



UvA-DARE (Digital Academic Repository)

Cell-based models

Tamulonis, C.V.T.

Publication date
2013

[Link to publication](#)

Citation for published version (APA):
Tamulonis, C. V. T. (2013). *Cell-based models*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

1

Introduction

The concept that matter is composed of simple elements beyond the reach of the naked senses dates back to the fourth century B.C., when Democritus and Leucippus established the atomic philosophy. The core idea was a simplification of the position of the Pythagoreans, who held that all matter was a combination of four basic elements: earth, water, air and fire. The four element theory is thought to have been based on the action of fire: burn a piece green wood and you see not only fire released, but also the smoke becomes air, the water boils off from the ends and the result is an earth-like ash. The atomists, however, believed that the processes and qualities observed at one level cannot be used to explain those at a deeper level, and so, for example, the color of objects cannot be explained simply by assuming the presence of “colored” elements (1). Instead, the atoms were assumed to be all of the same substance, but distinguished in size and shape and exhibiting different dynamic properties, such as oscillations. Democritus imagined an infinite vacuum filled with atoms in perpetual motion constantly colliding

and forming more complicated elements, which humans perceived as various different substances. In Democritus' words: "According to convention there is a sweet and a bitter, a hot and a cold, and according to convention there is color. In truth there are atoms and a void."

The atomic theory had no observed facts to support it, nor was there any means of testing its consequences. Theories of matter were more a matter of doctrine than science and subject to be pushed aside by changes in mindset or by a rival philosopher. In the history of science the Democritean atomic theory is closer to modern views than any other theories which preceded or replaced it, but was ultimately undone by the criticisms of Plato and Aristotle. The revival of the atomist theory would only come about two millennia later with the emergence of Newtonian physics¹.

Naturally we associate the abstract Greek atoms with the modern concepts of molecules, atoms, and subatomic particles. However the fundamental idea is more general than may appear: a seemingly complex phenomenon has at its root simpler interacting elements. Although the interactions between a few elements in isolation may be well understood, the sheer combinatorial possibilities of even a moderately sized system can cause it to behave in ways that cannot be explained simply by looking at its constituents. The system becomes an individual in its own right, with emergent qualities different than those of its constituents. In other words, these *complex systems* are "greater than the sum of their parts". The functioning of the brain as a whole is another thing entirely than that of the neurons that compose it. Consciousness, memory and emotion are emergent properties of the system, which are not obvious consequences of how individual neurons function. Similarly, the behavior of financial markets seems more complex than the actions of market players when observed in isolation, and traffic jams emerge not simply because there are too many cars, but because of the way drivers behave and interact.

Complex systems are abundant in biology. Biological systems are organized into a well defined hierarchy of scales, from the molecular to the ecological. At each scale we find a complex system of numerous agents from which the agents of the next scale emerge, and so from molecular agents we get cells, from cells we get tissues, from tissues we get organs, etc. A hallmark of complex systems is that the behavior of the system as a whole is not

¹Dampier (2) noted: "Plato was a great philosopher, but in the history of experimental science he must be counted a disaster."

evident based solely on observations of its constituents. It is often not clear, for example, how a certain tissue functions as a whole based on observations of individual cells. To understand complex systems, we need models that can link emergent properties with the underlying constituents.

Approaches to modeling

In science, “model” can mean different things to people in different fields. For a physicist, a model is a set of mathematical laws that describes the essential components of a system and can accurately predict its behavior. In biology, more often than not “model” is meant as “model organism” – a particular species studied in detail by many researchers and thought to be representative, at least physiologically, of a wide range of other related species. But biologists also work with abstract models, which we might call logical or functional models. These are the models that occupy the discussion sections of biological research and review papers. While physical models are expressed almost entirely in mathematics, the logical models are expressed in plain language. They consist of a narrative on how the system is thought to work and are usually accompanied by an illustration. These models don’t describe quantitative relationships, but discrete cause-and-effect relationships between actors in the system. Calculus and algebra are not really suitable for expressing the logical models used by cell biologists and, consequently, theoretical work in cell biology has not been as successful as it has in other fields.

The different approaches to modeling found in cell biology and physics have consequences for how the two fields operate. Physicists can be roughly divided into two camps: theoreticians and experimentalists. Whereas experimentalists produce interesting empirical observations through experiments, theoreticians are chiefly concerned with developing theoretical models that can explain the observations in terms of how the system is composed and how its elements interact. These models not only satisfy intellectual curiosity, explaining why experiments turned out the way they did, but any decent model should also be predictive, that is, able to foresee the outcome of experiments on the system (within some limited margin of error). When we have a predictive model of a system we can truly say we understand how it works, since the essential components and their interactions are expressed explicitly in the model, without the confounding detail of the real system.

For a model to be useful in physics it must be quantitative, so that its predictions can be measured against the real system and judged accordingly. Physics is only concerned with measurable quantities found in Nature, and so the accuracy of a model can only be judged by the magnitude of its errors. Temperature described as “hot” or “cold”, or speed as “fast” or “slow”, is meaningless in physics. Since physical models need to be quantitative, they are defined using mathematics, which is the only language known to us for describing and reasoning with quantitative relationships. Theoretical physicists can use the rich mathematical toolset to work the model and predict previously unknown phenomena. Often, all possible behaviors of the system can be extracted from the equations, or at least specific questions can be answered, before performing actual experiments. The model is validated by confirming these predictions experimentally. If some day an experiment is performed that contradicts the model, the theoreticians will refine it and the cycle is repeated. With each iteration of this cycle the knowledge of the system increases. Not only is the system better understood in terms of how it works, but also more accurate predictions can be made.

Contrary to physics, biology has a relatively small theoretical component, and one often has the feeling that there is no synergy between experimental and theoretical approaches. Yet, biologists are not content just documenting their experiments, they also seek to understand the systems they work with in terms of an theoretical models, albeit different than those found in physics. Biologists take experimental data, analyze it and distill it into a compelling narrative just complex enough to explain all the experimental results. This narrative, along with the accompanying figure, is the kind of theoretical model most biologists work with. They are typically logical models, proven or disproven by making binary predictions of the sort “if I knock out gene A, then gene B is not expressed” or “if I sequester molecule X, then the cell will lose motility”. It is uncommon for a biologist to predict “if expression of A is reduced 50%, then expression of B is reduced to 25%”. This may be because performing measurements on the cellular systems is very difficult, or perhaps because it is already difficult enough to try to understand “how?”, never mind “how much?”.

Although these logical models are not specified in a formal language such as mathematics, testable predictions can still be extracted from the model by deduction. Again, predictions are usually qualitative because they are concerned with functional aspects of the system, such as “Is gene A expressed

or not?", "Did the cells survive or not?", "Did the cells remain motile or not?". However, this approach scales poorly since biologists have little theoretical framework for making predictions other than what deductions they can work out in their minds. This is where the theoretician should come in. The theoretician's job is to make sense of experimental data by constructing a theoretical model that can explain all the observations of a given system. Models not only explain why certain phenomena occur, they also serve as a way of summarizing knowledge. A good model compresses and organizes large amounts of experimental data into a succinct description of the system. The ability of biologists to distill and compress information into logical models has not kept pace with the ever increasing amount of experimental data that is produced. Whereas the virtuous cycle between theory and experiment in physics has driven the field to the brink of a theory-of-everything, biology seems to be exploding with unanswered questions and overflowing with data.

A new modeling paradigm

The ability to extract answers from a model without performing full-blown experiments is one of the most powerful aspects of theory, and that ability is possible only by the use of formal languages equipped with tools that are suitable for exploring the structure and implications of the model. Key to any model is the language in which it is expressed. After all, what state would physics be in if physicists could only express their work in plain English? And how fast could biology progress if it were possible to express models in a formal language, such as mathematics?

The use of mathematics in biology for modeling and theory has been successful in some fields, such as in ecology and molecular biology. However, cell and microbiology have seen little theoretical work compared to the reams of experimental studies that have been published. Biologists in these fields do use models for understanding their results, but these models are specified informally in plain English, which gives us no mechanism to test the model apart from conducting more experiments. Why can't all of biology be described using mathematics, similar to physics?

Traditional mathematical models in biology are typically stated as differential equations in time (ordinary differential equations ODEs) or in time and space (partial differential equations, PDEs). These equations are constructed

from basic principles, taken as self evident, which are composed, substituted and transformed analytically to obtain the final equations. A key step in deriving differential equation models is taking the system to its continuous limit, in which it is assumed that the constituents are infinitesimally small and smoothly distributed over the domain. This works well when the constituents are essentially identical, numerous and whose individual shapes, quirks and behaviors (if any) do not have singular effects and can be abstracted away. In other words, the constituents can be seen as a smooth mass, rather than as a collection of individuals. Again, in biochemistry and ecology these assumptions often work well, and models based on the law of mass action and reaction-diffusion are used widely. The constituents (molecules, organisms) are sufficiently small, numerous, simple in function and smoothly distributed. Differential equation models have also found use in modeling cellular systems in which the cells are small compared to the domain size and can be reduced to infinitesimal points, and all behave in the same way. For example, the heart has been modeled with partial differential equations in which the tissue is modeled as a continuous excitable medium and cells are infinitesimally small (3). Yet there are many other systems in which cells cannot be reduced to points and abstracted away, but in which the individuality and complexity of cells is fundamental. For example, in developing embryos, the cells are not small compared to the embryo itself, and furthermore the shape and the behavior of a small subset of cells has a large impact on the evolution of the system. During the morphogenesis of *Dictyostelium discoideum*, the presence of a few dispersed auto-cycling cells is fundamental to the formation and movement of the crawling slug. To understand blood flow in small capillaries, the shape and cohesiveness of red blood cells must be taken into account. The roundworm *Caenorhabditis elegans* has only a thousand cells, yet is a complex multicellular organism with multiple organs and a nervous system.

There is another factor that sets biology apart and which makes quantitative reasoning more difficult than in physics. At its core, physics is concerned with elementary particles and matter composed of elementary particles. These particles slavishly follow a set of quantitative laws that makes their behavior predictable, even if only statistically. Particles of the same species are all identical and interchangeable, they have no identity, and they behave the same way according to universal quantitative laws. Mathematical modeling in biology works best when the biological system can

be approximated as such a system of elementary particles whose behavior and interactions can be specified as equations, such as algae in a population dynamics model or metabolites in a metabolic network model. Still, all biological systems are based on the same elemental matter as everything else, so why can't physics and chemistry fully explain biology? A significant difference between biology and more fundamental sciences is that in biology elementary particles combine to form "complex agents" – machines that perform tasks – and the behavior of these agents is often difficult to capture mathematically.

Loosely, anything that can be viewed as following a procedure, typically for a specific role in some grander scheme, can be interpreted as a complex agent. Biological systems are full of complex agents, from the macroscopic to the microscopic. Cells are powered by discrete complex agents such as molecular motors, tasked with carrying payloads, as well as exerting force on the cytoskeleton; transcription machinery to copy DNA, ribosomes to translate RNA into protein, molecular pumps to regulate concentrations. Agents can operate on other agents, as well as on passive structures such as membranes and the cytoskeleton. From these myriad agents and their interactions, larger scale agents emerge – cells – with cellular agents of different types performing specific roles in a larger system, giving rise to yet larger complex agents. Cells of different types form tissues and organs, again each with a specific role. At every scale in biology we find complex agents whose interactions between themselves and their environment give rise to a new level of complexity at a higher level.

Complex agents have both a physical aspect and a logical aspect. The physical aspect of complex agents is concerned with the mechanics of the agent, how the agent is implemented physically. This aspect is captured using physical models of e.g. protein folding, chemical reactions, viscoelastic materials, etc. These models explain to us exactly how the agent works on a nuts-and-bolts level, but do not really describe what the agent does. On the contrary, the logical aspect of an agent relates to the procedure the agent performs and its role in the overall system. Therefore, we can describe a cell mathematically as a collection of quantities such as concentrations, stiffnesses, viscosities etc. We can also view a cell as a complex agent and describe it algorithmically in terms of its "program". Whereas the logical aspect describes the cell program, the physical aspect describes how that program is implemented as a machine.

The difficulty of calculus to describe the logical aspect of agents is perhaps one reason why theoretical work in cell biology has met with limited success. It quickly becomes awkward to express the tasks an agent performs. Although in principle the mechanics of each agent could be modeled at the molecular level, a cell is clearly too complex to be understood as a purely mechanical system. With the computing revolution, however, an alternative form of studying biology theoretically has emerged. Theorists can now describe an agent's behavior using computer programming languages. These formal languages were designed to describe procedures and are well suited for describing the tasks agents perform. Furthermore, computer languages also have associated toolsets that, although different in nature from those of mathematics, also enable the theorist to answer questions through the model in the form of running computer simulations. By converting his/her plain English models into a so-called "agent-based model", a Biologist can now work in the abstract with both physical and logical models of the system. By running simulations, all the logical and physical consequences of the model play out and making predictions is no longer limited to the (often "linear") intuition of the researcher. An elegant example of this in practice is Odell and Foe's recent model for mitosis (4). The authors used physical models for the purely physical parts of the system, such as microtubules, and agent-based models for the active molecular agents, such as molecular motors. The agents are not modeled physically as molecules, but as micromachines following a procedure. The agents interact with the physical structures as well as with other agents. The result is a model that captures mitosis as the outcome of both physical processes and logical processes.

Agent-based models are computational in nature and so are specified in terms of an algorithm. Traditional mathematical models, such as diffusion of a morphogen or the equations for fluid flow, may be pieces of the model, but the model as a whole may not be amenable to calculus. The models assume that the system is composed of discrete interacting elements often with some internal state. The agents are programmed to change their state as a function of time and the interactions with other agents. A simulation consists of initializing the model domain with anywhere between hundreds up to millions of these agents, then stepping in time. In turn, a time step consists of updating the state of each agent according to its interactions at that point in time, and if the agents are motile, moving the agent a small step in the direction dictated by a combination of the internal state and the

external forces applied to it. Over time, an emergent pattern will (hopefully) arise from these minute interactions. By conducting many simulations for different system parameters, we can explore how the system behaves under different conditions. This is a major difference between computational and mathematical models. Predictions from computational models can only be made through induction from running simulations, whereas predictions from purely mathematical models can be deduced from the equations, and may not require running any simulations at all. Computational models may therefore seem to be a more brute force approach, but for most complex systems there is no other way. The collective behavior of individuals in those systems is too complex and chaotic. The emergent properties of the system can only be generated from the individuals themselves.

In this thesis we explore a particular kind of agent-based model called *cell-based models* (CBMs) that take biological cells as the agents. Cell-based models have become an important theoretical tool for understanding and predicting the behavior of cellular scale systems. Their natural representation of cells has made them increasingly popular in cellular and microbiology. While they do not constitute a theory per se, they are an important theoretical tool that can help drive experiments by making useful predictions and increase our understanding of complex systems. Like all models, cell-based models allow a researcher to pare down a cellular system to its essential working parts, to see past the complexity fog and reveal how the system really works in a clear and compelling manner. Cell-based models conjugate well with current experimental techniques, such as high resolution fluorescent microscopy and electron microscopy, which can be used to gather cell-level data on the system that can be directly compared to a cell-based model simulation. Finally, the modeler has absolute and complete control over every detail of the model system and also has complete visibility of simulation results. Simulations run autonomously and are extremely cheap compared to running actual experiments. A cell-based model can therefore act as a virtual lab, where experimentalists and theoreticians alike can play with the system and generate targeted ideas for new experiments.

In Chapters 2 and 3 we will introduce various techniques for designing cell-based models. Focus is given to general techniques for modeling large numbers of motile cells that interact through physical contact and collisions, amongst other processes. We focus purely on modeling the mechanics of cellular systems (cell geometry, deformation, motility, cell-cell adhesion, etc.)

since this is a common starting point for all cell-based models. Mechanistic models for the internal workings of the cell (metabolism, homeostasis, genetic regulation, etc.) are desirable, but as a first approach phenomenological models of cell behavior based on experimental observations are sufficient to observe emergent properties. One of the challenges of cell-based modeling is to incorporate mechanistic models of cell regulation, such that, for example, morphogenesis can be controlled by the parameters of a genetic regulatory model. Tying multiple space and time scales into so called *multi-scale* models is a huge challenge and has become a very active field of research. In Chapters 4, 5 and 6 we present novel cell-based models for two totally disparate systems, the gastrulation of the starlet sea anemone and pattern-formation in cultures of filamentous cyanobacteria, but with a common framework based on simple mechanical elements. The models illustrate the flexibility of the cell-based modeling paradigm, as well as its ability to perform virtual experiments and generate new hypotheses. We conclude with a brief discussion on the future of cell-based models.