Intuition and Innumeracy

Roger Brent
The Molecular Sciences Institute
Center for Genomic Education and Computation
2168 Shattuck Avenue
Berkeley, CA 94704

The editors of *Cell Biology Education* requested a perspective on the math that biologists should know. As it happens, *Science* magazine just provided such perspective, superbly, when it published on Feburary 4 a number of thoughtful articles on math and biology. One, by Bialek and Botstein (2004), describes what would amount to comprehensive curriculum reforms to better integrate mathematics into the teaching of undergraduate biology and biology into the teaching of natural sciences and math. Another, by May (2004), covers some of the dangers practicing biologists face when attempting to use computational methods without a good understanding of the underlying mathematics. Here, I take those two articles as the starting context and raise only a few additional points. I argue that, even for biologists who lack mathematical skills, replacement of what is in some cases near–functional innumeracy with a better intuition about math and computers is possible and that doing so could have significant positive effects. I also make the case that significant gains in biological understanding may come from biologists taking onboard Twentieth-Century formalisms not now part of the canon of undergraduate mathematics.

Some Biologists Today Know Almost No Mathematics

In my opinion, the teaching reforms proposed by Bialek and Botstein (2004) would be exciting and effective and could be well implemented by the end of this decade at a number of elite universities. I fear, though, that the authors may be assuming a level of intellectual competency that may not exist for all students and teachers at all universities.

Here I am relying on my own experience. Not all biology graduate students come from undergraduate universities as good as the Princeton the authors write from, and some students from those universities who did enter graduate school did not perform well in undergraduate courses in math and physics. I'm under the impression that, at least during the 1990s, some of the graduate students at elite research universities self-identified in high school as not being good with mathematics and opted in university for a graduate school rather than a premedical track precisely to avoid the necessity to take and do well in those courses. Perhaps as a consequence, it is rather easy to ask evaluative questions that elicit disappointing answers. One, which I started asking in 1994, is to ask students how many molecules of their favorite proteins they are visualizing in cell nuclei or cytoplasms. Open mouths and expressions of shock and alarm when asked that question are not a good sign and provoke follow-up questions. Not infrequently, such follow-up questions establish the point that the student is at least a little bit uncomfortable with powers of 10, let us say, the difference among 100,000, 1 million, 10 million, 10^8, and 10^{11}. A friend has suggested that this behavior is not discomfort with powers of 10 so much as discomfort at the idea of being asked to make numerical estimates. Whatever its source, I now take such discomfort as operationally defining functional innumeracy in the lab.

In the Absence of Understanding, Intuition Can Sometimes Help

Happily, I believe there is a good deal of hope for providing some level of mathematical competence, even for scientists (such as the author) who lack talent or an aptitude for the subject.

Part of the reason for optimism is the boost provided by humankind's friend, the computer. Computer aids to performance of simulations and symbolic processing of equations (one could call these Matlab and Mathematica, respectively) can endow people with insight beyond their native math aptitude. These computational prostheses are not without their downsides. May (2004) raises a number of problems with them. One is that, for the naive, there may be embedded assumptions, such as "All distributions will conform to the central limit theorem," which are in many cases not true. Working with simulations based on these assumptions can lead to results that are clearly bogus, but some biologists, lacking good intuition, won't see the falsehood.

At the higher end of math talent, there may be other losses, or at least hazards, caused by reliance on simulation rather than analytical solution. May cites as an example the fact that most meteorologists believe that the outcomes of the Navier–Stokes equations used in weather models become chaotic after many weeks. The belief that the solutions are chaotic rests on the results of simulations rather than on formal proof. At Molecular Science Institute we have a running joke (except I am not sure it is funny) that if Karl Friedrich Gauss were alive today, he would have been so dependent on simulations that he never would have invented some of the math tools we use in our current work.

But at least at the moment, the relative lack of math skills in biology hardly constitutes a crisis. That is largely because much of the math now touted has limited utility to the problems biologists are actually interested in. The broad wavefront of molecular, cellular, and developmental biology is not suffering from the lack of a widely distributed ability to write down and solve differential equations to make quantitative models of the processes under investigation. The progress of scientists exhaustively testing combinations of growth factors and feeder layers to identify conditions that keep embyronic stem cells from differentiating (rightly) is not held back by the fact that they lack understanding of ideas arising from consideration of the emergent properties of spin glasses.

At this point, the worst problem may be that without skills or intuition, individual biologists may be vulnerable to snake oil and hype. A scientist giving a research seminar might try to convince more classically trained colleagues about the rightness of the use of a measure of mutual information such at the "receiver operator characteristic" to compare sets of genomic data. This statistic, which comes from early information theory, may not be sensible in cases where there is no information transmission, for example, for different series of mass spec peaks from different samples from different patients. A startup biotech company might claim that its pathway modeling software will provide a royal road that should lead within two years to the identification of new drug targets. Absent intution born of working with models, this assertion might be difficult to evaluate.

Better Intuition Would Carry Real Benefits

Intuition of the kind I describe has been created before. Perhaps the single best example is the change in U.S. science education following the launch of Sputnik. The relentless efforts to teach science and math throughout the school curriculum helped develop a generation of citizens with a broad scientific education. I have no idea how to document the assertion I am about to make, but it seems to me quite likely that the whole field of biology benefited from budding biologists whose post-Sputnik education had given them some kind of understanding of physics and math, machines, electronics, and computers. But 1957 was a long time ago.

Today, development of broader intuition among biologists would have positive consequences for government decisions in democratic polities. The relentless increase in biological capability has already ensured that the impact of biological research on human affairs can no longer be confined to ghettos labeled "health care," "biotechnology," or "bioethics." Over the next several decades, the dominant impacts of science on human affairs will come from developments in biology (Benford, 1995). Crafting political decisions that are not counterproductive (or even outright dangerous) will require engagement by biologists. Just as, in the 1950s and 1960s, the issues of the day (nuclear weapons, long-term effects of fallout, satellite reconnaisance) called for physicists, today the issues call for biologists. But unless those biologists are conversant at some level with other disciplines, including math, they won't be able to speak sensibly to the range of scientific knowledge that will be needed for good policy.

Because biology is so important, its needs are so great, and its researchers so numerous and so relatively well funded, it is also possible that the current drive to create better mathematical intuitions among biologists may have positive impacts outside of biology. A visionary architect of human computer interaction, Brenda Laurel, has called for interfaces that allow humans to interact better with complex models (2001). In her example, Laurel refers to representations of material and energy flows that might contribute to global climate change. However, should efforts to understand the quantitative behavior of biological systems bear fruit, biologists will have the same need to interact with the world of numbers represented inside computers. Consider the calculus that underlies any first-order mathematical formalization of a dynamic biological system. The teaching of rates of change has traditionally been founded in analytical geometry, parabolas, and ballistics, the thrown rocks so beloved to the male hominid. But perhaps there are other entries into the world of first derivatives and tipping points as different from the ballistic representation as the computer desktop metaphor is from the command line. To give an extreme example, in 10 years, an alternative computer-aided way in to interacting with the first and second derivative might combine muscle tension and proprioception: effort as one rides a bicycle up a hill, diminution of effort as the slope becomes shallower, decrease of effort to zero at the crown of the hill. Success in generating alternate metaphors might broaden the net cast by science education, for example, here by providing alternative means to teach precalculus to larger numbers of girls. The benefits to biology and to broader human activities from the development of methods and interfaces to heighten intuition would be sufficiently great as to justify the engagement of effort from the best designers, artists, and architects of the virtual world.

Not All the Desired Intuition May Come from Classical Math

It's also worth noting that gains might not come from math as it developed up to the nineteenth century, but from other formalisms developed by mathematicians and computer scientists during the last century to deal with the human-created world. Although the systematic analysis of these formalisms is relatively new (Simon, 1981), it is possible to ascribe to them certain common features; at least to me, they all seem messier than classical math, less Cartesian, precisely as if they have not had the benefit of the logical rigor that comes from centuries of theorem proving. At least three of these may hold promise: "control theory" (e.g., Doyle *et al*. 1992), "qualitative physics" (e.g., Kluwer and de Kleer [1990], whose Aristoliean picture of the world of physics intuitively seems a good match to natural language descriptions of molecular events by biologists), and the "qualitative calculus" developed by Kuipers (1994), which permits certain differential equation-like operations even on data that are incomplete (one knows that one quantity is bigger than another but not by how much or one knows the temporal sequence of events but not the exact times). Testing the merits of such formalisms is extremely demanding because it requires tight collaborations between biologists and the formalists, and in my opinion it remains an open question how useful these might be. But any approach that does have explanatory power may then benefit from being articulated in ways that biologists can use to make it intuitive and might merit inclusion in the undergraduate biology curriculums of the future.

Bigger payoffs might come from formalisms that have not yet been devised. Sussman (2001) has pointed out that mathematics has its origins in a workaday human activity, the geometry invented and practiced by Egyptian surveyors. Just as systematization of insights from this human activity led to algebra and the rigorous symbolic methods derived from it, so the Twentieth-Century development of computers and the imperative languages used to control them (Do this! Now do that! OK, now, if this happens, do this again!) will likely give rise to a whole new field of knowledge that arises from formal imperative languages. Certainly, there are cases in which procedural imperative languages have simplified education in electronics and mechanics (see Sussman, 2001).

In biology, the prize from such formalisms might be greater still. That is because the DNA tape in the genome is a set of commands in a procedural imperative language. Insights from a new discipline arising from twentieth-century computer programming might be deeply revealing. As biologists, we cannot lose sight of the ground truths of our field, the wondrous sets of particular facts about living systems that our ceaseless experimentation has revealed. But it may be true that only by educating ourselves about concepts from other disciplines can we create abstractions that help us rise above these particulars. If that is true, then it is math and computer science that must show us the way.

ACKNOWLEDGMENTS

I am grateful to Andrew Gordon, Larry Lok, and Gerry Sussman for thoughtful discussions. Work on the Alpha Project at MSI is supported by a Center of Excellence grant from the National Human

Genome Research Institute. Computer work receives support from the Defense Advanced Research Projects Agency Information Processing Technology Office.

REFERENCES

Benford, G. (1995). The biological century. Reason *November*. (Revised version: *Reason Online*, July 2001; http://reason.com/9511/BENFORDfeat.shtml.)

Bialek, W., and Botstein, D. (2004). Introductory science and mathematics education for 21st-century biologist. Science *303*, 788–790.

Doyle, F., Francis, B., and Tannenbaum, A. (1992). Feedback Control Theory, New York: Macmillan.

Kluwer, W. D., and de Kleer (1990). Readings in Qualitative Reasoning About Physical Systems, San Mateo, CA: Morgan Kaufmann Inc.

Kuipers, B. (1994). Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge, Cambridge, MA: MIT Press.

Laurel, B. (2001). Confessions of a Utopian Entrepreneur, Cambridge, MA: MIT Press.

May, R. M. (2004). Uses and abuses of mathematics in biology. Science *303*, 790–793.

Simon, H. R. (1981). The Sciences of the Artificial, 1st ed., Cambridge, MA: MIT Press.

Sussman, G. J. (2001). The legacy of computer science. Presented at NCR/CSTB Symposium on the Fundamentals of Computer Science; www7.nationalacademies.org/cstb/project_fcs_agenda3.html.

DOI: 10.1187/cbe.04-03-0041

New Math for Biology Is the Old New Math

Ron Hoy
Department of Neurobiology and Behavior
Cornell University
Ithaca, NY 14853

It's true. If you live long enough, you will see good ideas (and bad) come, go, and return again. So if you're old enough to remember the early 1960s songster/musical prankster, Tom Lehrer, you might remember his clever satire called "New Math," which lampooned a popular curriculum reform of the same name. New math's proponents championed the teaching of set theory, Boolean algebra, vectors, matrices and Markov chains, combinatorics, and a little splash of game theory. The motivating factor behind this revolutionary outlook was that these skills would come in handy in the physical sciences, social sciences, and—yes—life sciences. Fast-forward four decades to 2004 and, sure enough, such skills are even more useful now than in those post-Sputnik driven days of science education. But today, most of our students don't possess them—and most of them didn't back then, either. What happened? Did the new math movement wither away out of neglect or no respect? No, not completely. In fact, new math did get integrated in bits and pieces into undergraduate and even into high school curricula, surviving the scorn of "back to basics" movements. Various components of new math, in the form of "finite" or "discrete" mathematics courses, were "grafted" into undergraduate courses, especially those substituting for or supplementing the (still) traditional year of calculus, particularly in courses for non-science majors. So is there a problem here? Now, in the age of the "new biology," especially, systems biology, I think there is. Our biology students still tend to be math-averse and many of us instructors could stand a brushing-up, if not outright resuscitation, of our own math skills, which may have atrophied from disuse.

Part of the problem is that the old new math is still taught largely as a math or computer science course by math professors, and although examples from applied disciplines provide ample problem set fodder, students don't seem to transfer those skills into their biology courses as facilely as we might hope. Students need to relearn matrix operations anew, although admittedly, cognitive "savings" may make the (re)learning curve less daunting a climb. It might be useful to take a quick glimpse at curriculum reform in mathematics and see how it has fared since, say, 1960.

Dartmouth College has long been a "hot spot" for mathematics curriculum innovations. The Dartmouth mathematician John G. Kemeny was the lead author of a remarkable textbook, *Introduction to Finite Mathematics* (IFN), first printed in 1957. This book began with symbolic logic and truth tables, then introduced the reader to set theory. This was followed by combinatorial math ("partitions and counting") that led directly to probability theory. The final section took up vectors and matrices, which led to game theory and linear programming. The last chapter integrated all this nice math into applications in genetics (Markov chains), economics (game theory), and even some graph theory (communication networks). The prerequiste for IFN was 2.5 years of high school math (1950s vintage) because it was aimed at college freshmen. The book was revolutionary and it went through several editions before ending up in "remainder heaven" by the 1980s, out of print; it had become unfashionable, sadly. Nonetheless, finite math morphed into discrete math in subsequent decades and, with it, a heavier dose of graph theory. Calculus is not a prerequisite for discrete math and its techniques are highly relevant for today's life sciences. I would never deny the utility and necessity of calculus for natural science students, including biology, but discrete math can be learned independently of calculus and, I believe, may draw on cognitive skills on the continuum of mathematical literacy other than calculus—perhaps more user-friendly to the practical mind of a biologist. To finish my digression into history, John Kemeny later became president of Dartmouth College, but even then he continued to contribute to undergraduate math education by inventing the well-known computer language, BASIC, which was intended to be more a means to teach and learn programming than a programmer's working language, such as FORTRAN or C; BASIC is still around and useful in an object-oriented incarnation (e.g., Visual Basic). In the 1990s, Dartmouth mathematicians, led by Dorothy Wallace, proposed a visionary math-based curriculum reform called "Mathematics Across the Curriculum" (MATC; http://www.math.dartmouth.edu/~matc/). MATC was enthusiastically supported by the National Science Foundation's Department of Undergraduate Education.