

## Comment

### Doctor Dunsel

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.  
E-mail: [petsko@brandeis.edu](mailto:petsko@brandeis.edu)

Published: 29 March 2004

*Genome Biology* 2004, **5**:104

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/4/104>

© 2004 BioMed Central Ltd

*"Dunsel ... is a term used by midshipmen at Starfleet Academy. It refers to a part which serves no useful purpose."*

Mr Spock, in *Star Trek*, Episode 53: "The Ultimate Computer"

Paranoia began morphing into depression with the arrival of the 15 January 2004 issue of *Nature*. On page 247 was a paper by King *et al.* entitled 'Functional genomic hypothesis generation and experimentation by a robot scientist'. The paper describes an automated system that uses techniques from artificial intelligence to formulate hypotheses to explain observations. The system then devises experiments to test these hypotheses, and actually carries out the experiments using a simple laboratory robot. But that's not all. It then interprets the results so as to falsify any hypotheses not consistent with the data. Moreover, it can iterate this process, making it capable of developing and testing quite extensive models.

In the paper, the authors used this system to probe the genetic control of aromatic amino-acid biosynthesis in yeast, using various growth conditions and auxotrophic strains. The robot scientist took a series of systematic gene deletion strains and tried growing each in nutritional medium that lacked one of the intermediates in the pathway. If the deleted gene was required to make that intermediate, the strain would not grow and a component of the pathway would have been identified. The machine automatically examined the cultures to see how opaque they were, returned the results to the artificial intelligence package, and then received instructions for what experiments to perform to validate the hypotheses based on the results of the first round, and so on. The final result was the assembled pathway: the set of genes coding for the enzymes that control each step. The authors claim in the end that the automated system carried out the project just as efficiently - and more cost-effectively - than scientifically trained human volunteers.

*Nature*, perhaps feeling guilty about the hordes of scientists who might be losing sleep over the prospect of having to go out and actually work for a living, tried to soften the blow with an editorial comment called 'Don't fear the Robot Scientist' (page 181 of the same issue) that completely missed the point. "Contrary to first impressions," the commentator says cheerily, "an automated system that designs its own experiments will benefit young molecular geneticists. At first glance, it seems to render obsolete the armies of postgrads and postdocs employed in the world's molecular-genetics laboratories."

That wasn't what was worrying me at all. Replacing my graduate students and postdocs with machines that would work around the clock and never pester me for more disk space on the computer or a new set of pipetmen; that would never complain about the temperature in the lab and never forget to clear up after themselves - that didn't sound so bad. It was the thought that it might eventually replace me that was frightening. After all, this thing didn't just carry out the experiments, it designed them and formulated hypotheses based on them. I thought I was supposed to do that.

*Nature* continued, "The team behind the Robot Scientist argues that such automation 'frees scientists to make the high-level creative leaps at which they excel'". Well, the thing already plans, performs and interprets experiments. Just what leaps would those be, guys - designing the next generation of software for the robot? Still, I decided after an initial bad moment or ten, the robot was carrying out functional genomics. As we all know, genomics doesn't require real thought, just the semblance of it. Maybe I would have to surrender my genomics projects to some machine, but that only represented a part of my research effort. The rest of my work is structural biology, a branch of science of such technical sophistication and intellectual rigor that it could never be automated.

Then the 10 February 2004 issue of *Proceedings of the National Academy of Sciences* arrived. On page 1537 - right

after a paper of my own, to add insult to injury - was an article by James Holton and Tom Alber (who was once my graduate student, to add injury to insult) entitled 'Automated protein crystal structure determination using ELVES'. It describes an expert system that can fully automatically determine the crystal structure of a protein from the primary X-ray data. True, individual steps in this process had been automated for some time, and the ELVES system had already been used to carry out such steps or even groups of steps, but always under the user's direction. This was different: there was no human intervention at all. The system was able to solve the structure of a 12,000 molecular weight coiled-coil protein from crystallographic data sets in two different crystal forms following a single command that launched the program and directed it to the location of the data files. The entire process, including interpretation of the resulting electron density map and refinement of the atomic model to convergence, took 9.5 hours on a multi-processor computer for one of the crystal forms, and 165 hours - the thing must have stopped for coffee or something - for the other form. The authors concluded that "high resolution structures with well-ordered metals can be determined automatically". To be fair, the protein structure, being all helical, did not present any real challenges in the model-building stage, and the authors are commendably candid about the limitations of the method: "ELVES is incapable of overcoming problems arising from poor data or inadequate phasing signal. Problems such as radiation damage, weak heavy atom signals, twinning, poor heavy atom models, low resolution, or crystal disorder that hinder crystallographic projects are not overcome by automation." Not yet, but just wait, I could hear them say *sotto voce*.

So, now I was about to be replaced as a crystallographer too. The year 2004 was sure turning out to be a terrific year. Well, strictly speaking I'm not paid just to do science anyway. Most of my salary comes from teaching undergraduates, and I consoled myself with the thought that I could always do more of that. Consoled myself, that is, until the arrival of last week's *Boston Globe* newspaper, with a story about a new effort at Massachusetts Institute of Technology (MIT) to revamp its undergraduate curriculum to take advantage of "innovations in educational methods". You know what that means - computer-based instruction. I could see it coming: once my lectures were all on the internet in interactive, self-test form, there would be no need for me to actually do any of the teaching myself anymore, or to be paid to do so - a fact I was sure would not be lost on any Brandeis administrator who might happen to read the article.

Feeling now very much like a horse might have felt about the time Henry Ford began turning out Model Ts, I tried to find something - anything - that I could do that a machine couldn't. Suddenly, it came to me: writing papers and grants. I probably spend half my non-teaching time writing things, things with highly technical content that also have to be comprehensible

to people in my field who aren't involved in the work I'm doing or am proposing to do. In fact, if I want to get a grant from a foundation or publish a paper in a high-profile, general journal like *Nature* or *Genome Biology*, I have to try to make this highly technical material comprehensible to people who aren't in my field at all. Automate that, if you can.

Well, that may not be far off, actually. As Clive Thompson has pointed out (*The New York Times Magazine*, 14 December 2003), the music business is making strides towards doing something very like that. An artificial intelligence program called Hit Song Science from the Barcelona-based company Polyphonic HMI tries to determine whether a new song is going to be a hit. It uses a clustering algorithm to locate acoustic similarities between songs, similarities like common bits of rhythm, harmonies or keys. It compares these features of a new tune with all the Top 40 hits of the last 30 years; the closer the features of a new song are to a 'hit cluster', the more likely it is predicted, by the software, to be a hit. Thompson reports that the algorithm produces some strange groupings - the rock group U2 is similar to Beethoven, for example - yet it seems to work. A number of record companies are now using it to help pick which songs on a new album they will promote heavily. And, perhaps ominously, others are using it in the studio to tweak new songs as they are being recorded, changing various aspects of them to bring them closer to the hits in the nearest cluster. All well and good for the record companies, but it seems to me that this process is likely to take the spontaneity - and much of the novelty - right out of the music business. Hit songs tend to sound too much alike as it is, at least to this jaded listener; now they are going to be forced to sound even more alike. And clearly the same approach could be used, theoretically at least, to produce grants with a high probability of being funded, and scientific papers guaranteed to be accepted by top-rank journals. Hot Paper Science would cluster the titles, author names and affiliations, title words and key concepts that are shared by papers published in *Cell*, for example. One then only has to input one's own initial effort, 'The complete sequence of the gerbil genome' by Gregory A Petsko, *et al.*, for example, and out would come 'Gerbil genome sequence: signal transduction pathways relevant to cancer, neurodegenerative diseases and apoptosis, with additional insights into systems biology and biodefense', plus a set of suggested coauthors that would help guarantee acceptance. The software would go on to write the paper, of course; submit it; and, if necessary, argue with the referees.

Well, that was it, I thought. Before long, even my writing functions would be taken over by machines. I was rapidly being made redundant, as they say in the UK - a twentieth-century equivalent of Captain Kirk in the *Star Trek* episode "The ultimate computer", his command capabilities handled more efficiently by a machine programmed to replace human beings in space exploration, his plaintive (and sexist)

cry, "But there are some things men must do to remain men!" drowned out by the footsteps of the relentless march of automation.

But then something happened to lift my gloom and restore my self-esteem. It was the arrival of an e-mail reminding me about the curriculum committee meeting scheduled for that afternoon. Of course! I wasn't useless after all. In fact, real human scientists are indispensable, and always will be. Computers may be better at solving crystal structures, and robots may be better at doing genome-enabled, hypothesis-driven experiments - may even be better at interpreting them - and eventually there will probably be software that writes better papers and grants, but we humans can still waste enormous amounts of time at interminable committee meetings. No machine will ever be stupid enough to do that.

comment

reviews

reports

deposited research

refereed research

interactions

information