Liber Amicorum
Cristian S. Calude 70

Editors
C. Câmpeanu
M. J. Dinneen
K. Svozil

Contents

A Pattern in Cris Calude's Work

Two decades ago, while one of us[20] was a PhD student, I discovered *Information and Randomness* [3]. I wrote to Cris with a question about the book. He was surprised to find out about applications of his work on algorithmic information theory in accounting and economics, far outside his core fields of mathematics and computer science, and corresponded with me at some length. He then agreed to serve on my dissertation committee and generously hosted me in Auckland for a research visit.

During this visit, when we talked shop, our conversations mostly focused on philosophical issues. (When we weren't talking shop, I believe I provided him with an unintended constructive proof about a partial order, demonstrating the existence of a bottom element of any ranking of Ping Pong playing talent.) He shared Calude [1] with me, prompting us to focus on an issue close to his heart: what constitutes a proof? Is it enough to know that something follows from some axioms and rules of inference, or is a proof something that provides deeper insight? In this note, we discuss how much of the body of his work relates to this philosophical question — on this, Paul Pedersen, the second author, joins in on the discussion.

We could go the opposite direction: Is it enough to know something without having any deeper insight? Big Data champions say yes: never mind about the whys and wherefores. Stable patterns tell us all we need to know.

Stable patterns in data are extremely rare. In Calude and Zamfirescu [2, 8], Cris shows us that almost all numbers have no consistent statistical properties. Restricting attention to the numbers in the unit interval $[0,1]$, expressed in some base b, Cris and his coauthor shows that the $\liminf$ of the probability of any numerical pattern (i.e., word over the base b) occurring is 0, and the $\limsup$ is 1. Because patterns are so rare, they attract our attention.

This does not mean, however, Big Data champions have it right, because they don't. As the amount of data increases, spurious correlations, even stable ones, appear at a much faster rate than correlations reflecting a relationship that occurs for some underlying reason. Cris and his coauthors

[20] Stecher

make this argument precise, using insights from ergodic theory, in Calude and Longo [6], Calude and Svozil [7]. It is not simply that as we get more data, we see more nonsense. As we get more data, almost all of the patterns we see are garbage. As Cris and his coauthor put it: “Too much information tends to behave like very little information” [6, p600].

Cris’ argument is reminiscent of one now famous in macroeconomics, due to Lucas, Jr. [10], later expanded upon by Cooley and LeRoy [9]: stable correlations in economic data are not enough for making policy decisions. Consider the relationship between the number of calls to an Auckland Fire Station and the frequency of houses burning to the ground. It would not surprise anyone to find a high positive correlation, one that gets stronger as data sets grow, and that is highly persistent. Without knowing the underlying mechanism, it would be easy to conclude that fire departments recruit arsonists and pyromaniacs; to spare your house, never call the fire department and remove your smoke detectors. A similar example often comes up in econometrics classes: sales of ice cream are said to predict homicides. Surely this is because ice cream sales are more common in weather conditions in which people are more likely to interact. But the relationship is also consistent with an argument that banning ice cream would reduce crime.

So if the numbers can’t speak for themselves, who will speak for them? Don’t look to observed correlations. Run experiments. The laboratory speaks for the numbers.

Cris uses experiments in several papers; we comment on two of them here. Calude et al. [5] estimates a Chaitin Omega number (i.e., a halting probability). Omega numbers are strongly noncomputable. Nevertheless, using experiments, they obtain an exact initial segment of the Omega number.

Outside of theoretical computer science, Cris has used experimental methods to address problems in pure mathematics. Finding the 6^{th} Taxicab number, for example, is known to be difficult. In Calude et al. [4], Cris and his coauthors propose a candidate for Taxicab(6). Their experiments show that their candidate is the right answer with a p-value of approximately 0.002.

Cris’ work spans many fields. All of his research aims to address the questions we discussed during my visit to Auckland as a PhD student. The problems he addresses in his papers are in computer science and mathematics, but his sights are always set on larger philosophical issues.

References

[1] Calude, A. (2001). The journey of the four colour theorem through time. *New Zealand Mathematics*, 38(3):1–10.

[2] Calude, C. and Zamfirescu, T. (1999). Most numbers obey no probability laws. *Publicationes Mathematicae Debrecen*, 54(Supplement):619–623.

[3] Calude, C. S. (2003). *Information and Randomness*. Springer-Verlag, second edition.

[4] Calude, C. S., Calude, E., and Dinneen, M. J. (2005). What is the value of Taxicab(6)? an update. Research Report CDMTCS-261, Centre for Discrete Mathematics and Theoretical Computer Science, University of Auckland Centre for Discrete Mathematics and Theoretical Computer Science.

[5] Calude, C. S., Dinneen, M. J., and Shu, C.-K. (2002). Computing a glimpse of randomness. *Experimental Mathematics*, 11(3):361–370.

[6] Calude, C. S. and Longo, G. (2017). The deluge of spurious correlations in big data. *Foundations of Science*, 22(3):595–612.

[7] Calude, C. S. and Svozil, K. (2019). Spurious, emergent laws in number worlds. *Philosophies*, 4(2):Article no. 17 (13 pp.).

[8] Calude, C. S. and Zamfirescu, T. (1998). The typical number is a lexicon. *New Zealand Journal of Mathematics*, 27:7–13.

[9] Cooley, T. F. and LeRoy, S. F. (1985). Atheoretical macroeconometrics: A critique. *Journal of Monetary Economics*, 16(3):283–308.

[10] Lucas, Jr., R. E. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester Conference Series on Public Policy*, volume 1, pages 19–46.

Jack Stecher
University of Alberta
email: stecher@ualberta.ca
Arthur Paul Pedersen
City University of New York
email: apedersen@cs.ccny.cuny.edu