

# Mathematics and biology

**AUTHOR:**

Jan-Hendrik S. Hofmeyr<sup>1</sup>

**AFFILIATION:**

<sup>1</sup>Centre for Complex Systems in Transition, Stellenbosch University, Stellenbosch, South Africa

**CORRESPONDENCE TO:**

Jan-Hendrik Hofmeyr

**EMAIL:**

jhsh@sun.ac.za

**KEYWORDS:**

modelling relation; simulation; metric biology; relational biology

**HOW TO CITE:**

Hofmeyr J-HS. Mathematics and biology. *S Afr J Sci.* 2017;113(3/4), Art. #a0203, 3 pages. <http://dx.doi.org/10.17159/sajs.2017/a0203>

I am a biochemist turned theoretical systems biologist, who, by my definition, is someone who attempts to explain the emergence of systemic functional properties of the living cell that result from the interactions of its components. Mathematical modelling and computer simulation is the bread and butter of my everyday work. However, my formal training in mathematics was 1 year of ‘mathematics for biologists’, which certainly did not equip me with the tools I needed when, early on in my research career, I embarked on a theoretical and computational analysis of metabolic networks. To study the time-dependent and steady-state network behaviour I had to master the theory of non-linear differential equations and their numerical analysis; the framework of metabolic control analysis required me to learn linear algebra and the theory of matrices, and I had to teach myself computer programming. Lately I have become fascinated with self-fabrication and the functional organisation of cell processes and found that I needed to know about abstract algebra, category theory and the theory of formal systems and formal languages. Although this journey through mathematics has been exciting and rewarding, I do wish that I had had the appropriate training as part of my undergraduate studies. Nevertheless, as a biologist who uses mathematics to explore biological phenomena (in contrast to a biomathematician, who mines biology for interesting mathematical problems), I have learnt a lot about the relationship between biology and mathematics – a relationship which through its entire history has always been an uneasy one. Why this is so and why this may be changing are the questions I wish to explore here.

Whereas physics and mathematics have cultivated a rich synergy from which they have both clearly benefitted, most biologists have regarded mathematics to be of marginal or no interest to their work or, at best, to be useful for everyday numerical and statistical treatments of experimental data but nothing more. How many of us who teach in some field of biology have not encountered students and colleagues that readily concede that they chose to study biology because they feared or disliked mathematics? Often these sentiments are expressed not as a resistance towards mathematics per se, but as against theory in general, as if it were possible to do theory-free experiments. This sad state of affairs is expressed in our tertiary-level biology curricula in which mathematics is usually relegated to a one-semester or two-semester first-year course, after which it is never to be heard of again. Even mathematics departments usually make short shrift of biomathematics; to the best of my knowledge Stellenbosch University is at present the only South African university to offer a biomathematics degree programme.

Biologists’ disregard of mathematics is even more unfortunate in the light of the essential role that mathematical formalisms fulfil in all of science. Science rests on the bedrock of natural law – the twin beliefs that there are causal entailments in the natural phenomena that we study and that they can be imaged in inferential entailments in our formal systems. This act of imaging causality in terms of inference is what we call ‘modelling’. When the theoretical biologist Robert Rosen wrote ‘I have been, and remain, entirely committed to the idea that modelling is the essence of science and the habitat of all epistemology’<sup>1</sup>, he emphasised his view that science is the art of establishing what he called ‘modelling relations’ between the natural world and the world of our formalisms. This view is one I fully share and one that has afforded me deep insight into what I do as a scientist.

What is a modelling relation? According to Rosen<sup>2,3</sup>, a modelling relation aims to establish congruence between two systems – more specifically, between the elements of each system and between the entailment structures of each system. By accomplishing both these aspects, the orderly nature of one system is made to correspond to another system to the extent that the two systems have a degree of correspondence. Let me unpack this definition. A modelling relation establishes a relation between that part of the natural world we choose as our object of study (let us call it ‘natural system N’) and a ‘formal system F’ – the inferential entailment structure  $i$  of which mimics the causal entailment structure  $c$  in the natural system (Figure 1). This is done by choosing a set of observables that we believe characterises the natural system, and then constructing an *encoding* dictionary  $\varepsilon$  that maps observables in the natural system to input variables in the formal system. The inferential rules for manipulating the entities in  $F$  are then supposed to image the causal relations in  $N$ . What we hope to achieve is to bring the entailment structures in the two systems into alignment, so that, given an input set, the result of the inferential process in  $F$  can be *decoded* via mapping  $\delta$  into the  $N$  to make a prediction about its behaviour. When our prediction matches the behaviour of the natural system, i.e. when the result of natural processes  $c$  matches the result of tracing the arrows through  $\varepsilon \rightarrow i \rightarrow \delta$  we can claim that  $F$  is a model of a particular aspect of  $N$  and that that aspect of  $N$  is a realisation of  $F$ . If this is achieved, the modelling relationship is said to satisfy a commutativity condition. In the language of category theory, a modelling relation establishes a functorial correspondence between  $N$  and  $F$ , i.e. objects in  $N$  map to objects in  $F$  and mappings in  $N$  map to mappings in  $F$  (Figure 2a).

The modelling relation also helps to clarify the difference between a ‘model’ and a ‘simulation’. The mapping  $g$  in Figure 2b is a simulation of  $f$  instead of a model of  $f$ ; it differs from mapping  $\varepsilon(f)$  in Figure 2a in that it is not an encoding of a mapping  $f$  in  $N$  into one in  $F$ . This implies that  $g$  does not model the causal entailment structure of  $f$ . More generally,  $i$  in Figure 1 is a simulation of  $c$  if no attempt has been made to bring inferential entailment in  $F$  into correspondence with causal entailment in  $N$ . A typical instance of a simulation is when a set of input-output data is fitted with a polynomial to generate the input-output curve. Such a simulation provides no insight into the causal structure of the system that generated the data.

With this understanding of what modelling entails, we can ask which formalisms mathematical biology has employed to model biological systems and phenomena. For biology the 20th century was very much the age of analysis in which the quantitative approach ruled. I call this the age of metric biology. In metric biology, all observables

are quantifiable state variables and the concept of a succession of state transitions forms the basis of the description of the time-dependent behaviour of biological systems. Mathematical biology took its cue from mathematical physics; the vast majority of mathematical models of biological processes were and still are cast in the language of dynamical systems theory, using deterministic non-linear ordinary differential equations, partial differential equations and difference equations. The last few decades have also seen the increasing incorporation of stochasticity into biological models, with the use of stochastic differential equations, the master equation and Markov processes. Hybrid models mix continuous/discrete and deterministic/stochastic differential equations.

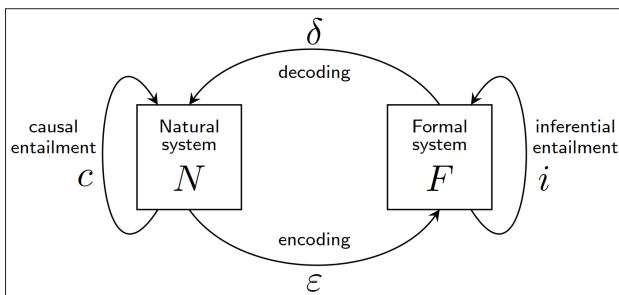


Figure 1: The modelling relation between a natural system  $N$  and a formal system  $F$ .

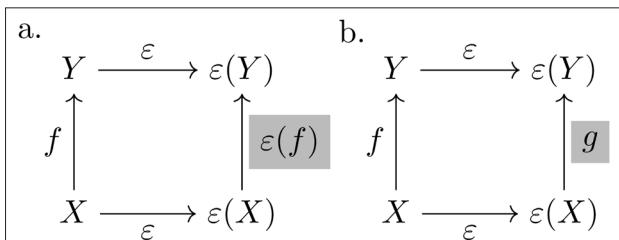


Figure 2: The difference between mappings between a natural system  $N$  and a formal system  $F$  in (a) model and (b) simulation.

The major problem with the formal systems used to model metric biology is that they are impoverished in the type of causal entailment that they can image. Almost everything of interest in a system of ordinary differential equations is unentailed from within the system and must be posited from outside; all that remains internally is the transition from one state to the next.<sup>3</sup> If one asks the Aristotelean question ‘why state  $x_{n+1}$ ?’ the only answer from within the system is ‘because state  $x_n$ ’, which amounts to an answer in terms of material causation. Efficient causation (the mappings) and formal causation (initial conditions and parameters) are posited from outside and are immutable with regard to internal changes. Final causation (function) does not exist in this picture. But, in living systems, all these causes can be produced or altered from within the system. Consider, for example, an enzyme that catalyses a biochemical reaction and is therefore its efficient cause. In the cell, this enzyme is itself produced by ribosomal protein synthesis (its efficient cause) informed by the sequence of nucleotides in its messenger RNA (its formal cause); the production of the enzyme is the final cause (the biological function) of ribosomal protein synthesis. From a mathematical point of view the enzyme is simultaneously an operator and also an object in the range of another operator. This *functional entailment* is a hallmark of living systems.<sup>3</sup> In the sense discussed above, systems of differential equations simulate the behaviour of biological systems; they are not models because they do not capture the rich causal entailment structure of the biological system. Of course this does not imply that these simulations are not useful. If that were so I would have to discard three decades of my own systems biological research in which the behaviour of enzyme-catalysed networks are all described in terms of differential equations and through which I have learnt a lot about how these systems are controlled and regulated. Their main weakness is that they cannot capture the essence of living systems.

Nowhere is this realisation that we need new mathematics to model life more evident than in the research career of the father of mathematical biology. Nicholas Rashevsky originally trained as a theoretical physicist but turned to the study of biological phenomena after emigrating to the USA from the Ukraine. He started the school of mathematical biophysics at the University of Chicago in the 1930s, and in 1939 he founded the *Bulletin of Mathematical Biophysics*, the international journal of mathematical biology, re-titled in 1973 to the *Bulletin of Mathematical Biology*. Although there were of course earlier landmarks in the development of 20th century mathematical biology, such as D'Arcy Thompson's *On Growth and Form* and Alfred Lotka and Vito Volterra's predator-prey model, it is generally accepted that Rashevsky's work forms the bedrock of mathematical and theoretical biology. Taking his cue from mathematical physics, in his early work he used its methods to make fundamental contributions to mathematical biology: he studied bifurcation phenomena; a Boolean version of his ‘two-factor theory’ for excitable elements led in the hands of Walter Pitts and Warren McCullough to the development of neural networks, and was also a forerunner of Hodgkin and Huxley's model for the propagation of action potentials in neurons.

In the 1950s, Rashevsky began worrying that the mathematical tools that he had employed so successfully could not answer the most basic question of biology: What is life? In his own words:

*All the theories mentioned above deal with separate biological phenomena. There is no record of a successful mathematical theory which would treat the integrated activities of the organism as a whole... And yet this integrated activity of the organism is probably the most essential manifestation of life... The fundamental manifestations of life mentioned above drop out from all our theories of mathematical biology... As we have seen, a direct application of the physical principles, used in the mathematical models of biological phenomena, for the purpose of building a theory of life as an aggregate of individual cells, is not likely to be fruitful. We must look for a principle which connects the different physical phenomena involved and expresses the biological unity of the organism and of the organic world as a whole.<sup>4</sup>*

It was this realisation that drove Rashevsky away from simulating state transitions to trying to model the functional organisation of processes in the organism. The emphasis shifted from the components of biological systems to the relations between them, and so was born what is now called ‘relational biology’. He used topology, set theory (‘organismic sets’) and propositional logic as his new tools. Relational biology was more fully developed by his PhD student, Robert Rosen, who used category theory to formulate his so-called metabolism-repair systems, which captured the circular causal structure of cell processes that he termed closed to efficient causation. Rosen is now recognised as one of the most important theoretical biologists of the 20th century. His development of relational biology can be said to have heralded the advent of the age of synthesis, of an integrative view in which we try to glue Humpty Dumpty together again. If the age of analysis was characterised by the unspoken motto ‘divide and conquer’, then perhaps the age of synthesis aims to ‘integrate and rule’. Today relational biology uses a rich array of mathematical tools: category theory, graph theory, network theory, automata theory, formal systems, to name the most important.

The perceptive reader may complain that I have not commented on the role of mathematics in fields such as evolutionary biology, population genetics, ecology and epidemiology. The reason is that, besides the crucial role of statistics in these disciplines, the mainstay has been population dynamics, which uses the mathematical tools of metric biology and therefore brings nothing new to the table.

As to the future of mathematical biology, Avner Friedman, who directed the Mathematical Biosciences Institute from 2002 to 2008, made this prediction:

*Viewing the present trends in mathematical biology, I believe that the coming decade will demonstrate very clearly that mathematics is the future frontier of biology and biology is the future frontier of mathematics.<sup>5</sup>*

I agree with Friedman: my reading is that biology in general has finally embraced mathematics as an indispensable partner, and that biology enriches and will continue to enrich mathematics. It is now up to us to ensure that our biology and mathematics curricula are commensurate with this vision.

## References

1. Rosen R. Essays on life itself. New York: Columbia University Press; 2000.

2. Rosen R. Anticipatory systems: Philosophical, mathematical and methodological foundations. 2nd ed. New York: Springer; 2012. <https://doi.org/10.1007/978-1-4614-1269-4>
3. Rosen R. Life Itself: A comprehensive inquiry into the nature, origin, and fabrication of life. New York: Columbia University Press; 1991.
4. Rashevsky N. Topology and life: In search of general mathematical principles in biology and sociology. Bull Math Biophys. 1954;6:317–348. <https://doi.org/10.1007/BF02484495>
5. Friedman A. What is mathematical biology and how useful is it? Notices of the AMS. 2010;August:851–857.

