

George *et al.*<sup>4</sup> (page 975) identify one of the same four germline mutations<sup>1–3</sup> in 2 of 96 patients with apparently sporadic neuroblastoma (as other members of the families to which these two individuals belonged were not examined). Moreover, these studies<sup>1–4</sup> report that 59 of 617 (9.6%) sporadic neuroblastoma cases they investigated have somatic single-nucleotide mutations in *ALK*. Such somatic mutations were associated with more aggressive tumours and lethal cases of this cancer.

These studies<sup>1–4</sup> also found that, in contrast to normal *ALK*, the mutated enzyme was variably phosphorylated and had increased kinase activity in a mutation-specific way. Similarly, downstream targets of *ALK* were activated by the mutations in this enzyme in a mutation-dependent manner. Could the increased activity of *ALK* in neuroblastoma be inhibited for treatment purposes? For example, a leucine for arginine substitution at position 1174 of *ALK* results in phosphorylation of the *STAT3* and *AKT* proteins, whereas the replacement of arginine by glutamine at position 1275 was associated with phosphorylation of *AKT* and the *ERK1/2* protein. Although these data hint that *ALK* inhibition<sup>3,4</sup> could be a viable strategy for the treatment or prevention of neuroblastoma, they also indicate that multi-agent targeted therapy as well as nonspecific kinase inhibitors make sense, given that multiple signalling pathways consisting of several kinases are involved.

Mossé *et al.*<sup>1</sup> and Janoueix-Lerosey *et al.*<sup>2</sup> find

that although members of 6 of the 16 families they examined do not harbour mutations in *ALK* and *PHOX2B*, neuroblastoma runs in these pedigrees. At least three possibilities, which are not mutually exclusive, could explain these observations. First, mutations in other genes, such as *MYCN*, might be involved. Second, because *ALK*-mediated neuroblastoma involves increased activity of this protein, germline mutations in promoter sequences that favour *ALK* expression are possible.

Third, large genomic deletions and rearrangements could occur in the germ line. Somatic translocations involving *ALK* have been reported<sup>8</sup>, and so germline rearrangements in the sequence of this gene are also plausible. *ALK* deletions associated with neuroblastoma seem counter-intuitive, however, because it is increased *ALK* activity — rather than its absence — that seem to lead to this cancer. But germline deletions in regions containing repressors of *ALK* expression, as well as partial deletions creating *ALK*-like proteins, new proteins distinct from *ALK*, or continuously active kinases, are among plausible mechanisms.

Chen *et al.*<sup>3</sup> and Mossé and colleagues<sup>1</sup> provide evidence that somatic *ALK* mutations associate with aggressive forms of neuroblastoma in sporadic cases. Several questions arise from these observations. For example, will somatic mutations make *ALK*-associated familial neuroblastoma more aggressive? If the answer is yes, should adjuvant therapy —

perhaps *ALK* inhibitors — accompany surgical removal of the tumour? Could it be that *ALK* inhibition is toxic rather than beneficial in cancer cases caused by germline mutations? After all, every single cell in the body will carry the mutation, albeit with differential expression in different tissues.

The past two years have seen an explosion in genome-wide association studies, which have shown that variations of certain genes with low penetrance account for a small subset of various common cancers. At present, these data are not associated with much clinical context, and so cannot meaningfully contribute to genetic counselling and cancer management. In this era of genomic medicine, the long-awaited discovery of a major non-syndromic neuroblastoma gene<sup>1–4</sup> is indeed a welcome advance for taking pre-emptive measures (Box 1). ■ Charis Eng is at the Genomic Medicine Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA.

e-mail: engc@ccf.org

1. Mossé, Y. P. *et al.* *Nature* **455**, 930–935 (2008).
2. Janoueix-Lerosey, I. *et al.* *Nature* **455**, 967–970 (2008).
3. Chen, Y. *et al.* *Nature* **455**, 971–974 (2008).
4. George, R. E. *et al.* *Nature* **455**, 975–978 (2008).
5. Amiel, J. *et al.* *Nature Genet.* **33**, 459–461 (2003).
6. Mossé, Y. P. *et al.* *Am. J. Hum. Genet.* **75**, 727–730 (2004).
7. Zbuk, K. M. & Eng, C. *Nature Rev. Cancer* **7**, 35–45 (2007).
8. Soda, M. *et al.* *Nature* **448**, 561–566 (2007).

## PHILOSOPHY OF SCIENCE

# Theories of almost everything

P.-M. Binder

**A provocative contribution to the logic of science extends the theorems of Kurt Gödel and Alan Turing, and bears on thinking about prediction, the standard model of particles, and quantum gravity.**

Since 1620, when Francis Bacon's *Novum Organum* set out the basic guidelines, the task of science has been to condense multiple observations into brief, general descriptions of natural phenomena. This process, called induction, helps us understand and predict the world around us. Enquiries into the limits of science<sup>1</sup> have involved asking questions such as "Can we know everything about the natural world?", which have so far gone unanswered. Writing in *Physica D*, David Wolpert<sup>2</sup> has made headway in this direction by demonstrating that the entire physical Universe cannot be fully understood by any single inference system that exists within it.

Various major scientific developments of the twentieth century have placed limits on different facets of knowledge. These include the measurement process (quantum mechanics,

through Heisenberg's uncertainty principle); the transmission of information (relativity, through the constancy of the speed of light); the ability to predict the future from less-than-perfect measurements in the present (chaos theory, through sensitive dependence on initial conditions); and the efficient prediction<sup>3</sup> of certain natural phenomena before they unfold (complex systems theory, through intractability).

Wolpert's work follows another path, that of Kurt Gödel and Alan Turing's theorems of incompleteness developed in mathematics and computation (Box 1), and he extends them to address the logic of science. He introduces the idea of inference machines — physical devices that may or may not involve human input — that can measure data and perform computations, and that model how we come to

understand and predict nature. He develops a formal description of all such inference machines in terms of two functions: one stipulates the initial state of a machine (the set-up function) and the other (the conclusion function) describes the observations, recollections or predictions it makes — in other words, a 'theory'.

In proving his theorems, Wolpert defines *U* as the space of all world-lines (sequences of events) in the Universe that are consistent with the laws of physics. He then defines strong inference as the ability of one machine to predict the total conclusion function of another machine for all possible set-ups. Finally, he uses 'Cantor diagonalization' (Box 1) to prove, among others, the following two statements:

(1) Let  $C_1$  be any strong inference machine for *U*. There is another machine,  $C_2$ , that cannot be strongly inferred by  $C_1$ .

(2) No two strong inference machines can be strongly inferred from each other.

The first of these statements posits that there is a portion of 'knowledge space' (that inferable by  $C_2$ ) that is not available to any  $C_1$  machine. The second is a statement about the non-equivalence of inference machines; it implies that, at most, only one machine at one

instant in time can infer all others. The two statements together imply that, at best, there can be only a ‘theory of almost everything’. Hence, they slam the door on Pierre-Simon Laplace’s ‘demon’, introduced in 1814. This hypothetical being has a “vast intellect”, such that, with full knowledge of the state of the Universe at one time, it can completely predict the future and recall the past with no uncertainty whatsoever. Wolpert’s results are particularly compelling because they are totally independent of both the details of the laws of physics and the computational characteristics of the machines.

What are the practical consequences of these findings for science? A prescription for deriving the laws of nature was proposed by Roy Frieden<sup>4</sup> a few years ago. He asserted that any attempt to measure a physical quantity elicits a transfer of information from the ‘source’ physical phenomenon to the observer. As a consequence, the observation of a physical phenomenon cannot be entirely accurate. By manipulating ‘Fisher information’, a measure of the quality of data, one can obtain information about the equations that govern the phenomenon in addition to the numerical value one seeks. Most fundamental equations of physics can be derived in this way (thus making physics a minor branch of statistics!). But Wolpert’s work warns us that Frieden’s recipe is bound to fail at least once, perhaps by producing multiple solutions in certain cases.

Another example of the relevance of Wolpert’s work to science is in predicting the behaviour of chaotic systems. Through the attractor reconstruction method, in which a time series is converted into a geometrical trajectory in higher-dimensional spaces<sup>5</sup>, one can forecast the evolution of fairly complex systems up to a specified prediction horizon. This method works quite well, and does not need the explicit knowledge of the system’s governing equations (see ref. 6 for an astrophysical example). However, when one tries to infer the equations themselves, the results are often unclear or ambiguous<sup>7</sup>, possibly as a manifestation of Wolpert’s theorems.

Finally, Wolpert’s findings have a bearing on the possible limitations of two theories in physics. One set of limitations concerns the standard model of particle physics, which has accumulated a long list of shortcomings<sup>8</sup>: these include an exceedingly high predicted cosmological constant, failure to predict the mass of the Higgs particle, and failure to account for the ‘dark matter’ in the Universe. The other limitation is our inability to bring quantum mechanics and gravity into a single theory, although several viable alternative theories are being studied<sup>9</sup>. Quantum electrodynamics, a refinement of quantum mechanics, is defined by just two parameters (the charge and mass of the electron), whereas quantum gravity would require infinitely many parameters, and hence infinite experiments to determine those parameters, making it so

### Box 1 Cantor, Gödel, Turing and the uncountable

In 1874, Georg Cantor published a proof of the existence of uncountable infinities. He started by labelling points in the interval  $[0, 1)$  with the countable infinite natural numbers  $(1, 2, 3, \dots)$  as follows:

$$a_1 = 0.d_{11}d_{12}d_{13}d_{14}d_{15}\dots \text{ (for example, } a_1 = 0.31415926\dots \text{ with } d_{11} = 3, d_{12} = 1, \text{ etc.)}$$

$$a_2 = 0.d_{21}d_{22}d_{23}d_{24}d_{25}\dots$$

$$a_3 = 0.d_{31}d_{32}d_{33}d_{34}d_{35}\dots$$

...

All  $d$ s are digits between 0 and 9. There is at least one number  $a_x = 0.d_{x1}d_{x2}d_{x3}\dots$  in the unit interval  $[0, 1)$  such that  $d_{x1} \neq d_{11}$ ,  $d_{x2} \neq d_{22}$ , and so

on, guaranteeing that it differs from each number in the list by at least one digit, and hence it cannot be in the list. This proves that the number of points in the unit interval is not countable, a proof known as Cantor diagonalization.

A related result, the diagonal lemma, has played an important part in the proof of several incompleteness theorems. In 1931, Kurt Gödel proved that any mathematical system that includes enough of the theory of natural numbers contains statements that cannot be proved to be either true or false, and is thus incomplete. The general argument for the proof is

based on Epimenides’ liar’s paradox — is ‘This sentence is false’ a true or a false statement? — but it replaces ‘false’ with ‘unprovable’. The construction of a number that represents a concrete, undecidable statement requires diagonalization techniques.

In the mid-1930s, Alan Turing used similar methods to prove that no general algorithm can determine whether a given Turing machine — a computer — halts for a given input. It is thus not surprising that Wolpert’s proofs also rely on diagonalization arguments.

P.-M.B.

far a meaningless theory. It is possible, though, that these various theories, along with all that we have learned in physics and other scientific disciplines, will yet merge into the best science can do: a theory of almost everything. ■

P.-M. Binder is in the Department of Physics and Astronomy, University of Hawaii, Hilo, Hawaii 96720, USA, and at the Center for Nonlinear Dynamics, University of Texas, Austin. e-mail: pbinder@hawaii.edu

1. Barrow, J. D. *Impossibility: The Limits of Science and the Science of Limits* (Oxford Univ. Press, 1998).
2. Wolpert, D. H. *Physica D* **237**, 1257–1281 (2008).
3. Wolfram, S. *Phys. Rev. Lett.* **54**, 735–738 (1985).
4. Frieden, B. R. *Science from Fisher Information: A Unification* (Cambridge Univ. Press, 2004).
5. Kantz, H. & Schreiber, T. *Nonlinear Time Series Analysis* (Cambridge Univ. Press, 2004).
6. Binder, P.-M., Crosson, I. J. & Cadmus, R. R. Jr *Astrophys. J.* **685**, L145–L148 (2008).
7. Crutchfield, J. P. & McNamara, B. S. *Complex Systems* **1**, 417–452 (1987).
8. Ellis, J. *Phys. World* **12** (12), 43–48 (1999).
9. Wilczek, F. *The Lightness of Being* (Basic Books, 2008).

### MOLECULAR BIOLOGY

## Bound to splice

Bruce Fitcher and Janet K. Leatherwood

**Messenger RNAs don’t usually correspond exactly to DNA — portions of the primary transcript, known as introns, are removed by splicing. A study reveals new ways in which splicing can be regulated.**

Complex eukaryotes, such as animals, have extensive RNA splicing to remove sequences that don’t encode proteins (introns) and to connect those that do (exons). Often, several messenger RNAs can be generated by a single gene, because different patterns of splicing place different exons into the final mRNAs. As an example, the *Dscam* gene of the fruitfly *Drosophila* contains 24 exons, and is thought to encode 38,016 different protein isoforms by alternative splicing (although we’re not sure anyone has counted)<sup>1</sup>. In this particular case, ‘docking’ and ‘selector’ sequences within the primary transcript help regulate splicing. But in general, understanding exactly how a cell picks and chooses among the many possible combinations of splices has been a long-standing problem<sup>2</sup>. Reporting on page 997 of this issue, Moldón *et al.*<sup>3</sup> investigate how a gene’s promoter region — the sequence that regulates

gene expression — can regulate splicing.

As ever, one approach is to work with a simpler eukaryote. The fission yeast *Schizosaccharomyces pombe* has a sophisticated splicing apparatus, and regulates splicing, although this regulation is somewhat different from that in more complex eukaryotes. In the latter, introns are humongous, and the splicing system may therefore focus on recognizing the relatively small exons. When splicing is altered, it may fail to recognize an exon, leaving that exon out of the final product, and giving alternative splicing. In yeast, introns are tiny, and the splicing system seems to focus on recognizing the introns. Thus, when splicing is altered, the system may fail to recognize an intron, retaining the intron in the final product. Whatever the reason, the usual nature of regulated splicing in animals is alternative splicing (an altered selection of exons), whereas the usual nature of