines in detail the Hopfield model, a model of the specific brain function of auto-associative memory that is situated at the interface of neuroscience, theoretical physics, and engineering. By drawing analogies to a model of disordered magnetic systems in physics, John Hopfield constructed a model that was able to mimic auto-associative memory. The model became an active field of research in theoretical physics and in engineering but not in neuroscience. According to neuroscientists, in the process of construction the model had lost touch with the biological system and the function that it was supposed to model. How did this happen? As will be shown, the process of constructing the model was guided by conceptualizing the biological organism as a computational, information-processing device. This construction process gained its own momentum that involved an interrelated development of a computational concept and a redescription of the biological system or phenomenon for bringing the computational concept to work. We will see how physicists and engineers tried to get models in touch with biological systems by constructing "a synthetic model."

Introduction

Recent studies of the scientific practice of modeling (Sismondo 1999, Merz 1999, Boumans 1999, Knuuttila and Voutilainen 2003), initiated by the book *Models as Mediators* (Morgan and Morrison 1999), revealed the large number of different functions that models are fulfilling in science. These studies also provided insight into how scientists are learning from the construction, application, and modification of models. According to Morgan and Morrison, the various ways models may function in science is due to their partial independence from theory and the world: "We believe there is a significant connection between the autonomy of models and their ability to function as instruments. It is precisely because models are partially independent of both theories

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

and the world that they have this autonomous component and so can be used as instruments of exploration in both domains" (ibid., 10).

How does this concept of models as mediators apply in the context of interdisciplinary research? I am going to argue that the description of models as boundary objects turns out to be very helpful in this respect. Boundary objects (Star and Griesemer 1989) are objects that are located at the interface between different communities of practice. They are shared by several different communities but viewed or used differently by each of them. Boundary objects are flexible enough to adapt to local needs and have different distinct identities in different communities, but at the same time they are robust enough to maintain a common identity across the boundaries to provide a place for shared work. By careful analysis of a specific example, the so-called Hopfield model (Hopfield 1982), which is located at the interface of neuroscience, theoretical physics, and computer engineering, I want to show how the concept of boundary objects can help us to analyze models in interdisciplinary research contexts.

Historical setting plays an important role in such an analysis. The Hopfield model is a model of auto-associative memory and belongs to a specific group of models in biology that approach structures and functions of biological systems as instances of computation or information processing. Examples of such models can be found in neural networks, artificial intelligence, and in new developing fields such as systems biology and synthetic biology. As Lily Kay has shown, the conceptualization of biological systems as computational or information processing systems was developed in the 1940s when cybernetics, information theory, digital computing, and control systems were being reconfigured together in the information discourse (Kay 2000). If we want to understand how models like the Hopfield model which are part of the information discourse become boundary objects in interdisciplinary research settings, we must look into the specific conceptualization that is expressed in the configuration of specific notions, metaphors, methods, and techniques coming from cybernetics, information theory, digital computing, and control systems. We must ask how these configurations influence the way biological systems and their functions are represented in models. The process of fitting together these different parts that originated in engineering, mathematics, physics, and biology, results in a complex object. Tarja Knuuttila described this process: "Thus, scientific models are by their construction linked in complex and intricate ways to many kinds of knowledge originating from different fields and materialized in the machinery and modelling procedures, . . . and link to other knowledges and artefacts that have made their compilation work possible, and they mediate information only to those who have some relevant background knowledge" (Knuuttila and Voutilainen 2003, 1486).

In the case of the Hopfield model and of related models that are based on the view of biological systems as information processing or computational systems, this construction process leads to considerable epistemological difficulties. The construction process involves translating a complex biological system into an abstract and much less complex computational system. This translation process raises the troublesome question

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

of whether the guiding idea of conceptualizing life in terms of information and the corresponding notions, methods, and techniques from cybernetics and information theory are appropriate in describing and modeling biological systems. Joan Fujimura in a recent article on systems biology writes about this point: "The reason for the disentangling the materialities of engineered and biological systems is to delineate the various border-crossing in order to *understand what is lost in the translation*. By 'lost', I mean more than 'loss'. *Translations can distort, transform, delete, and add*" (Fujimura 2005, 220; emphasis in original). Reservations against this concept of life are expressed not only by philosophers, historians, and sociologists of science; but scientists themselves are also worried and are asking: "Is there any substance to this metaphor relating algorithms and the mechanics of life?" (Winfree 2001).

The first part of this article focuses on the construction of the Hopfield model and on the question of how this model became a boundary object located at the interface between neuroscience, theoretical physics, and engineering. The second part focuses on the epistemic difficulties associated with the Hopfield model and similar models, and discusses how the agenda of synthetic biology can be understood as an attempt to overcome these difficulties.

The Hopfield model: How do biological systems compute?

John Hopfield introduced his model, a distributed model of auto-associative memory, in 1982 in a paper entitled: "Neural Networks and physical system with emergent collective computational abilities" (Hopfield 1982). A nice example of auto-associative memory is Marcel Proust's famous scene in his *Remembrance of Things Past* in which he describes how his childhood memories are called up by the taste of a madeleine cake soaked in linden-blossom tea (Proust 1981). A more mundane example is the problem of character recognition. Hopfield's model provides a mechanism that allows just this kind of retrieval of information from incomplete data. In the model, the human brain's neurons are binary variables that form an interconnected network. The network may be storing a certain number of patterns. In the retrieval phase, one starts with some initial configuration and the network dynamic evolves towards the stored pattern that is closest to the initial configuration. The model belongs to the class of attractor neural networks.

Hopfield's paper is a classic in the field of neural networks, and the model is a standard example discussed in many textbooks on neural networks. James A. Anderson, a well known neuroscientist, wrote about this paper:

> As far as public visibility goes, the modern era in neural networks dates from the publication of this paper by John Hopfield. This paper is short, clearly written, and brings together a number of strands that up to this point were somewhat separated in the neural network literature. Much of the reason for its impact was simply this: it presented a sophisticated, coherent theoretical picture of how a neural network could work, and that it could do so. (Anderson and Rosenfeld 1988, 457)

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

Anderson added another reason for the paper's success:

> We cannot avoid making a comment about the sociology of science, it is hoped without causing offense. John Hopfield is a distinguished physicist. When he talks, people listen. Theory in his hands becomes respectable. Neural networks became instantly legitimate, whereas before, most developments in networks had been the province of somewhat suspect psychologists and neuroscientists, or by those removed from the hot centers of scientific activity. (Ibid., 457)

We will see that this assertion is somehow overstated. Neural networks did not become instantly legitimate with the Hopfield model, especially not in neuroscience. Hopfield's model opened up an active field of research for theoretical physicists, but their investigations were regarded with great skepticism by neuroscientists. The abstract model did not reflect the complexity of the biological systems with which they were dealing in their everyday experiments.

It turns out that there is a long history behind the discussion between biologists and neuroscientists on the one hand and non-biologists or non-neuroscientists on the other hand about the question of the usefulness of abstract mathematical modeling. We may recognize the same kind of discussion in a series of conferences that took place during the years 1946 to 1953 sponsored by the Josiah Macy, Jr. Foundation to promote cybernetics and its application to biology and social sciences (Kay 2000). We will discuss this history in more detail below. In the case of the Hopfield model, the particular interest of theoretical physicists was due to the analogies that the model offered to models of disordered magnetic systems that were being intensively studied in the statistical mechanics of spin models. The interest of computer engineers, on the other hand, can be explained by the model's potential for constructing new computer architecture.

From complex biological systems to less complex computational systems

In the introduction to his paper, Hopfield described the basic computational concept on which his model is based:

> In physical systems made from a large number of simple elements, interactions among large numbers of elementary components yield collective phenomena such as the stable magnetic orientations and domains in a magnetic system or the vortex patterns in fluid flow. Do analogous collective phenomena in a system of simple interacting neurons have useful "computational" correlates? For example, are the stability of memories, the construction of categories of generalization, or time-sequential memory also emergent properties and collective in origin? This paper examines a new modeling of this old and fundamental question . . . and shows that important computational properties spontaneously arise. (Hopfield 1982, 2554)

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

According to the above description the model connects the brain's capacity of auto-associative memory to collective phenomena in physics as well as to functions of computation. Linking auto-associative memory with collective phenomena is based on a functional analogy and cannot be justified on the basis of empirical knowledge about biological neuronal networks and the way they fulfill the function of auto-associative memory. As we will see, Hopfield justified his way of proceeding by arguing that biology is a part of physics and, furthermore, by claiming that the performance of complex biological and physical systems is based on the same basic mechanism that had to be discovered. This heuristic argument led him away from the actual biological phenomena and into theoretical physics. When his attempts to identify basic mechanisms of complex behavior were not successful, he started to draw analogies to models of magnetic spin systems in which magnetic properties are described as cooperative phenomena based on the interaction between the magnetic moments. In a second step, the mechanism of cooperative phenomena was afterwards interpreted as an instance of computation. Hopfield pointed out that the way he realized the specific computational concept in his model opened up the possibility of using it for the construction of new computer architectures. He wrote: "The model could be readily implemented by integrated circuit hardware. The conclusion suggests the design of a delocalized content-addressable memory or categorizer using extensive asynchronous parallel processing" (ibid.).

In addition to offering a possible working description of the brain's function of auto-associative memory, the model was also intended to offer new ways of designing computer architectures. Thus the conceptualization of the computer as a model of the brain and vice versa, was a further important element in the construction of the Hopfield model. We will see more evidence for this fact in the next paragraph.

Linking auto-associative memory with collective phenomena in theoretical physics and computation is a basic step in the construction of the model as a boundary object. Let us look at this linking process in more concrete detail and look more closely at the process that transforms a complex biological system into a less complex computational system. First we will provide some biographical information about John Hopfield and elaborate on the discussion between neuroscientists and non-neuroscientists about the usefulness and appropriateness of abstract mathematical modeling of brain structures and functions. This will give us some insight as to why it was so difficult for models such as the Hopfield model to become accepted as a scientific tool for the exploration of biological systems and functions.

Biographical notes on John Hopfield

John Hopfield, currently professor for molecular biology at Princeton University, is a well known and influential scientist in molecular biology and especially in the field of neural networks. Hopfield started his career as a theoretical physicist and worked

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

in the field of solid state physics before he moved into molecular biology and later into neuroscience. Early in his career, in 1958 when he was a Ph.D. student, Hopfield became associated with Bell Labs, the research and development arm of the American Telephone and Telegraph Company. Hopfield said that Bell Labs was an intellectually inspiring place for him. After a stay at the Ecole Normal Superieure in Paris from 1960–1961 and three years as Associate Professor of Physics in Berkeley, Hopfield became Professor of Physics at Princeton in 1964. In 1973 he again became associated with Bell Labs. Here he came into contact with Robert Shulman, who had founded a biophysics group in the condensed matter division at Bell Labs. Shulman, too, was a physicist by training and a specialist in Nuclear Magnetic Resonance (NMR), Electronic Spin Resonance (ESR), X-Ray, and Moessbauer measurements. Using these techniques Shulman's group performed measurements on biological molecules such as hemoglobin, which are very large compared with typical inorganic molecules. Hopfield worked on the problem of the transportation of oxygen by hemoglobin and the ability of enzymes to recognize substrates in processes such as DNA replication or protein synthesis (Hopfield 1974, 1978, and 1980).

He became interested in biological neural networks and special functions of the brain when he was invited to meetings of the Neuroscience Research Program (NRP) in the mid-1970s. These meetings were originally started in 1962 by Francis O. Schmitt, who was a professor of biology at the Massachusetts Institute of Technology. The general aim of these meetings was to bring together neuroscientists and non-neuroscientists to search for new ideas, methods, and techniques in research on complex neurophysiological phenomena. Hopfield attended as a non-neuroscientist.

At the time that Hopfield began to attend these meetings, Schmitt identified a "quiet revolution in the neurosciences" in a 1976 article co-authored with Parvati Dev and Barry Smith: "This quiet revolution in our concepts of synaptic organization significantly alters our understanding of neuronal interactions because it suggests the concept of neuronal 'local circuits' whose components may include many neurons joined through dendrodendritic junctions" (Schmitt, Dev, and Smith 1976, 253). The catalyst for the quiet revolution was, according to these authors, the application of the new technology of electron microscopy in neuroscience, which required alterations of traditional concepts of neural and synaptic organization. A year later the NRP organized an Intensive Study Program (ISP) at Boulder, Colorado, emphasizing the basic mechanism of higher brain functions. Schmitt outlined the aim of the meeting in an introductory volume published in 1978: "Such theories [theories of higher brain functions as learning, memory, perception] based on studies of the properties of partial systems, are subject to the component-systems dilemma that bedevils all attempts at biological generalization: they failed to explain how distributive global functional capabilities emerge from complex interactions of functional circuits" (Schmitt and Worden 1978, 1586).

The fourth ISP held under the auspices of the NRP at Boulder, Colorado, in June 1977, was designed to accelerate the emergence of a new phase in neuroscience based

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

on new information about brain structure and on unconventional concepts that had emerged primarily during the previous decade (see Schmitt, Dev, and Smith 1976). It was hoped that the ISP would pinpoint the most promising directions in research for an understanding of basic mechanisms of higher brain functions. John Hopfield started to attend the NRP meetings when members were discussing new approaches of analyzing and describing higher brain functions. In an interview in 2004, he told me that when he attended these meetings he became interested in neuroscience and began to look around for an interesting problem. Hopfield is an exceptional scientist in the way he chooses his objects of investigations. He says about himself that he looks for what he calls "ground-zero problems." Problems where no explanations exist. He likes to prepare the ground, give a possible explanation, and leave it to others to work out the details.

This kind of trans-disciplinary meeting organized by NRP was not new. In fact Schmitt had attended the famous Ninth Washington Conference on Theoretical Physics (1946) organized by George Gamow,[1] which brought together biologists and physicists, mathematicians, and engineers. On the list of participants we find next to Schmitt, Goerge W. Beadle, Niels Bohr, Max Delbrueck, James Frank, Sol Spiegelmann, Leo Szilard, Edward Teller, and John von Neumann (Kay 2000). Other examples of such meetings are the Macy conferences (1946–1953) where participants came from backgrounds ranging from neurophysiology, medicine, psychology, psychiatry, social sciences, anthropology, philosophy, mathematics, and engineering. The main forces behind these meetings were Norbert Wiener and Warren McCulloch. Together with Walter Pitts, McCulloch had introduced the first neural network in an article entitled "A Logical Calculus of the Ideas Immanent in Nervous Activities" in 1943 (McCulloch and Pitts 1943). This neural network formed an important part and expression of the emerging information discourse. Lily Kay wrote, "their project was not merely an application of mathematical models to physiological and mental phenomena – such practices were certainly not novel. Rather, the novelty in their notion of mind lay in equating the operations of reason with those of binary-logic neurons." According to Kay, McCulloch could neither convince "psychologists nor psychiatrists (though he was an active member of the American Psychiatrist Association); the neurophysiologists, preoccupied with amassing empirical knowledge, were generally suspicious of theories of mind. . . . Instead the McCulloch-Pitts model attracted mavericks willing to cross boundaries among mathematics, electrical engineering, physiology, psychiatry, and psychology" (Kay 2001, 608).

Suspenseful discussions developed between those "mavericks" applying neural networks in their research on brain functions and neurophysiologists. Examples can be found in some of the conference proceedings, e.g., the one published after the symposium on Neural Theory and Modeling organized in 1962. Contributions to the

[1] George Gamow was a very colorful scientist who worked in theoretical physics and genetics and became well known for his book character, Mr. Tompkins.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

symposium were published two years later (Reiss 1964). The volume is divided into two parts. The first theoretical part includes contributions from non-neuroscientists. The second part is reserved for neuroscientists. Each of the parts is prefixed by an overview of the current state of the art. In a 1982 paper, Hopfield cited Leon Harmon's overview article, which was part of the first "theoretical" part. Harmon had a background in computer communication and was working at Bell Labs. He described the role of the migrating scientists in neural modeling: "With growing emphasis placed on the information-processing aspects of nervous systems, theoretical studies assume greater importance. In some instances questions pertaining to neurophysiology action appear to be unanswerable by any experimental techniques now known or easily envisioned. In such cases theoretical modeling can significantly complement direct physiological experimentation" (Reiss 1964, 54). Harmon sees the potential of theoretical modeling in offering an alternative way of finding explanations for processes and phenomena that cannot be examined in experiment, thus offering a different approach to the nervous system. In his formulation, viewing the nervous system as an information processor opens the study to scientists from other fields such as mathematicians, computer engineers, and theoretical physicists.

The introduction to the second part of the book, which had been reserved for neuroscientists, was written by Donald Kennedy, a neuroscientist from Stanford University who was working on the mechanisms by which animals generate and control patterned motor output. Kennedy reacted very cautiously to Harmon's optimism because he somehow feared a takeover of the field by theoreticians like Harmon. He wrote:

> It is certainly legitimate to ask whether such "poachers" are not actually ready to take over the entire preserve, converting the former owners into a class of touchingly old-fashioned ranger-naturalists whose main task is the collection of new specimens for the theorists to ponder. There are several reasons for believing that biologists must play a wider role than that. First, their education, though it needs remodeling in some respects, has given them a respect for organic diversity which is crucial to the understanding of living machinery. The biologist has a healthy skepticism of universals, he knows that organisms come up with the embarrassing number of adaptive solutions when challenged with a problem in evolutionary engineering.
>
> Second, the biologically trained neurophysiologist can provide a working knowledge of the constraints which the unit properties of neurons and junctions place upon possible models; in other words, he can keep models in some sort of touch with the systems they are meant to model. (Reiss 1964, 341)

Kennedy designs a vision of how the field could be reorganized. Tasks are distributed amongst the different disciplines. In his vision, the non-neuroscientists stay in their fields and the neuroscientists provide the knowledge, experience, and, as he says,

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

"keep models in some sort of touch with the systems they are meant to deal with" (ibid., 339).

Let us return to John Hopfield and his path into neuroscience. In addition to the meetings of the NRP, an important element in the elaboration of a possible explanation of the function of auto-associative memory was a class on physics and computation that Hopfield started to teach shortly after he came to Caltech in the early 1980s. The class was co-taught by Carver Mead, who worked in the electrical engineering department at Caltech. The goal of the course was, as Hopfield put it, "to consider digital computers, analog computers, and the brain as having things in common" (Hey 1999, 3). In the second year, Hopfield continued this class together with Richard Feynman. They invited scientists such as Marvin Minsky, John Wheeler, and Gerry Sussman to give talks about computation which afterwards would be critically analyzed by Feynman. They focused on the following three points: "First, what limits did the laws of physics place on computation? Second, why was it that when we wanted to do a hard problem a little better, it always seemed to need an exponentially greater amount of computer resources? Third, how did the human mind work, and could it somehow get around the second problem?" (ibid., 1999).

Computation as a physical phenomena

Hopfield's main interest concerned the third point, how the human mind works. Following the tradition of McCulloch and Pitts, and relating to the information discourse, Hopfield approached the problem with the understanding of biological functions (such as auto-associative memory) as instances of computation in biological systems based on physical processes. He elaborated on this point in an article entitled "Physics, Biological Computation and Complementarity" (Hopfield 1986) which is based on a talk he gave at the Niels Bohr centenary symposium in 1985. Hopfield began this article with the statement: "Biology as we know it lies in the domain of physics." He goes on:

> The laws of elementary particle physics and cosmology and the history of the universe serve merely to determine the nature of a planetary environment. The dynamical equations of quantum mechanics and quantum electrodynamics (and their classical equivalents when appropriate) are the essential elemental laws of physics which lead to biology. Some physicists make claims that "we shall never understand life until we understand the origins of the elementary particles." But the real mysteries of biology lie in the way in which these dynamical laws of physics, and the substrate of electrons, photons and nuclei on which they operate, produce the complex set of counter-intuitive phenomena labeled with the term biology. (Hopfield 1986, 52)

For Hopfield, the processes observed in biology can be reduced to the laws of physics in order to find out what laws of physics underlie the observed phenomena, for example, the transportation of oxygen by hemoglobin, protein synthesis, and

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

auto-associative memory. Hopfield elaborated further on the analogy between physical and biological systems by comparing physical systems that show complex behavior with biological systems that show complex behavior:

> Driven (non-equilibrium) physical systems of simple components already show complex and almost unpredictable behaviors. Turbulence in fluid flow, deterministic chaos and fractal forms in snowflakes are a few of the complex phenomena which arise from the same simple physical laws and substrates that rule biology. Some of the basic unsolved problems of biology, such as the generation of complex forms in large systems, can already be seen in these simple systems. (Ibid., 53)

In this reductionistic approach, where biological phenomena can be reduced to physical laws, Hopfield connects complex physical and biological systems by stating that the complex behavior in both systems are caused by the same physical laws. This is the basic assumption on which Hopfield built and by which he justified the analogy drawn between physical and biological systems. It created the starting point of the actual model construction. Hopfield's goal at this very early stage in the construction process was to search for a mechanism which underlies the behavior of complex systems and which would be true for all systems showing complex behavior and properties. It should be mentioned that research on complex systems had just started to develop as a field of research. For example the Santa Fe Institute devoted to research on complex systems was founded in 1986.

Before I describe Hopfield's search for the underlying mechanism of complex systems, let me describe in one last point how Hopfield connected his approach on biological systems as complex systems to the property of biological systems as computational systems. In his paper from 1986 he writes:

> Biology can be seen as a hierarchy of computations and computational devices. The translation of DNA into protein structure is a kind of computation. The construction of a complex organism from the instructions in DNA is also the following of an algorithm. The recognition of a familiar object is a neural computation. (Ibid., 296)

For Hopfield, biological systems are complex computational systems. They share the basic mechanism with complex physical systems, which produce complex behavior and underlie various processes such as DNA replication, protein synthesis, learning, and pattern recognition. Looking for universal mechanisms shared by biological and physical systems and based on physical laws is the approach of a theoretical physicist, and this is how Hopfield approached problems in neurophysiology and biology. This approach is very similar to the research program of another famous physicist who turned to biology, Max Delbrueck. Lily Kay describes Delbrueck's research program thus: "Delbrueck saw himself as a natural philosopher in pursuit of new universal laws, liberated from the traditional division that had prejudiced man's study of nature and

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

seeking knowledge in its most integrated form" (Kay 1985, 209). There is an extensive debate about the role of physicists in biology, especially in molecular biology (see for example Morange 1998, Part I, chap. 7). But I do not intend to get into this debate or make a contribution to it because for our specific object of investigation, the process of substantiating the conceptualization of life and living organisms by means of models, we are interested only in how being a physicist affects the way a model is constructed.

Hopfield's approach to the question of how biological systems fulfill all their different functions or more specifically how the brain fulfills the task of auto-associative memory, diverged from earlier approaches, such as the McCulloch-Pitts model of viewing the brain as an information processor based on mathematical logic. Following the tradition in which the mind and the computer are models for each other, McCulloch and Pitts had based information processing on mathematical logic. Hopfield pointed out that in order to gain insight into the processes of higher brain function, one has to proceed the opposite way and concentrate on how the brain *differs* from the conventional computer. More precisely, Hopfield thought that computation in conventional computers and biological systems differ in the way in which they ensure reliability. He wrote: "Biology compounds this problem by adding a unique view of the nature of errors and of logic itself" (Hopfield 1986, 297). This realization was mainly a result of his research on the question as to how nature reassures the high accuracy by which enzymes find the right amino acid in the process of protein synthesis. Hopfield developed a specific model, the kinetic proofreading model (Hopfield 1974, 1978, 1980), in which he gave an explanation for this effect by drawing on analogies to non-equilibrium processes in physical systems. The main elements of the model are described in the following section.

The kinetic proofreading model or why the brain works differently

In an enzyme mediated reaction an amino-acid C is bound to its adequate tRNA. This reaction is described by the reaction sequence:

$$E + C \rightleftharpoons EC \rightarrow P_C, \tag{1}$$

where E stands for the enzyme, C for the amino-acid, and P_C for the product formed out of the amino-acid and the tRNA. In addition to this reaction another reaction occurs in which the wrong amino-acid D is bound to the tRNA

$$E + D \rightleftharpoons ED \rightarrow P_D. \tag{2}$$

The minimal error made in the process of recognition is determined by:

$$f_0 = \exp(-\Delta D_{CD}/RT), \tag{3}$$

where ΔG_{CD} is the maximal difference between the free energy of C and D along the reaction path. The error rate observed in protein synthesis is much lower than f_0.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

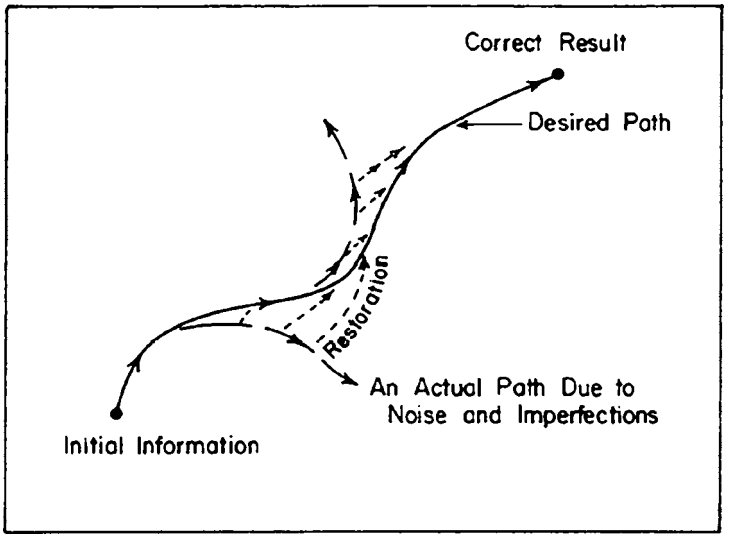


Fig. 1. The trajectory moving in the phase space (Hopfield 1986).

Hopfield began his analysis of how biological systems assure high accuracy by describing how the process would be performed in the manner of a computer program. The instruction set would look like the following:

- Read the next three bases on the mRNA molecule "tape."
- Look up the corresponding amino acid in the genetic code "dictionary."
- Add the amino acid to the protein.
- Shift the input tape by three bases.

Following these instructions, the competing reaction simply does not occur. But the competing reaction does happen in the biological organism. From the point of view of chemistry, the competing reaction can happen because "the reaction is rather similar to the one desired, and any enzymatic system which allows the correct reaction to take place must also allow the incorrect reaction, albeit with rather slower rates." The result will be "useless proteins being formed because a particular incorrect amino acid is as likely to be added as a correct one."

Hopfield continued: "The logic which we would like the biological system to display – correct amino acid only – is not possible without errors, and is possible with low but finite error only by deliberately running the system out of equilibrium" (Hopfield 1986, 55).

But what does it mean to run a system out of equilibrium? Hopfield used the following illustration to explain what he means (see fig. 1). The trajectory visualizes the desired course of the reaction by which the enzyme chooses the correct amino acid in the physical phase space that is spanned by all the possible states of the system.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

The second similar reaction takes a trajectory that diverges from the correct path and the biological system requires a restoration mechanism to return to the pathway. That restoration can be provided by introducing "a physical process which squeezes the system back down onto the proper track in the state space." In his kinetic proofreading model he does this by adding a high energy intermediate in the pathway of the reaction. Without going into technical details, this high energy intermediate runs the reaction out of equilibrium, thus limiting the space occupied in the phase space. This ensures that the system is able to be restored, and, as Hopfield noted, even with restoration, biological computers are inaccurate. But: "In biology, being as accurate as possible in computations may be a losing proposition! Biology is not interested in perfect logic" (ibid., 298). Hopfield is using the computer as a model for the mind in the sense that he is focusing on the difference.

Concretization of the general picture

Drawing the analogy between complex biological and complex physical systems and looking for basic underlying mechanisms responsible for complex behavior is a very general idea. Somehow implicit was the idea that the mechanisms causing complex behavior are simple. Hopfield started to concretize this general idea by focusing on the concrete function of auto-associative memory which he displayed in the simple and abstract picture of trajectories moving into attractors (see fig. 2). The trajectories, lines marked with arrows, represent the process of Proust's recollection of his childhood that was released by the taste of the madeleine. The trajectories move into attractors associated with Proust's childhood memories.

The great challenge of this simple picture was to find the underlying dynamics. Hopfield used the Game of Life for exploring different possibilities of such dynamics. The Game of Life is not a game in the conventional sense. It was introduced in 1970 by the mathematician John Conway and simulates the birth, reproduction, and death of a living organism (Conway 1970). In this game, one has a rectangular grid (the exact form of the grid is not of importance) on which one has cells which can be in different states; they can live, die, or multiply. In addition to this grid of cells in different states, one has rules which determine how the states of the cells change over time. What makes the Game of Life, and cellular automata in general, attractive is that they are examples of emergence and self-organization. Complex behavior is reproduced starting from simple elements and rules, which scientists use as tools for studying the complex behavior of systems such as weather, climate, stock markets, social dynamics, traffic, and communication. For Hopfield, the Game of Life represented a possibility of finding a realization of the dynamic process in the framework of complex systems.

In the case of the biological neuronal network, the cells of the Game of Life became neurons. And instead of cells that were either dead or alive, the neurons were either active or inactive. Hopfield deviated from the rules by changing the dynamical

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

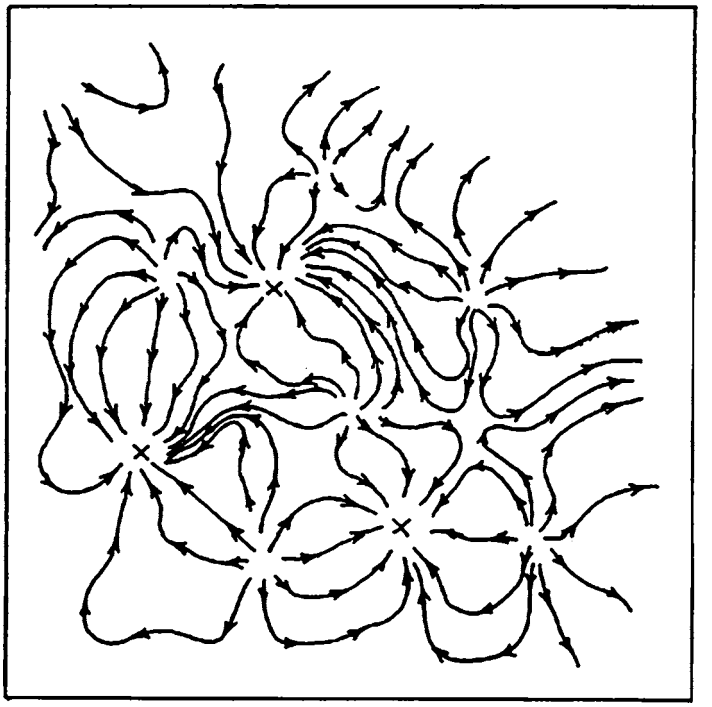


Fig. 2. Visualization of the dynamics. The attractors represent the stored pattern into which the system moves in the course of dynamics (Amit 1992).

rules from synchronous to asynchronous. In the synchronous case, the states of all the cells are updated at the same time. In the case of the brain, this would mean that there is something like a central pacemaker giving the signal for the global update of the neurons. Asynchronous dynamics, where at each step the state of only one neuron is calculated, seemed to be more realistic. Experiments with different rules did not produce satisfying results. The process Hopfield was looking for could not be reproduced by the Game of Life. The model was not able to represent processes that, starting from an incomplete input, the incomplete memory would develop into a stable state: the recollected memory. A severe drawback was the fact that, at that time, one could only look at static pictures that were produced very slowly. Technological restrictions made it impossible to thoroughly investigate the dynamics resulting from the different rules in order to see if the dynamic would lead to a stable state or not.

In experimenting with the Game of Life, the complex biological network and the function of auto-associative memory were redescribed as a network of interconnected binary variables that performed actions based on simple rules. This approach – trying to simplify and to abstract the system into the model in such a way that details are left out even if they are known in order to find the essential properties that explain the

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

phenomena under investigation – is a well established and widely used modeling tradition in theoretical physics (Horner and Kuehn 1998). We will see that Hopfield, in the further construction process, held fast to this modeling tradition. In the following step, Hopfield started to draw analogies to models in statistical physics for investigating magnetic properties, such as ferromagnetism where the magnetic moments tend to align in one direction below a critical temperature. In these models, ferromagnetism is described as a cooperative phenomenon based on the interaction between the magnetic moments.

Hopfield had not been the first to try to use such models to describe brain functions such as pattern recognition, learning, or short-term memory. As early as 1954, the physiologist B. G. Cragg and the theoretical physicist H. N. V. Temperley investigated the possibility that the organization and function of the brain is structured similarly to complex interacting systems investigated in statistical physics. They characterized their basic idea as "a possible analogy between the organization of neurons and the kind of interaction among atoms which leads to co-operative processes." As an example, they discuss "the phenomenon of ferromagnetism, which arises specifically from *interaction* of atoms that would not otherwise be ferromagnetic" (Cragg and Temperley 1954). And in 1974 William Little, a theoretical physicist like Hopfield, introduced a memory model by drawing analogies to the Ising model which models the property ferromagnetism as a cooperative phenomena (Little 1974).

Interlude: The Ising model

In the standard Ising model, the magnetic moments S_i are fixed on a rectangular lattice and can take on values $S_i = \pm 1$, corresponding to the different spin directions. Only the spins located on the nearest neighbor lattice points interact in the model. The interaction is described by a single constant J, representing the interaction energy. This quantity is a function of S_i and S_j and is negative if S_i and S_j have the same sign and it is positive if they have opposite signs. The total interaction energy E is given by:

$$E = \sum_{\langle i,j \rangle} -J\, S_i\, S_j, \tag{4}$$

where $\langle i, j \rangle$ runs over all nearest neighbors and is at its minimum at maximal order of the system i.e., when all spins point in the same direction. This means that the simple Ising model has two energy minima corresponding to the case of all spins up or all spins down.

The system represented by the Ising model shows what physicists call critical behavior around a critical temperature T_c, the so-called Curie temperature, where the system undergoes a phase transition from the paramagnetic, the disordered state, into the ferromagnetic, the ordered state. This behavior of magnetic systems below a critical temperature is an example of collective behavior.

The Ising model is part of the tradition in statistical physics of investigating the properties of complex systems such as crystals, gases, liquids, and magnetic systems,

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

by models that are based on the assumption that the *interaction* between the elements are responsible for the observed phenomena, such as ferromagnetism. For most of the collective behavior of such systems, the properties of the individual elements are not important. As has already been discussed by others such as R. I. G. Hughes, the Ising model has been used extensively in the study of critical phenomena (Hughes 1999). In this context the fact that the model shows only little correspondence to real magnetic systems was not regarded as a disadvantage. The Ising model functions as a *representation* for some class of other objects or processes in the world that share similar features (Morgan 2003). The unspecific and abstract structure of the Ising model enabled physicists to use it to represent various systems that share the property of showing critical behavior, such as gases, alloys, and liquids.

Little used this specific property of the Ising model for modeling the ability of the human brain to connect two events that are separated in time. For example, we are able to connect an event that took place in our childhood with another event that took place when we were a teenager. This means, on the one hand, we have to be able to remember those events and, on the other hand, we also have to be able to connect them. The fact that Little's paper is very abstract and mathematical and lacks possible applications explains why it did not gain much recognition outside the theoretical physics community. The same is true for the work of, e.g., Steven Grossberg, Shun-Ischi Amari, and Teuvo Kohonen, who worked on competitive and cooperative mechanisms to describe dynamics in neural networks in the late 1960s and during the 1970s. These models did not become boundary objects because they could not be shared by different communities, i.e., scientists from different scientific backgrounds could not link to the models.

In the following construction process, Hopfield integrated both concepts, competition and cooperation, when he started to draw an analogy to models of disordered magnetic systems, such as the Sherrington-Kirkpatrick spin glass model.

In these models, the disorder arises from the fact that both ferromagnetic and antiferromagnetic couplings are present. A classical example for a spin glass is an alloy of gold and iron (*AuFe*). Here gold, *Au*, a non-magnetic substance is weakly diluted with iron, *Fe*. Such a system does not have a regular lattice structure such as one finds in conventional crystals. The *Fe* atoms are randomly distributed in an *Au* matrix. This also means that their magnetic moments are randomly distributed. The couplings are a function of the distance between the magnetic moments $J_{ij} = J(R_i - R_j)$, with R_i and R_j as the positions of the spins. The values of the couplings are randomly distributed and can be positive, i.e., ferromagnetic, or negative, i.e., antiferromagnetic. The concurrent existence of ferromagnetic and antiferromagnetic couplings leads to a "conflict" for some spins of the system, when they cannot satisfy at the same time both ferromagnetic and antiferromagnetic coupling. Ferromagnetic couplings tend to align the spins, antiferromagnetic couplings make them point in opposite directions. Marc Mezard, Giorgio Parisi, and Miguel Angel Virasoro explain the phenomena by the following illustrative example:

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

> In a classical tragedy the scenario may be the following: there is a fight between two groups and the various characters on the scene have to choose sides. In addition they all have strong personal feelings, positive or negative, towards each other (it is a tragedy!). Some of them are friends and some of them are enemies. For simplicity we will assume that all feelings are reciprocal; otherwise the system may never reach equilibrium . . . . Let us consider three characters (A, B, and C); if A and B, B and C, A and C do like each other, there is no problem: they will all choose the same side. In a similar way, if A and B are friends and C is an enemy of both, then A and B can be on one side and C will be on the other. Frustration follows, instead, if A, B, and C hate each other because two personal enemies must then fight on the same side. (Mezard, Parisi, and Virasoro 1987, 5)

The competing interaction between the ferromagnetic and antiferromagnetic couplings is at the origin of the observed complex behavior in spin glasses. But as the simple example shows, this form of competition could also be the origin of complex behavior in other systems, where instead of a competing interaction one has, for example, competing goals.

The complex behavior of spin glasses appears at low temperatures after the system undergoes something like a phase transition. The systems exhibit a "freezing transition" to a state with a new kind of "order" in which the spins are aligned in random directions (Binder and Stauffer 1979). At high temperatures the system is in a paramagnetic phase. Cooling down the substance leads to the freezing of the spins. This transition is not an ordinary phase transition because it does not lead to an ordered phase. Scattering with neutrons, for example, shows that the material in the spin glass phases does not have a periodical order.

One of the most debated questions concerning spin glasses is whether the transition is a genuine phase transition and, if so, whether it is a new kind of phase transition. Experimentally spin glasses are very hard to investigate. One of the main difficulties is the very long relaxation time. They can be of the order of several generations of scientists. Consider for example measurements of the magnetization m. Without an external magnetic field the magnetization m is zero. But if a probe is cooled down in an external magnetic field and then the field is switched off, the substance remains magnetized but the magnetization decreases only very, very slowly.

The detailed understanding of the physics of this class of magnetic substance is a lively and very active part of contemporary physics. Since the 1970s, spin glasses have been intensively studied and a great number of materials have been found which show spin glass properties and behavior.

The Sherrington-Kirkpatrick (SK) model

In 1975 David Sherrington and Scott Kirkpatrick introduced their model for spin glasses in an article titled "Solvable Model of a Spin-Glass" (Kirkpatrick and Sherrington 1975). This seminal paper was followed in 1978 by a longer article

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

"Infinite-ranged models of spin-glasses" (Kirkpatrick and Sherrington 1978) in which they undertook a more detailed analytical and numerical examination of the properties of their model.

The SK model was very attractive in condensed matter physics because it seemed to open up the prospect of allowing analytical solutions and thus an exact understanding of the properties of spin glasses, such as phase transitions, and an understanding of the nature of the spin glass phase at low temperatures. One of the questions under debate was whether at low temperatures the system relaxes very slowly into a global energy minima or whether it becomes trapped in one of the many local energy minima. For a while the situation was very promising but the analytical treatment of the model also had its limits. These arose mainly as a consequence of a mathematical trick, the so-called replica trick, which was introduced in order to deal with the disorder in spin glasses. An in-depth discussion of the technical details of the analytical treatment of spin glasses goes beyond the scope of this article.

The basic structure of the model is relatively simple: a system of interacting binary variables where the binary variables represent the spins, S_i, or equivalently the magnetic moments. The S_i can take either value $+1$, which corresponds to spin direction "up," or -1, which corresponds to spin direction "down." The coupling between two spins S_i, and S_j is described by parameters J_{ij} which enter into an energy function of the form:

$$E = -\sum_{i,\,j} J_{ij} S_i S_j. \tag{5}$$

The disorder is modeled by choosing the J_{ij} Gauss distributed,

$$P(J_{ij}) = \frac{1}{\sqrt{2\Pi J^2}} \exp\left[-\frac{(J_{ij} - J_0)^2}{2J^2}\right]. \tag{6}$$

$P(J_{ij})dJ_{ij}$ is the probability that a given coupling takes a value between J_{ij} and $J_{ij} + dJ_{ij}$. Positive values of J_{ij} correspond to ferromagnetic and negative values to antiferromagnetic couplings.

These are the basic elements of the model. Remarkable about the model is its simple structure. It is a strong simplification of the actual situation in spin glasses. For example, magnetic moments are very complex objects if fully described in the context of quantum mechanics. The interaction between the magnetic moments is also much more complex than the model would have it.

Hopfield first learned about spin glasses and attempts to model them when he was participating in Phil Anderson's group seminars at Bell Labs. Later, after his experiments with the Game of Life did not yield any satisfying results, Hopfield started to draw analogies to the SK spin glass model. This meant extending the property of the SK model of being a representation for a class of systems with which it shares specific features to the modeling of biological systems and functions.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

A comparison between the SK and the Hopfield models shows that the two models agree in their basic structure. The Hopfield model consists of a system of N model neurons. The states of the neurons $\sigma_i(t)$ are described by binary variables, where $\sigma_i(t) = +1, (0)$ characterizes an active (passive) neuron. This representation of the state of the neurons agrees with the neuron model introduced in 1943 by McCulloch and Pitts (McCulloch and Pitts 1943) and at the same time agrees with the modeling of the spins in the Ising and SK model. The model neurons were connected by exchange couplings J_{ij}. The exchange couplings in the case of the Hopfield model are used for storing the memory item. Hopfield used the Hebb rule which states: a simultaneous activation of two connected neurons results in a strengthening of the synaptic coupling between the two neurons (Hebb 1949). This rule is formalized in the Hopfield model as follows

$$J_{ij} = \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu}, \tag{7}$$

where the ξ_i^{μ} are variables that describe a pattern, i.e., a given configuration of active and inactive neurons. The number of patterns stored in the network is given by p and in each pattern the number of neurons is equal to the total number of neurons in the network, N.

The structural similarity between the two models is obvious. In his 1982 paper, Hopfield explained this similarity in the following way: "When T_{ij} [in our notation J_{ij}] is symmetric but has a random character (the spin glass) there are known to be many (locally) stable states." If the energy minima correspond to stored information, a large amount of information could be stored in the model system. The second advantage of the SK model resulted from a specific property of its energy function. In order to see this property, one must take a closer look at the model dynamics. In the Hopfield model, the post-synaptic potential $h_i(t)$ is produced by the activating signals arriving from all the other neurons to which the neuron i is connected is given by:

$$h_i(t) = \sum_{j=1}^{N} J_{ij} \sigma_j(t). \tag{8}$$

Hopfield introduced the following dynamic into his model:

$$\sigma_i(t) = \begin{cases} +1, & \text{if } hi > 0, \\ 0, & \text{if } hi > 0. \end{cases} \tag{9}$$

In the course of the dynamics, a neuron σ_i is randomly chosen and updated according to the update rule. Under the right conditions (symmetric couplings, no self couplings, and a limited pattern number), the so described dynamics will drive the system from an incomplete pattern into the associated complete pattern. This property

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

is due to the energy function which, similar to the Ising and SK model, is given by:

$$E = -\frac{1}{2}\sum_{i \neq j} J_{ij}\sigma_i\sigma_j, \tag{10}$$

and which is, as Hopfield could show, a Lyapunov function which decreases in the course of the dynamics. Starting from an incomplete input pattern the dynamics will lead into the nearest energy minimum, the one that corresponds to the stored pattern.

The dynamic rule corresponds to the computational algorithm underlying the function of auto-associative memory. This algorithm is tied to the specific model system, with its specific simplified structure and restrictions, such as symmetric and no self-synaptic couplings. It mimics auto-associative memory, i.e., the system starting from an incomplete pattern is able to recollect the associated pattern but there is no way to prove that biological neural systems function in an analogous way. So it may not be astonishing that the validation of the model took place mostly in the context of theoretical physics.

Validation of the model

The validation of the Hopfield model started about two to three years after the model was published in 1982 when theoretical physicists working on spin-glasses realized the similarity between the Hopfield and the SK model (Amit, Gutfreund, and Somolinsky 1985a and 1985b).

The Hopfield model can be understood as the inverse spin glass problem because in the Hopfield model the energy minima are given and in the SK model they are the unknown. Theoretical physicists therefore used the Hopfield model as a different approach to explore the properties of the SK model. In the following years, an interesting dynamic between the Hopfield model and SK model developed which contributed to a validation of the Hopfield as well as the SK model. In the case of the SK model, the specific "mathematical trick" (replica trick) which was introduced to study the behavior of the system at low temperatures had produced results which suggested that the "trick" was not appropriate. The Hopfield model was used to explore both the SK model itself and the mathematical trick. By a controlled variation of the two parameters – number of stored patterns and noise (temperature) – the Hopfield model could be transformed into an SK model. This allowed a controlled experimentation on the SK model and its behaviors and properties, and, at the same time, an examination of the behavior and properties of the Hopfield model and its function as an auto-associative memory. Theoretical physicists used the Hopfield model for their specific purposes. The same is true for the computer engineers who worked on electronic and optical realization of the model.

With this exploration of the Hopfield model, the investigations of the theoretical physicists did not leave the proper sphere of the model. They were concerned about

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

the inner consistency of the model, i.e., whether it could be understood with the concepts and laws of statistical physics. They did not relate the model to the biological object and function that it was supposed to represent.

The case of the investigation of the model by engineers is very similar. Engineers worked on the electronic and optical realization of the model. Their hope for new computer architectures that would be faster and more efficient aroused some public interest. For example, in March 1986 an article on the front page of the *Los Angeles Times* published with the title "Associative Memory Technology: The Brain May Yet Meet Its Match in a Computer," claimed:

> By redesigning the blueprints of computers to mimic the architecture of the brain, the Caltech biophysicist [John Hopfield] hopes to produce computers that will perform tasks thousands, perhaps even millions, of times faster than the best existing machines and be able to solve problems that conventional computers now solve only with great difficulty.

A number of new companies were founded that attempted to build neural computers. Tony Materna, an engineer working for the Hecht Nielsen corporation. in San Diego, said in a 1987 interview in the *Los Angeles Times* that at the end of the century a neuronal computer will be in every place where a conventional computer is now used.

The Hopfield model as a boundary object

The specific way the Hopfield model was constructed makes it a boundary object between theoretical physics and computer engineering. Neuroscientists did not include the model in their research on biological neural networks and their specific functions. The new concept of computation, computation as a competitive or emergent property, had been introduced but the model was not able to relate this form of computation to the specific biological function of auto-associative memory. What are the reasons for this?

There are, I believe, several partly connected reasons. First, Hopfield approached the problem very much as a theoretical physicist using knowledge, experiences, methods, and techniques from statistical mechanics. He made use of the specific tradition of modeling many body systems in statistical mechanics that tries to find the important parameters in the system that are responsible for the observed phenomena, leaving aside all the others. The Ising and SK spin glass models are prominent representations of this tradition, and, in addition, they have the important attribute of being representations for a class of systems that share specific properties in biology. Hopfield imported this tradition and the related models into biology. He extended the scope of the models into the realm of biology in accordance with his reductionistic approach. Physicists could easily link to the Hopfield model, they could use their analytical instruments

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

from statistical physics to explore the model and to apply it in the investigation of spin glasses. They dipped under the simple and clear structured surface of the model into the complex underlying structure of statistical mechanics with its open question about critical point behavior of spin glass systems, phase transitions, structure of energy landscapes, and dynamics. Computer engineers remained at the surface and translated the structure of the model into electronic networks consisting of transistors and resistors.

The Hopfield model had been flexible enough to adapt to the special needs of theoretical physicists and computer engineers. It also maintained a common identity by retaining its basic structure of being a network of interconnected binary variables. All this is not true for neuroscience and can be explained by the fact that the construction process as described earlier involved an interrelated development of a computational concept and a redescription of the biological system or phenomenon in a way that enables the computational concept to work. This resulted in an inferential gap (Morgan) between the model and the actual object that it was supposed to represent. The fact that this gap could not be bridged was the main reason that neuroscientists were prevented from linking to the Hopfield model. In the words of Kennedy, the model "lost its touch" with the system it was meant to represent.

Given these circumstances one may find it astonishing that Hopfield's ideas about computation in biological systems did not remain in theoretical physics and computer engineering but did, after all, enter into molecular biology and neuroscience. The special graduate program, the Center for Computation and Neural Systems, founded in 1989 with the strong support of John Hopfield at the California Institute of Technology, played an important role. Its aim was to introduce physicists and engineers to neuroscience, and enable its students to use theories, methods, and techniques from physics and engineering to approach problems from neuroscience. The Center describes itself as:

> . . . an interdisciplinary Ph.D. program to study problems arising at the interface between neuroscience, electrical engineering, computer science, and physics. The unifying theme of today's program is the relationship between the physical structure of a computational system (molecular, neuronal or electronic hardware), the dynamics of its operation, and the computational problems that it can efficiently solve.

Similar centers and programs were subsequently founded at many other universities and became essential in the further exploration, application, and extension of Hopfield's research program of finding computational concepts based in physics. These special graduate programs can be understood as a way of standardizing methods on a more specific level than the information discourse. Physicists and computer engineers worked on establishing a link to neuroscience and molecularbiology by bringing the abstract computational concepts in touch with the biological systems. The new field of synthetic biology can be understood as a result of these efforts.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

The synthetic approach

Synthetic biology can at least partly be understood as a result of the described program of introducing students from physics and engineering into neuroscience and as a further development of Hopfield's program of searching for new computational concepts.

Within this development, one can recognize an interesting movement to reduce the inferential gap by constructing and using synthetic models. In synthetic biology, starting from an abstract mathematical model describing, for example, the circadian clock of Drosophila that oscillates with a 24-hour rhythm and self-synchronizes to the day and night cycle (Elowitz and Leibler 2000), networks are engineered out of DNA sequences and proteins (genes and cells) and are consequently transformed into a synthetic model. The behavior of synthetic models is investigated by implementing them into cells. These kinds of models seem to be located between very abstract mathematical models, like the Hopfield model, and biological systems. The exploration of the properties of these synthetic models has the character of an experiment on material objects. In 2005 a report "Synthetic Biology Applying Engineering to Biology" was conducted for the European Program NEST (New and Engineering Science and Technology).[2] In this report, synthetic biology is described as being "concerned with applying the engineering paradigm of systems design to biological systems in order to produce predictable and robust systems with novel functionalities that do not exist in nature." To make this shift visible, I will discuss the example of the Repressilator, which was introduced in 2000 by Stansilav Leibler and Michael Elowitz.

The Repressilator

The Repressilator is a synthetic oscillating network that was constructed in order to gain insight into the function of the circadian clock, regulating the day and night rhythm in Drosophila. To explain why they used the synthetic approach, Leibler and Elowitz wrote:

> Networks of interacting biomolecules carry out many essential functions in living cells, but the "design principles" underlying the functioning of such intracellular networks remain poorly understood, despite intensive efforts including quantitative analysis of relatively simple systems. Here we present a complementary approach to this problem: the design and construction of a synthetic network to implement a particular function. We used three transcriptional repressor systems that are not part of any natural biological clock to build an oscillating network, termed the Repressilator, in *Escherichia coli*. (Elowitz and Leibler 2000, 335)

[2] The report can be found under: www.univ-poitiers.fr/recherche/documents/pcrdt7/syntheticbiology.pdf.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

According to Leibler and Elowitz the main motivation for constructing synthetic networks was to gain an understanding of the *design principle* of regulating cell mechanism. The Repressilator is a specific network linked to a specific function. Examining one such synthetic network will not be enough to infer the general possible *design principles* that underlie the regulating mechanism of the various cell functions. But synthetic biologists hope that by studying a number of such systems and comparing the different *design principles*, they may be able to identify general elements in the structure of the network and the accompanied regulating mechanism.

The basic structure of the Repressilator is a DNA ring containing three genes: *tet*R, *lac*I, and λ*c*I, which encode the proteins TetR, LacI, and λcI. For a gene to be translated into a protein, the enzyme polymerase must bind to a region of the DNA, called promotor. Polymerase then transcribes the gene into messenger RNA, which in turn is translated into a protein. In the Repressilator the proteins selectively bind to one another's promoter and inhibit the expression of the gene and respectively the production of the protein. Such an arrangement is called a negative feedback loop and leads to temporal oscillations in the protein concentrations. These oscillations in protein production can be made visible when one of the protein products in the feedback loop is linked to a gene for making a green fluorescent protein. If the circuit is introduced into the bacteria *E. coli* the oscillation in the protein production can be observed by the blinking of the bacteria.

Following this description, the construction of the synthetic model contains two essential steps, first the construction of the synthetic network and second the observation of the behavior of the network *in vivo*. In a special issue of *Nature*, published in 2005 and addressing the question of contents, methods, and aims, an overview article appeared entitled "Reconstruction of genetic circuits" published with a number of other articles addressing the question of contents, methods, and aims. In it, the authors, Sprinzak and Elowitz describe this construction process as a reconstruction. They write: "In this approach, one constructs synthetic replicas of natural genetic circuits out of well-characterized elements, such as genes, proteins, regulatory sequences, and so on, and observes their dynamics in living cells." As an illustration of their approach they provide the following example: "By taking apart an old clock, you could probably come up with a pretty good guess at how it works. But more concrete understanding of the clock mechanism might be obtained by designing and building one's own clock" (Sprinzak and Elowitz 2005, 443).

A careful reading of this quotation shows that the notion of "reconstruction" has to be taken with some caution. It does not mean the replication of known genetic circuits and corresponding regulating mechanisms. The regulating mechanisms are the unknown which synthetic biologists aim to understand by their approach. Leibler and Elowitz also wrote in the 2000 article that the synthetic genetic circuits "are not part of any natural biological clock." The construction process usually starts out from a mathematical model, which can be the differential equations of an oscillator. In the next step, the mathematical model is transformed into a synthetic network by making

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

use of the available empirical knowledge about the components of the network. After introducing it into a cell the behavior of the network is then observed *in vivo*.

From general to more specific

Why could synthetic networks represent a step forward in understanding functions of biological systems? If one compares Leibler's and Elowitz's repressilator with Hopfield's model of auto-associative memory a couple of important differences become visible. First of all, the repressilator focuses on a much more specific biological system and function than the Hopfield model, where specific experimental and bio-chemical data are available that can be used in the construction process. Second, the repressilator has the same materiality as the natural genetic networks. This means the constraints in constructing and manipulating the network are not, as in case of the Hopfield model, given by mathematics but by nature itself. This represents a step in the direction of reducing the inferential gap between the model and the natural system.

In constructing and using synthetic models, scientists have left abstract mathematical space and are moving in the direction of experiments. Synthetic models can be best described as hybrid systems, consisting of models and experiments. But does this approach help to overcome the worries connected with viewing and describing biological systems and their functions as information processing or computational systems? This question is difficult to answer. Construction of the underlying mathematical models is still guided by the information approach, as can be seen by the fact that control theory coming from engineering plays an important role in this process. These elements from information theory find their expression in such structures as negative and positive feedback loops in the synthetic models. Synthetic biologists say that these kinds of feedback loops have been found in experiments on regulating mechanisms in cells. The question of whether concepts of information processing and control are adequate for describing basic regulating cell mechanisms requires further analysis of the construction process and use of synthetic models.

Concluding remarks

Such attempts as the Macy conferences, the Neuroscience Research Program, and the Meeting on "Neural Theory and Modeling" discussed at the beginning of this article have revealed the difficulties of establishing models, like neural networks, as research tools in the neurosciences. Using Donald Kennedy's words, these models missed "the touch" with the systems they are meant to represent. Neuroscientists could not link to the models, or to put it the other way around, the models could not adapt to the local needs of neuroscientists. Models like the Hopfield model contradicted the complexity observed in experiments. Neural networks, such as the Hopfield model,

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

were located at the interface of neuroscience because they were supposed to represent biological neural networks and associated functions but in the understanding of the neuroscientists they failed to fulfill this function.

Examination of the construction process of the Hopfield model and further developments connected with the model, such as the foundation of interdisciplinary graduate programs and the emergence of new fields of synthetic biology, shows that efforts to overcome the inferential gap between models and biological systems have taken place in communities of physicists and engineers working on questions of the structure and function of biological systems, such as in the case of the repressilator. This first analysis shows that the difficulty of getting abstract mathematical models in touch with biological systems and functions, and the resulting failure of these models to function as boundary objects in the case of biology, led to synthetic models whose materiality and resulting hybrid structure were supposed to ensure the touch. This then led to the creation of the new field of synthetic biology. But even if the synthetic models have the same materiality as the biological systems the question still remains whether the inferential gap really can be overcome by this approach. As earlier, in the case of the Hopfield model, the construction of the models is based on a concept of biological systems as information processing or computational systems, a concept that functions as a guiding principle in both construction processes. Validation of the synthetic model, which means experimenting on the model and observing the behavior *in vivo* turns out to be a rather complex enterprise. At this moment, it still seems unclear whether synthetic biology will lead to new insights into the basic cell regulating mechanism. Further analysis of the construction and application of these models will be necessary.

References

Amit, Daniel, J. 1992. *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge: Cambridge University Press.

Amit, Daniel J., Hanoch Gutfreund, Haim Somolinsky. 1985a. "Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks." *Physical Review Letters* 55(14):1530–1533.

Amit, Daniel J., Hanoch Gutfreund, Haim Sompolinsky. 1985b. "Spin-Glass Models of Neural Networks." *Physical Review* A32(2):1007–1018.

Anderson, James A. and Edward Rosenfeld. 1988. *Neurocomputing*. Cambridge, Mass. And London: MIT Press.

Binder, Kurt and Dietrich Staufer. 1979. "Monte Carlo Studies of Systems with Disorder." In *Monte Carlo Methods in Statistical Physics*, edited by K. Binder, 225–260. Berlin-Heidelberg-New York: Springer.

Boumans, Marcel. 1999. "Built-in justification." In *Models as Mediators*, edited by Mary Morgan and Margaret Morrison, 66–97. Cambridge University Press.

Conway, John. 1970. "Mathematical Games." *Scientific American* 223:120.

Cragg, B. G. and H. N. V. Temperely. 1954. "The organization of neurons: a cooperative analogy." *EEG and Clinical Neurophysiology* 6:85.

Elowitz, Michael and Stanislas Leibler. 2000. "A synthetic oscillator network of transcriptional regulators." *Nature* 40:335–338.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

Fujimura, Joan. 2005 "Postgenomic futures: Translations across the machine-nature border in systems biology." *New Genetics and Society* 24(2):195–225.
Hebb, Donald. 1949. *The Organization of Behavior*. New York: Wiley.
Hey, Anthony. G. 1999. *Feynman and Computation: Exploring the Limits of Computers*. Boulder: Westview Press.
Hopfield, John J. 1974. "Kinetic Proofreading: A New Mechanism for Reading Errors in Biosynthetic Processes Requiring High Specificity." *Proceedings of the National Academy of Science USA* 71:4135–4139.
Hopfield, John J. 1978. "Origin of the genetic code: A testable hypothesis based on tRNA structure, sequence, and kinetic proofreading." *Proceedings of the National Academy of Science USA* 75:4334–4338.
Hopfield, John J. 1980 "The energy delay: A proofreading scheme based on dynamic cooperativity and lacking all characteristic symptoms of kinetic proofreading in DNA replication and protein synthesis." *Proceedings of the National Academy of Science* 77:5248–5252.
Hopfield, John J. 1982. "Neural Networks and physical system with emergent collective computational abilities." *Proceedings of the National Academy of Science USA* 79:2554–2558.
Hopfield, John J. 1986. "Physics, Biological Computation, and Complementarity." In *The Lesson of Quantum Theory, Proceedings of the Niels Bohr Centenary Symposium* edited by D. de Boer, E. Dahl, O. Ulfbeck, 295–314. Amsterdam: Elsevier Science Publishers.
Horner, Heinz. and Reimer Kuehn. 1998. "Neural Networks." In *Intelligence and artificial inteligence: an interdisciplinary debate*, edited by U. Ratsch, M. M. Richter, I. O. Stamatescu, 5749–5762. Heidelberg: Springer.
Hughes, R. I. G. 1999. "The Ising model, computer simulation, and universal physics." In *Models as Mediators*, edited by Mary Morgan and Margaret Morrison, 97–146. Cambridge: Cambridge University Press.
Kay, Lily E. 1985. "Conceptual Models and Analytical Tools: The Biology of Physicist Max Delbrueck." *Journal of the History of Science* 182:207–246.
Kay, Lily E. 2000. *Who Wrote the Book of Life?* Stanford: Stanford University Press.
Kay, Lily E. 2001. "From Logical Neurons to Poetic Embodiments of Mind: Warren S. McCulloch's Project in Neuroscience." *Science in Context* 4(4):591–614.
Kirkpatrick, S. and D. Sherrington. 1975. "Solvable Model of a Spin-Glass." *Physical Review Letters* 35(26):1792–1796.
Kirkpatrick, S. and D. Sherrington. 1978. "Infinite-ranged models of spin-glasses." *Physical Review B* 17:4384.
Knuuttila, Tarja and Atro Voutilainen. 2003. "A Parser as an Epistemic Artefact: A Material View on Models." *Philosophy of Science* 70(5):1484.
Little, W. A. 1974 "The Existence of Persistence States in the Brain." *Mathematical Biosciences* 19:101.
McCulloch, Warren S. and Walter A. Pitts. 1943. "A logical calculus of the ideas immanent in neural nets." *Bulletin of Mathematical Biophysics* 5:115–133.
Merz, Martina. 1999. "Multiplex and Unfolding: Computer Simulation in Particle Physics." *Science in Context* 12(2):293.
Mezzard, Marc, Giorgio Parisi, Miguel Angel Virasoro. 1987. *Spin Glass Theory and Beyond*. Singapore, New Jersey, Hong Kong: World Scientific.
Morange, Michel. 1998. *A History of Molecular Biology*. Translated by Matthew Cobb. Cambridge, Mass. and London: Harvard University Press.
Morgan, Mary S. and Margaret Morrison. 1999. *Models as Mediators*. Cambridge University Press.
Morgan, Mary. 2003. "Experiments without Material Intervention." In *The Philosophy of Scientific Experimentation*, edited by Hans Radder, 103–116. Pittsburgh: Pittsburgh University Press.
Proust, Marcel. 1981. *Remembrance of Things Past*. Translated by C. K. Scott Moncrieff and Terence Kilmartin. New York: Random House.
Reiss, Richard F. 1964. *Neural Theory and Modeling*, Proceedings of the 1962 Oja Symposium. Stanford: Stanford University Press.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press

Schmitt, Francis O., Parvatt Dev, and Barry H. Smith. 1976. "Electronic processing of information by brain cells." *Science* 193:114.

Schmitt, Francis O. and Frederic G. Worden. 1978. *The Neurosciences*. Cambridge and London: M.I.T. Press.

Sismondo, Sergio. 1999. "Models, Simulations, and Their Objects." *Science in Context* 12(2):24.

Sprinzak, David and Michael B. Elowitz. 2005. "Reconstruction of Genetic Circuits." *Nature* 438:443–448.

Star, Susan L. and James R. Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–1939." *Social Studies of Science* 19:387–420.

Winfree, Eric. 2001. "The DNA and Natural Algorithm Group: Research Perspectives." http://www.dna.caltech.edu/DNAresearch_perspective.html.

https://doi.org/10.1017/S0269889706001153 Published online by Cambridge University Press