Simplicity

Paul van Helden

Perhaps the most famous equation in science is $E=mc^2$; it also represents one of the best-known scientific insights. But why does Einstein's equation about the conservation of energy enjoy such popularity? Is it because it is easily and instinctively understood, or because it is widely taught in schools? I doubt that either explanation is true; rather, its appeal is its elegant simplicity within the complexity of Einstein's general theory of relativity.

There are various other discoveries of similar simplicity: the laws of thermodynamics, the finding that the double-helical structure of DNA with all genetic information is made up of only four bases, and the generation of ATP from glucose via the Krebs cycle. Of course, reality is more complicated than those simple answers suggest: DNA can exist as a single strand, a triple helix or a Holliday junction, and its bases can be further modified by methylation. Similarly, one molecule of glucose might not always be fully metabolized to yield six molecules of ATP under certain conditions. Nonetheless, many major advances in science have been based on such simple, uncomplicated models of reality.

So why do we still strive for a full and complex explanation or model that takes every detail into account? Why pursue a 'theory of everything'? Our world and our lives have become so complex that perhaps we have forgotten the basic simplicity of life and instead look for additional complexities. "It's too simple to be true, so it's probably wrong", is a common conclusion. The catastrophic explosion of the Challenger Space Shuttle in 1986, which was caused by problems with O-ring seals on the rocket boosters, should serve as a reminder: the simplest explanation might often be the correct one.

It seems we are obsessed with uncertainties: we use advanced statistics to calculate

every possible deviation. This is necessary for sending a satellite probe into outer space because a tiny error could result in a huge deviation from the desired path. Attention to detail is sometimes vital, but the basic principle remains simple. Perhaps (and I shudder at the thought) we should borrow a concept from law: we should not add unnecessary details, but rather work on "what the reasonable (man) might do." A model that produces a curvilinear output, for example, might generate a 'noisy' line with deviations and outliers, but further analysis does not increase the knowledge we gain from it. To analyse highly complex databases, we need mathematics and statistics to determine the significance of the findings. But this has little significance for the individual, only for the group as a whole. For example, the Kaplan–Meier method for calculating survival or death in clinical studies is a perfectly adequate tool and requires no further mathematics or statistics. If we apply additional statistical analysis to such data to show significance at the 5% level, one might argue that this is of questionable relevance and has no real meaning for any individual in the relevant study.

However, natural systems, whether large ones such as in ecosystems, or small ones such as signal transduction pathways, are highly complex. We do need some methods to understand and express this. The relatively new approach of systems biology is one promising attempt: the system might have many components, which appear to make it complex, yet each component is simple, and the outcome or product of the system as a whole is simple. The key lies in defining the rules that run our world: the story of Newton and the apple epitomizes this reductionist approach. Any attempts to understand the world should focus on identifying the most

important drivers in any system. Adding too much detail (fine-tuning) is not always an improvement and might stifle imagination and progress.

The Latin language has logical rules, but it is also infamous for the number of exceptions to the rules. It is essentially a dead language today, perhaps because it was unnecessarily complex. Any model that is too complex is likely to go the same way. Some time ago, the approach to finding new drugs was to construct three-dimensional structures of targets and to design compounds that match these. This approach has largely failed and most researchers have returned to simply testing a range of potential compounds followed by optimization.

Occam's razor, which is commonly misquoted as "the simplest explanation is most likely the correct one", is a principle that suggests we should tend towards simpler theories, even if experience shows that the simplest explanation might not be completely accurate. In such cases, we have to trade some simplicity to more closely approximate the truth. But there is a limit because increasing complexity might generate a parallel increase in uncertainty. Moreover, we scientists increasingly have to communicate our research to the public to avoid alienating ourselves from society. This is not made easier if we have to explain unnecessarily complex models: we need to distil simple take-home messages for explaining our work to laypeople.

*Paul van Helden is a professor and head of the Department of Biomedical Science at Stellenbosch University, Faculty of Health Science, Tygerberg, Cape Town, South Africa.
E-mail: pvh@sun.ac.za*

EMBO reports (2012) 13, 172; published online
3 February 2012; doi:10.1038/embor.2012.6