Science AAAS

Introductory Science and Mathematics Education for 21st-Century Biologists

William Bialek, *et al.*

Science **303**, 788 (2004);
DOI: 10.1126/science.1095480

Introductory Science and Mathematics Education for 21st-Century Biologists

William Bialek[1,3] and David Botstein[2,3]*

Galileo wrote that "the book of nature is written in the language of mathematics"; his quantitative approach to understanding the natural world arguably marks the beginning of modern science. Nearly 400 years later, the fragmented teaching of science in our universities still leaves biology outside the quantitative and mathematical culture that has come to define the physical sciences and engineering. This strikes us as particularly inopportune at a time when opportunities for quantitative thinking about biological systems are exploding. We propose that a way out of this dilemma is a unified introductory science curriculum that fully incorporates mathematics and quantitative thinking.

Dramatic advances in biological understanding, coupled with equally dramatic advances in experimental techniques and computational analyses, are transforming the science of biology. The emergence of new frontiers of research in functional genomics, molecular evolution, intracellular and dynamic imaging, systems neuroscience, complex diseases, and the system-level integration of signal-transduction and regulatory mechanisms require an ever-larger fraction of biologists to confront deeply quantitative issues that connect to ideas from the more mathematical sciences. At the same time, increasing numbers of physical scientists and engineers are recognizing that exciting frontiers of their own disciplines lie in the study of biological phenomena. Characteristic of this new intellectual landscape is the need for strong interaction across traditional disciplinary boundaries.

Biology curricula at our colleges and universities have not kept pace with these developments (*1*). Even though most biology students take several years of prerequisite courses in mathematics and the physical sciences, these students have too little education and experience in quantitative thinking and computation to prepare them to participate in the new world of quantitative biology. At the same time, advanced physical science students who become interested in biological phenomena can find it surprisingly difficult to master the complex and apparently unconnected information that is the working knowledge of every biologist. These barriers to communication between disciplines become significant early, such that advanced undergraduates in the different disciplines already speak noticeably different languages. Effective solutions to this problem will require collaboration between university faculty in biology and traditionally mathematical sciences. We believe that the needs of biology should provide a stimulus to re-examine the teaching of all science.

[1]Department of Physics, [2]Department of Molecular Biology, [3]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA.

*To whom correspondence should be addressed. E-mail: botstein@princeton.edu

Quantitative Courses as Prerequisites

During the last century, the educational path leading to professional degrees in the biological and biomedical sciences (i.e., Ph.D., Sc.D., or M.D.) in the United States became rather standard. Undergraduates interested in biology and medicine begin their studies with a set of "prerequisite" courses, typically one or two semesters each of mathematics and physics and two to four semesters of chemistry. For most biologists and physicians, this early college experience, most of it preceding serious enagagement with biology itself, is the end of their education in mathematics and the physical sciences.

For reasons of history, this prerequisite mathematics, physics, and chemistry education is delivered by departments as a service to students who take them because it is required for a degree in biology or for entry into medical school. Many of the students taking these courses have no real enthusiasm for mathematics, physics, or chemistry per se and perceive these courses simply as obstacles to be overcome on the way to a career in biology or medicine. Not surprisingly, the faculty who teach these service courses are ill prepared to make connections between what is presented in the prerequisites and what is exciting in the biological sciences. Almost without exception, larger universities teach students hoping to major in mathematics, physics, chemistry, or engineering separately from those hoping to be biologists or physicians. The difference in sophistication (and difficulty) of the quantitative content of these separate tracks can be startling.

These traditions have resulted in a deep bifurcation in culture and quantitative competence among the scientific disciplines. On one branch are mathematics, the physical sciences, and engineering. Scientists educated along this branch achieve a high level of quantitative expertise: They generally have some mastery over and comfort with not only multivariate calculus and differential equations, but also linear algebra, Fourier analysis, probability, and statistics. All scientists in these areas are expected to be able to program as well as to use computers themselves. Projects of any size, whether theoretical or experimental, require custom software, for example, to acquire and analyze data and to carry out simulations. Beyond textbook knowledge of mathematical and computational methods, quantitative thinking is the daily essence of the science to which this educational path leads, and this is expressed in a rich interplay of theory, experiment, and computation.

On the other branch are biology and medicine. With significant exceptions (e.g., population genetics, structural biology, and some areas of neuroscience), biologists today rarely achieve mathematical competence beyond elementary calculus and maybe a few statistical formulae. Although everybody uses a computer, biologists rarely use anything but standard commercial software. Virtually all biologists today must use some sophisticated programs (e.g., sequence comparison at the National Center for Biotechnology Information's Web site), yet distressingly few academic biologists feel comfortable teaching the underlying principles to their students, and fewer are able to program even a rudimentary software implementation of such an algorithm themselves. Most biologists require consultations with biostatisticians in order to do anything but the simplest statistics, and all too often mathematical or statistical analysis in published biological papers is inadequate or omitted entirely.

The cultures of students following the two paths, not surprisingly, are also different. Whereas the students (and their teachers) on the physical science branch are focused on principles and reasoning as the goal of their education, students (and teachers) on the biology-medicine branch find themselves focused more on mastering huge arrays of facts. Although this characterization is partly a stereotype and good teaching can help bridge these cultures, undergraduates are strongly influenced by these ideas.

Toward a New Level of Understanding

The reader may well ask whether the bifurcation of scientific cultures is really a problem, especially in view of the much-heralded success of the biological sciences over the last half century, of which the sequencing of the human and other genomes is emblematic. After all, this success was achieved by scientists in different disciplines working together, each mostly educated along one or the other of the educational branches described above. Why not continue as before?

Our answer is that the basic nature and goals of biological research are changing fundamentally. In the past, biological processes and the underlying genes, proteins, other molecules, and environmental factors were of necessity studied one by one in relative isolation. In contrast, today we are already no longer satisfied with studies or answers that do not place each of these in a larger context. We now know that there are tens of thousands of genes encoded in the genomes and that simple perturbations, such as a change in nutrition or a heat shock, alter the expression of thousands of them. Similarly, in the past regulatory systems were of necessity studied in a limited way, resulting in largely intuitive, one-bit explanations (e.g., genes are turned on or off, proteins are inhibited or not); today we cannot and should not be satisfied with explanation of phenomena that are not fully quantitative. We know from experience that boxes and arrows or even more formal wiring diagrams are not sufficient to specify the function of a network. These architectural descriptions must be completed by models of the underlying dynamics. New goals are in sight, namely robust mathematical models and computer simulations that faithfully predict behavior of entire biological systems. One might even hold out hope for the discovery of theoretical principles that transcend detailed models and unify our understanding of seemingly different systems. Many of these ideas were articulated 20 years ago in the context of neuroscience; with the emergence of much wider possibilities for system-level analysis, they have migrated into many diverse areas of biology.

In order to participate fully in the research of the future, it will be essential for scientists to be conversant not only with the language of biology but also with the languages of mathematics, computation, and the physical sciences. It is important to recognize that the problem cannot be solved by specifying minimal mathematical expertise for future biologists and assigning our colleagues in the mathematics department the task of inculcating this expertise in our students. Although physics students, for example, often take many mathematics courses, the fundamental idea that mathematics describes the natural world is something that is taught in the physics department. Understanding how to reason in the language of mathematical symbols is essential, but one must go further to appreciate that these symbols actually stand for the variables of the natural world, that these variables can be measured quantitatively in the field or in the laboratory, and hence that abstract mathematical relations among symbols can become concrete relations among the results of experiments. There is an enormous challenge in raising a generation of scientists who are equally at home with this quantitative mode of thought and with the complexities of real organisms.

Progress requires unflinching honesty about the depth of the problem. Forty years ago, Snow wrote eloquently of the difficulties raised by the emergence of two cultures, one scientific and one humanistic. Today we have two cultures within science itself, one mathematical and the other not. If biology is to assimilate into the world of quantitative science, biologists and nonbiologists alike will need a different kind of education than we provide today (*2*).

An Integrated Introductory Quantitative Science Curriculum

What is to be done? The answers may not be the same for all students. We propose here an approach aimed specifically at students interested in a research career in the biological sciences, whether in academia, industry, or medicine. For them, we advocate an integrated introductory curriculum in which mathematics, the physical sciences, and biology are introduced together. Instead of separate prerequisite courses in mathematics, physics, chemistry, and computation, the fundamental ideas of each of these disciplines should be introduced at a high level of sophistication in context with relevant biological problems. Indeed, we think that the choices made in such a curriculum should be motivated in no small part by connections with biology. The emphasis on integration is particularly important at the introductory (i.e., first year or two of college) level, because a large part of the goal is to show the students how each discipline contributes to understanding the phenomena of life, how these phenomena illustrate and reinforce our quantitative understanding of phenomena in the inanimate world, and how the boundaries between disciplines are becoming arbitrary and irrelevant. Integration will allow students to learn the languages of the different disciplines in context.

The goal should be students with a mastery of a broad set of skills and the confidence to approach biological phenomena quantitatively. Such an integrated science curriculum, when done well, should allow students to continue their education in any area of science. Scientists educated in this way, regardless of their ultimate professional specialty, would share a common scientific language, facilitating both cross-disciplinary understanding and collaboration.

There are many challenges in designing such a program. Clearly it cannot and should not be the sum of everything in introductory courses in mathematics, physics, chemistry, computer science, and biology, including the history of each idea. It must be reasonable in length yet provide a serious and useful introduction to all of these disciplines in context with each other. The primary challenge in designing an integrated curriculum, therefore, is to identify each of the individual intellectual concepts, methods, and facts that are fundamental and generalizable, independent of their history. Whereas advanced courses in the individual disciplines can reinforce material treated briefly in an introductory course, the first-year science curriculum offers a special opportunity to convey the intellectual point of view and the quantitative attitude toward the natural world that is embodied by Galileo's dictum.

Any attempt to create a multidisciplinary curriculum leads to difficult questions about what will be left out. Seldom emphasized, however, are the opportunities for synergy. Teaching mathematical methods in the context of the natural science problems that motivate their development is an obvious example and already happens in most physics curricula; for example, in an integrated curriculum, calculus would be taught together with basic physics. Making the most of these opportunities cannot help but facilitate both the teaching and learning of all the relevant disciplines. Thermodynamics, kinetic theory, and the rudiments of statistical physics appear in different guises in introductory physics and chemistry courses, and even introductory biology courses make reference to binding constants and chemical potential. Most introductory physics courses include some approach to "modern physics," and introductory chemistry courses provide at least the outlines of quantum theory to describe electrons, orbitals, and chemical bonding. In these cases, unification of the curricula would convey a clearer picture both of the underlying principles and the diversity of their applications.

More subtly, but perhaps most crucially, there are commonalities of the mathematical structures that summarize our understanding of seemingly disparate topics. Classical mechanics presents a model of the world's dynamics based (in the introductory account) on simple differential equations, but chemical kinetics and even the dynamics of populations provide models of the same general form. Although the different systems have important special features (e.g., the conservation laws), surely we would like to communicate the more general idea that dynamics are described by differential equations and encourage students to discover the applicability of this approach to the dynamics of more complex biological systems through well-designed laboratory exercises. In a similar spirit, statistical physics and kinetic theory provide probabilistic models of the world, but Mendelian genetics is also a probabilistic model and an understanding of probability is at the heart of all practical data analysis.

Today, not only can we integrate subjects that share common mathematical structures, we can also integrate these abstract structures with their practical implementation through computation. If the students are taught to program and to use simple algorithms and if they learn to use high-level languages (e.g., Matlab or Mathematica), they can visualize and verify for themselves the mathematical ideas and thereby become comfortable with those they find less intuitive or more abstract. In statistics, for example, it is possible to begin by applying simulation and bootstrap algorithms (e.g., for finding P values). By starting in this way, students will more easily come to appreciate parametric methods and closed-form solutions and learn to understand and to use them appropriately.

We believe that integrating mathematics, computation, and the scientific context for these ideas will allow students in an introductory course practical access to conceptual tools that are much more sophisticated than those currently taught in the standard first-year mathematics courses. Although real mastery over these ideas will require continuing reinforcement throughout the undergraduate curriculum (as is currently done for physical science students), a unified introduction can empower the students to explore ideas far beyond what is currently accessible to them.

A final and, in the context of biology, possibly the most important synergy derives from the judicious use of nonstandard examples for basic principles and methods of physics and chemistry. For example, it makes sense, in modern times, to introduce students to the idea of molecular motion and thermodynamics in solution rather than focusing only on the world of ideal gases. With affordable modern instrumentation, students can observe and record Brownian motion in a microscope, for example, and satisfy themselves quantitatively how this motion derives from invisible molecules bouncing around in the solution and even how many such molecules there must be. This hands-on approach has the advantage that the phenomena (and of course the underlying principles) are directly and obviously relevant to research in biology. In a similar vein, much of basic combinatorics, probability theory, and statistics can be presented in tandem with basic genetics, resulting a substantial saving in overall time when compared with separate courses in different departments. Again, the concurrent use of computation will provide students with tools that will serve them well in all of their scientific careers thereafter.

Our proposal for an integrated introductory education for quantitatively oriented biologists really is an experiment in a more general problem: science education in the modern world. This is a problem whose solution will require collaborations among scientists who now reside in quite different departments and cultures; enthusiastic as we are, we also are cognizant of the difficulties that will no doubt arise. On the other hand, the necessary collaborations among the faculty from several disciplines may well set a wonderful example for students.

To conclude, we believe there is a great opportunity to construct a unified, mathematically sophisticated introduction to physics and chemistry, which draws on examples from biology wherever possible. Such a course would provide a coherent introduction to quantitative thinking about the natural world, and invite all students, including biologists of the future, to partake of the grand tradition, which flows from Galileo's vision.

References and Notes

1. Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, Board on Life Sciences, Division of Earth and Life Sciences, National Research Council, *BIO 2010: Transforming Undergraduate Education for Future Research Biologists* (National Academies Press, Washington, DC, 2003).
2. C. P. Snow, *The Two Cultures and the Scientific Revolution* (Cambridge Univ. Press, New York, 1959).

VIEWPOINT

Uses and Abuses of Mathematics in Biology

Robert M. May

In the physical sciences, mathematical theory and experimental investigation have always marched together. Mathematics has been less intrusive in the life sciences, possibly because they have until recently been largely descriptive, lacking the invariance principles and fundamental natural constants of physics. Increasingly in recent decades, however, mathematics has become pervasive in biology, taking many different forms: statistics in experimental design; pattern seeking in bioinformatics; models in evolution, ecology, and epidemiology; and much else. I offer an opinionated overview of such uses—and abuses.

Darwin once wrote "I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics; for men thus endowed seem to have an extra sense." With the benefit of hindsight, we can see how much an "extra sense" could indeed have solved one of Darwin's major problems. In his day, it was thought that inheritance "blended" maternal and paternal characteristics. However, as pointed out to Darwin by the engineer Fleeming Jenkin and others, with blending inheritance it is virtually impossible to preserve the natural variation within populations that is both observed and essential to his theory of how evolution works. Mendel's observations on the particulate nature of inheritance were contemporary with Darwin, and his published work accessible to Darwin. Fisher and others have suggested that Fleeming Jenkin's fundamental and intractable objections to *The Origin of Species* could have been resolved by Darwin or one of his colleagues, if only they had grasped the mathematical significance of Mendel's results (*1*). But half a century elapsed before Hardy and Weinberg (H-W) resolved the difficulties by proving that particulate inheritance preserved variation within populations (*2*).

Today, the H-W Law stands as a kind of Newton's First Law (bodies remain in their state of rest or uniform motion in a straight line, except insofar as acted upon by external forces) for evolution: Gene frequencies in a

Zoology Department, Oxford University, Oxford OX1 3PS, UK.