On the “mysterious” effectiveness of mathematics in science

J Gerard Wolff*

December 11, 2017

Abstract

This paper notes first that the effectiveness of mathematics in science appears to some writers to be “mysterious” or “unreasonable”. Then reasons are given for thinking that science is, at root, the search for compression in the world. At more length, several reasons are given for believing that mathematics is, fundamentally, a set of techniques for information compression via the matching and unification of patterns (ICMUP), and their application. From there, it is argued that the effectiveness of mathematics in science is because it provides a means of achieving the compression of information which lies at the heart of science. The ‘anthropic principle’ provides an explanation for why we find the world—aspects of it at least—to be compressible; and the ‘contingency principle’ provides a justification for being prepared for compressibility, if or when it arises. ICMUP may be seen to be important in both science and mathematics, not only as a means of representing knowledge succinctly, but as a basis for scientific and mathematical inferences—because of the intimate relation that is known to exist between information compression and concepts of prediction and probability. Since ICMUP is a key part of the *SP theory of intelligence*, evidence presented in this paper strengthens the already-strong position of the SP theory as a unifying principle across artificial intelligence, mainstream computing, mathematics, human learning, perception, and cognition, and neuroscience. The evidence and ideas in this paper may provide the basis for a “new mathematics for science” with potential benefits and applications in science and science-related areas.

*Dr Gerry Wolff, BA (Cantab), PhD (Wales), CEng, MBCS, MIEEE; CognitionResearch.org, Menai Bridge, UK; jgw@cognitionresearch.org; +44 (0) 1248 712962; +44 (0) 7746 290775; *Skype*: gerry.wolff; *Web*: www.cognitionresearch.org.

Keywords: information compression, philosophy of mathematics, philosophy of science, computing, learning, perception, cognition.

1 Introduction

Although mathematics is a phenomenally successful "handmaiden" of science,[1] the reason that it is so effective in science has been described as a "mystery" that is "unreasonable". Thus:

- Roger Penrose writes:

> "It is remarkable that *all* the SUPERB theories of Nature have proved to be extraordinarily fertile as sources of mathematical ideas. There is **a deep and beautiful mystery** in this fact: that these superbly accurate theories are also extraordinarily fruitful simply as *mathematics*." ([37, pp. 225–226], bold face added).

- In a similar vein, John Barrow writes:

> "**For some mysterious reason** mathematics has proved itself a reliable guide to the world in which we live and of which we are a part.[2] Mathematics works: as a result we have been tempted to equate understanding of the world with its mathematical encapsulization. ... **Why is the world found to be so unerringly mathematical?**" ([6, Preface, p. vii], bold face added).

- And Eugene Wigner [50] writes about "The unreasonable effectiveness of mathematics in the natural sciences":

> "The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which **we neither understand** nor deserve. We should be grateful for it and hope that it will remain valid

[1]The slightly whimsical idea that mathematics might be some kind of servant of science, and the use of the curiously archaic word "handmaiden", seems to have originated with *The Handmaiden of the Sciences*, a book by Eric Bell [8].

[2]It is clear that, in this quote, the expression "the world" is intended to mean "everything in the observable universe", in accordance with normal usage. That expression is intended to have the same meaning elsewhere in this paper.

> in future research and that it will extend, for better or for worse, to our pleasure, even though perhaps also to **our bafflement**, to wide branches of learning." (*ibid*, p. 14, bold face added).

This paper, which draws on, and considerably expands, some of the thinking in [52, Chapter 10], offers some reasons why mathematics can be effective in many areas of science.

1.1 Overview of arguments to be presented

The gist of the arguments presented in this paper is that:

- Much of mathematics, perhaps all of it, may be understood as a set of techniques for the compression of information[3] via the matching and merging or 'unification' of patterns,[4] and their application. In this view:
  - Mathematics is not merely some kind of tool for compressing information like WinZip or Gzip. Instead, mathematics is seen here as being, at its most fundamental level, about the compression of information.
  - As described in Section 2.6, the search for patterns that match each other, and their merging or unification, is an effective approach to the compression of information. As described in [60], this approach is prominent in human learning, perception, and cognition.
  - As argued in Sections 3 and 4, compression of information in mathematics, and in the mathematics-related disciplines of logic and computing, is achieved largely via the matching and unification of patterns.

  For the sake of brevity, the expression "information compression via the matching and unification of patterns" will be referred to as 'ICMUP',

[3]This paper treats terms like 'information', 'data', and 'knowledge' as being equivalent, although 'information' is intended to be most general, 'data' is intended to suggest raw data from the world, with or without some low-level processing, and 'knowledge' is intended to suggest information that has been encoded via human perception and cognition.

[4]Here and throughout this paper, the term 'unification' will mean a simple merging of two or more matching patterns to make one. This use of the term is only loosely related to the meaning of the term 'unification' in logic and computer science.

the expression “mathematics as a set of techniques for ICMUP, and their application” will be referred to as ‘MICMUPA’, and the expression “human learning, perception, and cognition” will be referred to as ‘HLPC’..

- Since “Science is, at root, just the search for compression in the world” [6, p. 247], and since mathematics has an important role in science, MICMUPA may be seen to be an important means, amongst others, for describing scientific knowledge in a compressed form.

- The idea that science is fundamentally about compressing information about the world rests on the assumption that that information is compressible, with the implied question: “Why should we assume that the world is compressible?” An answer to that question is, in accordance with the anthropic principle, that the world must be at least partly compressible because, if it were not, everything in it, including ourselves, would be a soup of randomness.

- In view of the close connection that is known to exist between information compression and concepts of prediction and probability (Section 6), MICMUPA may be seen as a driver for the making of mathematically-based inferences.

An important point in this paper is that, for reasons given in Section 2.6, ICMUP differs sharply from mathematically-oriented approaches to information compression including those derived from algorithmic information theory (AIT) [28] and such techniques as Huffman coding, arithmetic coding, and wavelet compression.

1.2 Presentation

The next section provides some background to the main body of the paper including: a summary of variants of the concept of ‘pattern’; a sketch of the *SP theory of intelligence* and its realisation in the *SP computer model*; research relating to the effectiveness of mathematics in science and elsewhere, to mathematics as compression of information, and to information compression in science; and an outline of seven techniques for ICMUP.

Section 3 describes how mathematics may be seen as a set of techniques for the compression of information via the matching and unification of patterns, and their application.

Section 4 outlines how mathematics-related disciplines—specifically logic and computing—may be seen as techniques for the compression of information via the matching and unification of patterns, and their application.

On the strength of evidence presented in the preceding sections, Section 5 describes a solution to the mystery of why mathematics is effective in science. This section also contains a brief discussion of the above-mentioned anthropic principle as an explanation of why the world that we inhabit should be compressible.

Section 6 discusses briefly how MICMUPA is in keeping with other evidence that mathematics is fundamentally probabilistic, and how this may be reconciled with the all-or-nothing, 'exact', forms of calculation or inference that are familiar in mathematics.

Section 7 outlines some possible implications of ideas presented in this paper.

2 Background

This section outlines some preliminaries to the sections that follow.

2.1 Concepts of 'pattern'

Since, in this and related areas of research, the word 'pattern' has several different meanings, they are distinguished here and elsewhere in this paper by the use of different names:

- *Pattern.* The word 'pattern', without any distinguishing prefix, is intended to mean an array of atomic symbols in one or two dimensions. Here, an atomic symbol is one that may be matched in an all-or-nothing manner with any other such symbol, and has no other meaning. Unless otherwise stated, the meaning of 'pattern' is generic and may encompass any or all of the meanings that follow.

- *E-pattern.* An 'E-pattern' is intended to mean a pattern that is derived relatively directly from the environment of a person or artificial system, perhaps with the conversion of analogue information to an array of atomic symbols but without any other processing.

- *R-pattern.* An 'R-pattern' is one that has been created by the merging or 'unification' of two or more patterns and thus reflects informational redundancy in the information from which it was derived. It is much like a 'chunk' of information in the chunking-with-codes technique for compression of information (Section 2.6.2).

- *A-pattern*. An 'A-pattern' is an abstract pattern that may be seen to be derived from E-patterns but via a more complex route than an R-pattern.

- *SP-pattern*. An 'SP-pattern' is any discrete pattern within the SP system, as described in Appendix A.

2.2 The SP theory of intelligence

The ideas and arguments presented in this paper are part of a long-term programme of research developing the *SP theory of intelligence* and its realisation in the *SP computer model*. This subsection outlines some key ideas in this research. There is more detail about the SP system in Appendix A with pointers to where fuller information may be found.

The main aim in this research is, in accordance with Ockham's razor, simplification and integration of observations and concepts across a wide area. In this research, that wide area encompasses artificial intelligence, mainstream computing, mathematics, and HLPC. And in this research, ICMUP is a unifying theme.[5]

The focus on ICMUP is because there is substantial evidence that much of HLPC may be understood in those terms [60].

An important idea in the SP system is the powerful concept of *SP-multiple-alignment*, borrowed and adapted from the concept of 'multiple sequence alignment' in bioinformatics. This concept, outlined in Appendix A.1, is the key to the SP system's versatility in modelling diverse aspects of human intelligence, in the representation of diverse kinds of knowledge, and in the seamless integration of diverse aspects of intelligence and diverse forms of knowledge, in any combination.

2.3 Research relating to the effectiveness of mathematics in science and elsewhere

This subsection briefly reviews some studies with a bearing on the effectiveness of mathematics in science and elsewhere.

[5]Since people often ask what the name "SP" stands for, it is short for *Simplicity* and *Power*, two ideas that, together, mean the same as information compression. This is because compression of a body of information, **I**, means maximising the *simplicity* of **I** by reducing, as much as possible, repetition of information or *redundancy* in **I**, whilst retaining a percentage of its non-redundant descriptive or explanatory *power*: 100% in the case of lossless compression, or some smaller percentage in the case of lossy compression.

Mark Colyvan [15], writing about the effectiveness of mathematics in population ecology (as an example of an empirical science that is not physics), argues that mathematics can play a number of useful roles in ecological theory, that contrary to superficial impressions, mathematics can represent some biological facts, and that it can help to raise the level of abstraction above "the push and shove of base-level causal processes".

Without disagreeing with any of those arguments, some comments are offered here that relate to themes of this paper:

- With regard to the raising of levels of abstraction, there is nothing in the concept of MICMUPA to say that processes for compressing information must necessarily be applied to *all* the available information in a given domain. In the same way that people can and frequently do drive cars without worrying about the details of how a car works, scientists can and often do abstract away from some aspects of a domain and concentrate on others.

- Although it is probably true that the mathematics of today can represent some aspects of the biology of living things, it is probably also true that it may struggle or fail with some other aspects, and likewise in other areas of science (more below).

- Although this matter has not yet been explored in any detail, it appears that the SP system, with its focus on compression of information, may provide an explanatory framework for causal processes. How the SP system may be applied in causal analysis is discussed in [52, Section 7.9], with an example from the SP computer model.

Roughly in order of dates, other studies relating to the effectiveness of mathematics include:

- With reference to Wigner's [50] paper on "The unreasonable effectiveness of mathematics in the natural sciences", Richard Hamming [22] concludes that, for a variety of reasons, "I am forced to conclude both that mathematics is unreasonably effective" (p. 90) but that "we ... must continue to try to explain why the logical side of science—meaning mathematics, mainly—is the proper tool for exploring the universe as we perceive it at present." (p. 90).

- Mark Steiner, in his book *The Applicability of Mathematics as a Philosophical Problem* [48], discusses, as the name of the book suggests, philosophical aspects of the "applicability" of mathematics, taking account of alternative meanings for that word.

- In a paper called "The unreasonable uncooperativeness of mathematics in the natural sciences" Mark Wilson [51] writes in support of "mathematical opportunism", the idea that "... it is the job of the applied mathematician to look out for *special circumstances* that allow mathematics to say something useful about physical behaviour. ... Descartes frequently declared ... that many natural phenomena are too complicated to submit to *any* mathematical description." (ibid., p. 297, emphasis in the original).

- Mark Colyvan, in his book *The Indispensability of Mathematics* [14], writes that "One of the most intriguing features of mathematics is its applicability to empirical science. ... Not only does mathematics help with empirical predictions, it allows elegant and economical statement of many theories. Indeed, so important is the language of mathematics that it is hard to imagine how some theories could even be stated without it. Furthermore, looking at the world through mathematical eyes has, on more than one occasion, facilitated enormous breakthroughs in science." (p. 6). He goes on to discuss several aspects of the idea that mathematics is indispensable for empirical science and how that idea relates to several issues in the philosophy of mathematics, especially the validity or otherwise of the 'Platonist' or mathematical 'realist' idea that "... mathematical entities such as functions, numbers, and sets have mind- and language-independent existence ..." (p. 2).

- Alan Baker [2, 1] and Aidan Lyon with Mark Colyvan [30], like Colyvan [15] and elsewhere, argue that mathematical models can provide genuinely mathematical explanations of biological facts. As noted above, the likely truth of that conclusion does not exclude the possibility that the effectiveness of mathematics in modelling the world may be improved.

- Contrary to Wigner's [50] perception of the "unreasonable" effectiveness of mathematics in science, Ivor Grattan-Guinness [21] argues that the effectiveness of mathematics in science is: 1) because much mathematics was brought into being by the need to model one or more aspects of the world; and 2) because, for several reasons, mathematics is connected to the natural sciences via rational links.

- Robert Batterman [7] discusses the problem of providing a coherent account of how mathematical idealisations can play explanatory roles in physical theory. By contrast with arguments by Otávio Bueno and Mark Colyvan [11], Batterman argues that it is not always clear that,

> when mathematics gets applied in empirical science, it is because the relevant part of mathematics reflects physical structures. Because of difficulties like that, he suggests that a completely new approach is needed with the world as the "driving influence" for how mathematics gets applied.[6] For those kinds of reasons he takes issue with what [50] implies: that mathematics is always appropriate as a language for science.

A general point that emerges more-or-less explicitly from several of these studies is that, while mathematics can be very effective in science, it is by no means perfect. Mathematics is excellent for expressing such things as $E = mc^2$ but it is less useful, for example, in expressing the key ideas in the Darwin/Wallace theory of evolution by natural selection, and there are many other aspects of science that are not easily described with mathematics.[7] If it were a universal language for science, there would be no need for scientists to use natural language, still and moving pictures, diagrams, and more, to express scientific knowledge and scientific concepts.

These points relate to the possible development of a "new mathematics for science" with enhanced capabilities for modelling scientific observations and concepts, noted in the fourth bullet point in Section 7.

2.4 Research relating to mathematics as compression of information

Three recent books about the philosophy of mathematics [29, 31, 35] make no mention of anything resembling MICMUPA. More generally, the idea that information compression might be part of the foundations of mathematics appears to be invisible in writings about the nature of mathematics.

Keith Devlin's academic book, *Logic and Information* [18], aims to develop a mathematical theory of information, a goal which is related to but

[6] Reading between the lines, this looks a bit like a "new mathematics for science" mentioned in Section 7.

[7] In this connection: "... against Wigner's 'unreasonable effectiveness' statement (based on success in the physical sciences) one must ask why maths is often so unreasonably ineffective in the human and social sciences of behaviour, psychology, economics, and the study of life and consciousness. These complex sciences are dominated by non-linear behaviour and only started to be explored effectively by many people (rather than only huge well-funded research groups) with the advent of small personal computers (since the late 1980s) and the availability of fast supercomputers. Some complex sciences contain unpredictabilities in principle (not just in practice): predicting the economy changes the economy whereas predicting the weather doesn't change the weather." John Barrow, personal communication, 2017-04-06, with permission.

distinct from the central idea in MICMUPA, that mathematics may be seen as a set of techniques for ICMUP.

Devlin's later book for the general reader, *Mathematics: The Science of Patterns* [19], discusses things like "patterns of symmetry [such as] the symmetry of a snowflake or a flower" (p. 145) and "the patterns involved in packing objects in an efficient manner" (p. 152) which are somewhat like the concept of 'R-pattern' or 'A-pattern' described in Section 2.1. But the key ideas in MICMUPA, as described in this paper, are not discussed.

Amongst the several "isms" in the philosophy of mathematics—foundationism, logicism, intuitionism, formalism, Platonism, neo-Fregeanism, and more—the three which are perhaps most closely related to MICMUPA are: *psychologism* (mathematical concepts derive from human psychology); *embodied mind theories* (mathematical thought is a natural outgrowth of human cognition); and *intuitionism* (mathematics is a creation of the human mind). This is because the latter three views are broadly consistent with the aforementioned evidence that much of HLPC may be understood as compression of information [60]. But it appears that there is nothing like MICMUPA in any of those three views or any other school of thought in the philosophy of mathematics.

2.5 Research relating to information compression in science

This section briefly reviews other research relating to information compression in science, first considering the role of information compression in science, and then considering the apparent absence of any recognition of MICMUPA, or anything like it, as a principle in science.

2.5.1 Compression of information in science

Few people will dispute that much of science is concerned with observational studies of aspects of the world or conducting experiments to obtain empirical data. And most people with a knowledge of science will agree that compression of information, also described as simplification, is an important part of science. Here are some relevant topics:

- *Ockham's Razor*. Ockham's razor, the principle attributed to William of Ockham, is widely seen as a key principle in the development of theories, in science and other areas. It is often expressed as "Entities are not to be multiplied beyond necessity."—meaning that, when there are two or more competing theories that explain a given set of phenomena, we should choose the simplest. More generally:

> "There is a widespread philosophical presumption that simplicity is a theoretical virtue. This presumption that simpler theories are preferable appears in many guises. Often it remains implicit; sometimes it is invoked as a primitive, self-evident proposition; other times it is elevated to the status of a 'Principle' and labeled as such (for example, the 'Principle of Parsimony'). However, it is perhaps best known by the name 'Occam's (or Ockham's) Razor.' Simplicity principles have been proposed in various forms by theologians, philosophers, and scientists, from ancient through medieval to modern times. [3].

- *Views of Scientists.* Respected scientists have often described the goals of science in similar terms. Isaac Newton wrote that "Nature is pleased with simplicity" [33, p. 320]; Ernst Mach [4] and Karl Pearson [36] suggested independently of each other that scientific laws promote "economy of thought"; Albert Einstein wrote that "A theory is more impressive the greater the simplicity of its premises, the more different things it relates, and the more expanded its area of application.";[8] cosmologist John Barrow has written that "Science is, at root, just the search for compression in the world" [6, p. 247];[9] and Ming Li with Paul Vitányi have written that "Science may be regarded as the art of data compression" [28, p. 585]. It is pertinent to mention that George Kingsley Zipf developed the related idea that human behaviour is governed by a "principle of least effort" [64].

- *Some Examples of Simplifications in Science.* Carlo Rovelli [40, Location 812] provides an excellent summary of some of the more important simplifications that have been achieved in physics:

> "... as space and time fuse together in a single concept of spacetime, so the electric field and the magnetic fields fuse together in the same way, merging into a single entity which

[8] Quoted in [24, p. 512].

[9] It may seem inappropriate in an academic paper to quote from a book like *Pi in the Sky* which is intended for the general reader. But the author, John Barrow, is a respected cosmologist, theoretical physicist, and mathematician, who is currently Research Professor of Mathematical Sciences at the University of Cambridge. Even in a non-academic book, what he has to say about the nature of science carries weight.

today we call the electromagnetic field. The complicated equations written by Maxwell for the two fields become simple when written in this new language. ... The concepts of 'energy' and 'mass' become combined in the same way as time and space, and electric and magnetic fields, are fused together in the new mechanics. ... Einstein realizes that energy and mass are two facets of the same entity, just as the electric and magnetic fields are two facets of the same field, and as space and time are two facets of the one thing: space-time. This implies that mass, by itself, is not conserved; and energy—as it was conceived at the time—is not independently conserved either. One may be transformed into the other: only one single law of conservation exists, not two. What is conserved is the sum of mass and energy, not each separately. Processes must exist that transform energy into mass, or mass into energy."

- *Discussions of Simplicity in Science*. A book-length review and discussion of *Simplicity in Science* may be found in [42]. A shorter but relatively full discussion of the role of simplicity in science may be found in [44]. The main thrust of the latter publication appears to be a plea for greater precision in definitions of 'simplicity' in science.

- *Simplicity and Power*. Since competing theories rarely address exactly the same set of phenomena, Ockham's razor may be adapted to be "In the development of a scientific theory from a body of data, **I**, we should try to maximise the *simplicity* of **I** by reducing, as much as possible, repetition of information or *redundancy* in **I**, whilst retaining a percentage of its non-redundant descriptive or explanatory *power*: 100% in the case of lossless compression, or some smaller percentage in the case of lossy compression." In this connection:

  - As noted in a footnote to Section 2.2, the concepts of 'simplicity' and 'power' are the origin of the name "SP" for the *SP theory of intelligence*.
  - As previously noted, achieving a good combination of *simplicity* and *power* in a theory derived from a body of data, **I**, is equivalent to the lossless or lossy compression of **I**.
  - The concepts of 'simplicity' and 'power' in the SP theory appear to be similar to the concepts of 'simplicity' and 'strength' in [27, 63,

13]. However, a key difference between the two pairs of concepts is:

* On the one hand, [27, 13] refer to a "balance" between simplicity and strength, while [63] refers to a "balance" and a "trade-off" between simplicity and strength.
* On the other hand, by contrast, there is no "balance" or "trade-off" between simplicity and power in the SP theory. This is because, with respect to the compression of a target body of data, **I**: 1) 'simplicity' is about the extraction of redundant information from **I**; 2) 'power' is about the percentage of the non-redundant information in **I** that is to be retained; and 3) in any given body of information, **I**, redundant information in **I** and non-redundant information in **I** are mutually-exclusive subsets of **I**.

- *Algorithmic Probability Theory.* Ray Solomonoff [46] has argued that the great majority of problems in science *and* mathematics may be seen as either 'machine inversion problems'[10] or 'time limited optimization problems',[11] and that both kinds of problems may be solved via algorithmic compression from algorithmic probability theory [45],

- *Concepts of 'Pattern'.* In research relating to the fundamentals of science and of mathematics, concepts of 'pattern' have been discussed by some authors:

  - Daniel Dennett [17] writes about the concept of "real patterns" and links it to the concept of algorithmic compressibility in AIT. This and related concepts are further discussed at some length in, for example, [39].
  - James Ladyman and Don Ross [26] write that "The deep structure of the world is arguably not mathematical but statistical, ... The principles of statistics are generalizations of *recurrent patterns* found in data; and such structuring of data is the core business of both science and its metaphysical unification." (p. 108, emphasis added).

[10] "Inversion problems are the P and NP problems of computational complexity theory; i.e., given a machine M, that maps finite strings onto finite strings and given the finite string, x, how can we find in minimal time, a string, p, such that $M(p) = x$?" [46, p. 80].

[11] "Suppose we have a machine, M, whose inputs are finite strings, and whose outputs are numbers. We are given a fixed time T. The problem is to find within time T an input string, s, such that $M(s)$ is as large as possible." ([46, p. 81].

- As noted in Section 2.4, Keith Devlin [19] writes about *Mathematics: The Science of Patterns* in a book for non-specialists.

It seems that these three concepts of 'pattern' are, at least approximately, the same as the concept of 'R-pattern', or perhaps 'A-pattern', described in Section 2.1, reflecting informational redundancy in the domains from which they are drawn.

2.5.2 The possible role of MICMUPA in science

As with the philosophy of mathematics, it appears that there is nothing like MICMUPA in the philosophy of science:

- Apart from some brief discussions of Ockham's razor and the concept of 'simplicity' in science, a recent authoritative book on the philosophy of science [43] makes no mention of anything resembling MICMUPA anywhere.

- Similar things can be said about four other books about the philosophy of science: [47, 20, 25, 9]. Although these are described as "introductory" books, they are intended for university students and one would expect MICMUPA, or something like it, to be featured, at least briefly, if that kind of idea was recognised in the philosophy of science.

2.6 Seven techniques for the compression of information via the matching and unification of patterns

Readers who are acquainted with techniques for the compression of information will know that many of them, such as Huffman coding, arithmetic coding, and wavelet compression, have a mathematical flavour (see, for example, [41]). Much the same may be said about algorithmic compression in the framework of AIT [28].

Since ideas of that kind have a good pedigree and have proved their worth in many applications, one might suppose that they would be the starting point for any discussion of how mathematics may be understood in terms of information compression. But:

- The SP programme of research attempts to reach down below the mathematics of other approaches, to focus on the relatively simple, 'primitive' idea that information compression may be understood in terms of the matching and unification of patterns.

- In any discussion of the fundamentals of mathematics, it would not be appropriate to use mathematics itself.
- Since ICMUP is a relatively 'concrete' idea, less abstract than much of mathematics, it suggests avenues that may be explored in understanding possible mechanisms for information compression in artificial systems and in brains and nervous systems.

ICMUP and its variants may appear too childishly simple to merit attention in any discussion of the fundamentals of mathematics. But ICMUP is bedrock in the powerful concept of SP-multiple-alignment, outlined in Appendix A.1, which has proven capabilities, not only in modelling the six other versions of ICMUP described here [61, Appendix B], but also, more importantly, in diverse aspects of human intelligence, the representation of diverse forms of knowledge, and the seamless integration of diverse aspects of intelligence and diverse forms of knowledge, in any combination [54, 52]. And, as described in [60], ICMUP is prominent in HLPC.

While care has been taken in this programme of research to avoid unnecessary duplication of information across different publications, the importance of the following seven variants of ICMUP has made it necessary, for the sake of clarity, to describe them quite fully both in this paper and also in [60].

2.6.1 Basic ICMUP: information compression via the matching and unification of patterns

The simplest of the techniques to be described is to find two or more patterns that match each other within a body of information, **I**, and then merge or 'unify' them so that multiple instances are reduced to one. This is illustrated in the upper part of Figure 1 where two instances of the pattern '`INFORMATION`' near the top of the figure has been reduced to one instance, shown in the middle of the figure, with '`w62`' appended at the front, for reasons given in Section 2.6.2, below.

Here, and in subsections below, we shall assume that the single pattern which is the product of unification is placed in some kind of dictionary of patterns that is separate from **I**.

The version of ICMUP just described will be referred to as *basic ICMUP*.

A detail that should not distract us from the main idea is that, when compression of a body of information, **I**, is to be achieved via basic ICMUP, any repeating pattern that is to be unified should occur more often in **I** than one would expect by chance.

2.6.2 Chunking-with-codes

A point that has been glossed over in describing basic ICMUP is that, when a body of information, **I**, is to be compressed by unifying two or more instances of a pattern like '`INFORMATION`', there is a loss of information about the location within **I** of each instance of '`INFORMATION`'. In other words, basic ICMUP achieves 'lossy' compression of **I**.

This problem is overcome in the *chunking-with-codes* variant of ICMUP:

- A unified pattern like '`INFORMATION`', which is often referred to as a 'chunk' of information,[12] is stored in a dictionary of patterns, as mentioned in Section 2.6.1. A chunk in this context is much like an 'R-pattern' described in Section 2.1.

- As before, the unified chunk is given a relatively short name, identifier, or 'code', like the '`w62`' pattern appended at the front of the '`INFORMATION`' pattern in the middle of Figure 1.

- Then the '`w62`' code is used as a shorthand which replaces the '`INFORMATION`' chunk of information wherever it occurs within **I**. This is shown at the bottom of Figure 1.

- Since the code '`w62`' is shorter than each instance of the pattern '`INFORMATION`' which it replaces, the overall effect is to shorten **I**. But, unlike basic ICMUP, chunking-with-codes achieves 'lossless' compression of **I**.

- A detail here is that compression can be optimised by giving shorter codes to chunks that occur frequently and longer codes to chunks that are rare. This may be done using some such scheme as Shannon-Fano-Elias coding, described in, for example, [16].

2.6.3 Schema-plus-correction

A variant of the chunking-with-codes version of ICMUP is called *schema-plus-correction*. Here, the 'schema' is like a chunk of information and, as with chunking-with-codes, there is a relatively short identifier or code that may be used to represent the chunk.

What is different about the schema-plus-correction idea is that the schema may be modified or 'corrected' in various ways on different occasions.

[12]There is a little more detail about the concept of 'chunk' in [60, Section 2.4.2].

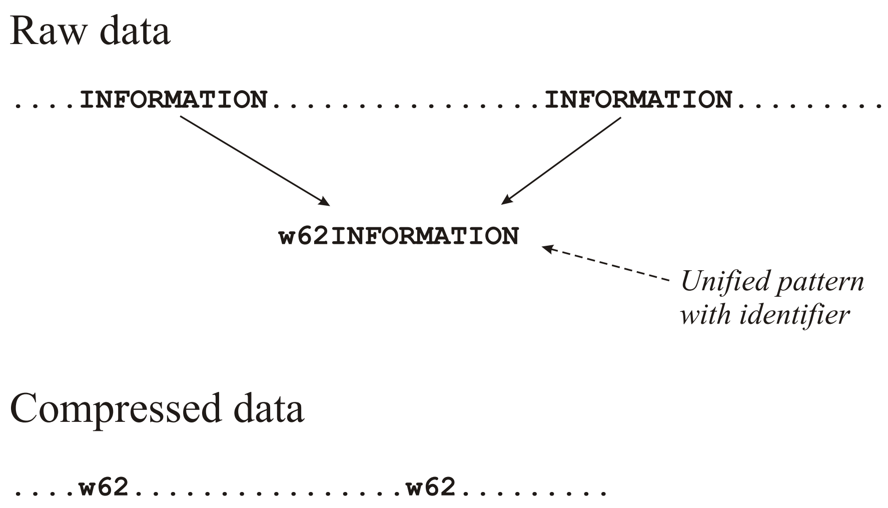

Figure 1: A schematic representation of the way two instances of the pattern '`INFORMATION`' in a body of data may be unified to form a single 'unified pattern', with 'w62' as an identifier assigned by the system. The lower part of the figure shows how the data may be compressed by replacing each instance of '`INFORMATION`' with a copy of the corresponding identifer. Reproduced with permission from Figure 2.3 in [52].

For example, a menu for a meal in a cafe or restaurant may be something like '`MN: ST MC PG`', where '`MN`' is the identifier or code for the menu, '`ST`' is a variable that may take values representing different kinds of 'starter', '`MC`' is a variable that may take values representing different kinds of 'main course', and '`PG`' is a variable that may take values representing different kinds of 'pudding'.

With this scheme, a particular meal may be represented economically as something like '`MN: ST(st2) MC(mc5) PG(pg3)`', where '`st2`' is the code or identifier for 'minestrone soup', '`mc5`' is the code for 'vegetable lassagne', and '`pg3`' is the code for 'ice cream'; another meal may be represented economically as '`MN: ST(st6) MC(mc1) PG(pg4)`', where '`st6`' is the code or identifier for 'prawn cocktail', '`mc1`' is the code for 'lamb shank', and '`pg4`' is the code for 'apple crumble'; and so on. Here, the codes for different dishes serve as modifiers or 'corrections' to the categories '`ST`', '`MC`', and '`PG`' within the schema '`MN: ST MC PG`'.

2.6.4 Run-length coding

A third variant, *run-length coding*, may be used where there is a sequence of two or more copies of a pattern, each one except the first following immediately after its predecessor like this:

‘`INFORMATIONINFORMATIONINFORMATIONINFORMATIONINFORMATION`’.

In this case, the multiple copies may be reduced to one, as before, with something to say how many copies there are (eg, ‘`(INFORMATION)`×5’), or where the repetition begins and ends (eg, ‘`[(INFORMATION)*]`’ where ‘`[`’ and ‘`]`’ are the beginning and end symbols, and ‘`*`’ signifies repetition), or, more vaguely, that the pattern is repeated without anything to say when the sequence stops (eg, ‘`(INFORMATION)*`’).

In a similar way, a sports coach might specify exercises as something like “touch toes (×15), push-ups (×10), skipping (×30), ...” or “Start running on the spot when I say ‘start’ and keep going until I say ‘stop’ ”.

With the ‘running’ example, “start” marks the beginning of the sequence, “keep going” in the context of “running” means “keep repeating the process of putting one foot in front of the other, in the manner of running”, and “stop” marks the end of the repeating process. It is clearly much more econonomical to say “keep going” than to constantly repeat the instruction to put one foot in front of the other.

2.6.5 Class-inclusion hierarchies

A widely-used idea in everyday thinking and elsewhere is the *class-inclusion hierarchy*: the grouping of entities into classes and the grouping of classes into higher-level classes through as many levels as are needed.

This idea may achieve ICMUP because, at each level in the hierarchy, attributes may be recorded which apply to that level and all levels below it. This can mean great economies because, for example, it is not necessary to record that cats have fur, dogs have fur, rabbits have fur, and so on—it is only necessary to record that mammals have fur and ensure that all lower-level classes and entities can ‘inherit’ that attribute. In effect, multiple instances of the attribute ‘fur’ have been merged or unified to create that attribute for mammals, thus achieving compression of information.[13]

This idea may be generalised to cross-classification, where any one entity or class may belong in one or more higher-level classes that do not have

[13] The concept of class-inclusion hierarchies with inheritance of attributes is quite fully developed in object-oriented programming, which originated with the Simula programming language [10] and is now widely adopted in modern programming languages.

the relationship superclass/subclass, one with another. For example, a given person may belong in the class 'woman' and 'doctor' although 'woman' is not a subclass of 'doctor' and *vice versa*.

2.6.6 Part-whole hierarchies

Another widely-used idea is the *part-whole hierarchy* in which a given entity or class of entities is divided into parts and sub-parts through as many levels as is needed. Here, ICMUP may be achieved because two or more parts of a class such as 'car' may share the overarching structure in which they all belong. So, for example, each wheel of a car, the doors of a car, the engine or a car, and so on, all belong in the same encompassing structure, 'car', and it is not necessary to repeat that enveloping structure for each individual part.

2.6.7 SP-multiple-alignment

The seventh version of ICMUP, the *SP-multiple-alignment* construct outlined in Appendix A.1, encompasses all the preceding six versions of ICMUP and much more besides.

How the preceding six versions of ICMUP may be modelled within the SP-multiple-alignment framework is described in [61, Appendix B]. The strengths and potential of the SP-multiple-alignment construct in modelling aspects of human intelligence and the representation of knowledge is outlined in [62] with pointers to where fuller information may be found. The potential of this construct in modelling aspects of mathematics is described in [52, Chapter 10].

3 Mathematics as a set of techniques for the compression of information via the matching and unification of patterns, and their application

The first step in the argument outlined in Section 1.1, depends on evidence that mathematics is fundamentally about the compression of information. The following subsections present evidence in support of this idea.

In what follows, the main emphasis is on ICMUP, as described in Section 2.6. This may be seen to have an impact on the structuring of mathematics, and on the dynamics of mathematical calculation or inference.

3.1 An example of information compression via mathematics

The equation $s = (gt^2)/2$, one of several that can be derived from Newton's Second Law of Motion, is a very compact means of representing any table, including large ones, showing the distance travelled by a falling object (s) in a given time since it started to fall (t), as illustrated in Table 1.[14] The constant, g, is the acceleration due to gravity—about $9.8m/s^2$. That small equation would represent the values in the table even if it was a 1000 times or a million times bigger, and so on. Likewise for other equations such as $a^2 + b^2 = c^2$, $PV = k$, $F = q(E + v \times B)$, and so on.

Distance (m)	*Time (sec)*
0.0	0
4.9	1
19.6	2
44.1	3
78.5	4
122.6	5
176.5	6
240.3	7
313.8	8
397.2	9
490.3	10
593.3	11
706.1	12
828.7	13
961.1	14
1103.2	15
1255.3	16
Etc	*Etc*

Table 1: The distance travelled by a falling object (metres) in a given time since it started to fall (seconds).

To make these points, it is not strictly necessary to show Table 1. But the table helps to emphasise the contrast between the potentially huge volumes of data in such a table and the small size of the equation which describes those data—and, correspondingly, the potentially high levels of information

[14] Of course, the law does not work for something like a feather falling in air.

compression that may be achieved with ordinary mathematics which is not specialised for compression of information.

3.2 How ICMUP may be seen in the structure and workings of mathematics

The subsections that follow describe how some of the basic principles and techniques for the compression of information that were outlined in Section 2.6 may be seen in the structure and workings of mathematics.

In themselves, these examples do not prove that mathematics may be understood as being entirely devoted to the compression of information. But there are reasons to think that compression of information is fundamental in mathematics:

- Since the techniques to be described are low-level techniques that are part of the foundations of mathematics and widely used in more complex forms of mathematics, it seems likely that mathematics may indeed be understood in its entirety to be a set of techniques for compressing information and their application.

- As described in Section 4.1.1, the workings of simple logical functions, including the NAND logical function, may be understood in terms of ICMUP. Since it is widely accepted that, in principle, the computational heart of any general-purpose digital computer may be constructed entirely from NAND gates [34], it appears that, within the bounds imposed by computational complexity, ICMUP has the generality to support any kind of computation, including mathematical computations.

3.3 Basic ICMUP

The 'basic' version of ICMUP, "basic ICMUP"(Section 2.6.1) may be seen in mathematics whenever one identifier is matched with another, with implicit unification of the two.

3.3.1 The matching and unification of identifiers

In mathematics, ICMUP may be seen wherever there is a need to invoke a named entity. If, for example, we want to calculate the value of z from these equations: $x = 4$; $y = 5$; $z = x + y$, we need to match the identifier x in the third equation with the identifier x in the first equation, and to unify the two so that the correct value is used for the calculation of z. Likewise for y.

In a similar way if we wish to invoke or 'call' a function such as '$\log x$' (the logarithm of a number), there must be a match between the name of the function in the call to the function (such as '$\log 1000$' and the name of the function in its definition, '$\log x$'. Unification of the call to the function with the definition of the function may be seen to have the effect of assigning the number in the call (1000 in this example) to the variable x in the definition of the function.

3.3.2 The execution of a function

At an abstract level, any function may be seen as a table in which each row shows the connection between one or more input values and one or more output values. And simple functions, such as a one-bit adder, may be specified in exactly that way, as shown in Table 2.

Input (1)	*Input (2)*	*Sum*	*Carry*
1	1	0	1
1	0	1	0
0	1	1	0
0	0	0	0

Figure 2: A table to define a function for the addition of two one-bit numbers in binary arithmetic, with provision for the carrying out of one bit.

In the workings of this adder, basic ICMUP may be seen, for example, in the matching and unification of input values '1' and '0' with corresponding values in 'input' columns of the table. In this case, the matches which achieve the greatest compression (both '1' and '0' in one row) will be to select the second row in the table, with the sum '1' and the carry digit '0', which are of course the correct outputs for those two inputs.

3.3.3 Matching and unification of patterns with peano's axiom for natural numbers

The sixth of Peano's axioms for natural numbers—for every natural number n, $S(n)$ is a natural number—provides the basis for a succession of numbers: $S(0)$, $S(S(0))$, $S(S(S(0)))$ ..., itself equivalent to unary numbers in which $1 = /$, $2 = //$, $3 = ///$, and so on. Here, S at one level in the recursive definition is repeatedly matched and unified with S at the next level.

3.4 Chunking-with-codes

This subsection describes aspects mathematics that may be seen to exemplify the chunking-with-codes technique for information compression, as described in Section 2.6.2.

3.4.1 Named functions

If a body of mathematics is repeated in two or more parts of something larger then it is natural to declare it once as a named 'function', where the body of the function may be seen as a 'chunk' of information, and the name of the function is its 'code' or identifier. This avoids the need to repeat that body of mathematics in two or more places.

An example of this kind of thing is the calculations needed to find the square root of a number, often provided as a ready-made square-root function with the non-alphabetic name '$\sqrt{x}$'. That name may be used to invoke the function wherever it is needed, like this: '$\sqrt{16}$'. Similar things may be done with functions such as '$\sin(x)$', '$\cos(x)$', and '$\log(x)$'.

Although they are not commonly seen as 'functions', all of the operations of addition, subtraction, multiplication, the power notation, and division, may be cast in that mould as, for example, '*plus*(x,y)', '*subtract*(x,y)', and so on. As such, they may be seen as examples of chunking-with-codes and schema-plus-correction (Section 3.5). As we shall see in Section 3.6, they may also be seen as examples of run-length coding.

3.4.2 The number system

Number systems with bases greater than 1, like the binary, octal, decimal and hexadecimal number systems, may all be seen to illustrate the chunking-with-codes technique for compressing information. For example:

- A unary number like '///////' may be referred to more briefly in the decimal system as '7'. Here, '///////' is the chunk and '7' is the code.
- A unary number like '/////////////////' may be split into two parts: '//////////' and '///////'. Then, in the decimal system, the first part may be represented by '1' and the second part by '7', giving us the decimal number '17'. The convention is that the right-most digit represents numbers less than 10, and the next digit to the left represents the number of 10s.

- Of course, this 'positional' system can be extended so that a digit in the third position from the right represents the number of 100s, a digit in the fourth position represents the number of 1000s, and so on.

Here, we can see how the chunking-with-codes technique allows us to eliminate the repetition or redundancy that exists in all unary numbers except '/'. This means that large numbers, like 2035723, may be expressed in a form that is very much more compact than the equivalent unary number.

3.5 Schema-plus-correction

Most functions in mathematics, like those mentioned above, are not only examples of chunking-with-codes: they are also examples of the schema-plus-correction device for compressing information. This is because they normally require input via one or more 'arguments' or 'parameters'. For example, the square root function needs a number like 49 for it to work on. Without that number, the function is a very general 'schema' for solving square root problems. With a number like 49, which may be regarded as a 'correction' to the schema, the function becomes focussed much more narrowly on finding the square root of 49.

3.6 Run-length coding

Run-length coding appears in various forms in mathematics, often combined with other things. The key idea is that some entity, pattern, or operation is repeated two or more times in an unbroken sequence. Here are some examples:

- Since all numbers with bases above 1 may be seen to be compressed representations of unary numbers (Section 3.4.2), unary numbers may be regarded as more fundamental than non-unary numbers. If that is accepted, then, for example, '$3+7$' may be seen as a shorthand for the repeated process of transferring one unary digit from a group of seven unary digits to a group of three unary digits. Thus the expression '$+7$' within '$3+7$' may be seen as an example of run-length coding.

  Subtraction may be interpreted in a similar way when a smaller number is subtracted from a larger number.

- Multiplication is repeated addition. So, for example, '3×10' is the 10-fold repetition of the operation '$x + 3$', where 'x' starts with the value '0'. Then '$\times 10$' within '3×10' may be seen as run-length coding.

Since addition is itself a form of run-length coding (as described in the preceding bullet point), multiplication may be seen as run-length coding on two levels.

- Division of a larger number by a smaller one (eg, '12/3') is repeated subtraction which, as with multiplication, may be seen as run-length coding. Of course there will be a 'remainder' if the larger number is not an exact multiple of the smaller number. As with addition as a part of multiplication, subtraction as a part of division means that division may be seen as run-length coding on two levels.

- The power notation (eg, '10^9') is repeated multiplication, and is thus another example of run-length coding. Since multiplication, as repeated addition, is a form of run-length coding, and since addition may be seen as run-length coding (the first bullet point above), the power notation may be seen as run-length coding on three levels!

- A factorial (eg, '25!') is repeated multiplication and subtraction.

- The bounded summation notation (eg, '$\sum_{i=1}^{5} \frac{1}{i}$') and the bounded power notation (eg, '$\prod_{n=1}^{10} \frac{n}{n-1}$') are shorthands for repeated addition and repeated multiplication, respectively. In both cases, there is normally a change in the value of one or more variables on each iteration, so these notations may be seen as a combination of run-length coding and schema-plus-correction.

- In matrix multiplication, 'AB', for example, is a shorthand for the repeated operation of multiplying each entry in matrix 'A' with the corresponding entry in matrix 'B'.

3.7 Class-inclusion hierarchies

Classes and subclasses (Section 2.6.5) feature in mathematics as 'sets', both as a sometimes-disputed foundation for mathematics and as a branch of mathematics.

The notion of 'inheritance' does not have the prominence in set theory that it does in object-oriented programming, but, nevertheless, ICMUP may be seen in other concepts associated with sets, described in Section 4.1.

3.8 Part-whole hierarchies

It seems that part-whole hierarchies are not much used in mathematics, except perhaps in set theory, but, as we shall see in Section 4.2, they are quite prominent in the mathematics-related discipline of computing.

3.9 SP-multiple-alignment

Preliminary work described in [52, Chapter 10] shows that the SP system, with SP-multiple-alignment centre-stage, has potential to model mathematical constructs and mathematical processes. This should not be altogether surprising since, as noted in Section 2.6.7, SP-multiple-alignments can do everything that can be done with the six variants of ICMUP described in Sections 2.6.1 to 2.6.6, and it provides for their seamless integration too.

Other reasons for believing that the SP system has potential to model many and perhaps all concepts and processes in mathematics are:

- The generality of information compression as a means of representing knowledge in a succinct manner.
- The central role of information compression in the SP-multiple-alignment framework.
- The versatility of the SP-multiple-alignment framework in aspects of intelligence and the representation of knowledge (Appendix A.2).
- The close connection that is known to exist between information compression and concepts of prediction and probability (Section 6).

3.10 Some equations

It seems that most equations that have become established in mathematics and science may be interpreted in terms of some combination of the techniques for compressing information described in Section 2.6. Thus:

- Einstein's equation, $E = mc^2$, illustrates run-length coding in its power notation (c^2) and in the multiplication of m with c^2.
- Newton's equation, $s = (gt^2)/2$, that featured in Section 3.1, illustrates run-length coding in its power notation (t^2), in the multiplication of g with t^2, and in the division of (gt^2) by 2.
- Pythagoras's equation, $a^2 + b^2 = c^2$, illustrates run-length coding via the power notation in a^2, b^2, and c^2, and via the addition of b^2 to a^2 (the first bullet point in Section 3.6).
- Boyle's law, $PV = k$, illustrates run-length coding in the multiplication of P by V.

- The charged particle equation, $F = q(E + v \times B)$, illustrates run-length coding in the multiplication of v by B, in the multiplication of $(E + v \times B)$ by q, and in the addition of $v \times B$ to E.
- One of special relativity's equations for time dilation, $\Delta t' = \Delta t/\sqrt{1 - v^2/c^2}$, illustrates chunking-with-codes and schema-plus-correction in its use of the square root function, and it illustrates run-length coding in the division of v^2 by c^2, in the subtraction of v^2/c^2 from 1, and in the division of Δt by $\sqrt{1 - v^2/c^2}$.
- In its use of bounded summation ($\sum$), Shannon's equation for entropy, $H = -\sum_i p_i \log_2(p_i)$, illustrates a combination of run-length coding and schema-plus-correction (as noted in Section 3.6). It also illustrates chunking-with-codes in its use of the $\log_2$ notation.

Since, addition, subtraction, multiplication, the power notation, and division, may each be seen as an example of chunking-with-codes and schema-plus-correction (Sections 3.4 and 3.5), as well as run-length coding (Section 3.6), the same can be said about the appearance of those notations in each of the examples above.

4 Mathematics-related disciplines as techniques for the compression of information, and their application

It seems that, to a large extent, what has been said about mathematics in Section 3 also applies to the mathematically-related disciplines of logic and computing.[15] The following two subsections present some examples in support of that idea.

4.1 Logic

Subsections that follow describe some evidence for ICMUP in logic.

4.1.1 XOR and other logical operations

The XOR logical function, and other simple logical functions, may be defined and interpreted in much the same way as the one-bit adder shown in Figure 2, as shown in Table 3.

[15] Where computing has its modern sense of computation by machine.

Input (1)	*Input (2)*	*Output*
1	1	0
0	1	1
1	0	1
0	0	0

Figure 3: A table to define the XOR logical function.

As with the one-bit adder, the operation of the XOR function may be understood in terms of basic ICMUP. Input values such as 1 (first) and 0 (second) may be matched and unified with values in the corresponding 'input' columns of the table. With those two input values, the third row is selected because it yields most matches—which, with unification, also means the greatest compression of information. And of course the third row yields the correct output value, which in this example is 1.

There are two points of interest here:

- *The XOR Function and Artificial Neural Networks.* As is well known, Marvin Minsky and Seymour Papert [32] demonstrated that basic perceptrons of the kind that were available in the late 1960s could not produce correct results with the XOR function, a demonstration which, for a time, led to a fall in interest in artificial neural networks.

- *The Generality of the NAND Logical Function.* As noted in Section 3.2, the fact that the NAND logical function may, like XOR and other simple logical functions, be understood in terms of ICMUP, and the generally-accepted idea that the computational heart of any general-purpose computer may, in principle, be constructed entirely from NAND gates, provide evidence in support of the idea that compression of information is fundamental in all kinds of computation including mathematical computations.

4.1.2 Deriving a set from a multiset

In logic and mathematics, a 'multiset' or 'bag' is like a set but any element within the multiset may be repeated as, for example, in the multiset {a, b, a, c, b, b, c, a, c}.

Conversion of any such multiset into the corresponding set means matching each element within the multiset with every other element and, wherever a match is found, unifying the two elements, including elements that are the result of earlier unifications, thus achieving ICMUP. In this case, the multiset {a, b, a, c, b, b, c, a, c} is reduced to the set {a, b, c}.

4.1.3 The union and intersection of sets

In much the same way that a set may be derived from a multiset (Section 4.1.2), the union and intersection of two sets may be found by the matching and unification of elements, yielding a reduction in the overall size of the two sets when unification has been achieved. Thus, for example, the union of the sets {b, f, d, a, c, e} and {e, g, i, f, d, h} is {a, b, c, d, e, f, g, h, i}, with the intersection {d, e, f}. In accordance with ICMUP, the union is smaller than the two sets from which it was derived.

4.1.4 ICMUP in Prolog

Further evidence for the significance of ICMUP in logic is that systems like Prolog—a computer-based version of logic—may be seen to function largely via the matching and merging of patterns.

Here, the meaning of 'unification' in Prolog—comparing two terms to see if they can be made to represent the same structure—is quite close to the meaning of 'unification' in this paper.

4.1.5 Versatility in reasoning with the SP system

Since SP-multiple-alignment is a generalised form of ICMUP (Section 2.6.7), and since SP-multiple-alignment is an important part of the SP system, it is pertinent to say that the SP computer model demonstrates several kinds of reasoning including: one-step 'deductive' reasoning; chains of reasoning; abductive reasoning; reasoning with probabilistic networks and trees; reasoning with 'rules'; nonmonotonic reasoning and reasoning with default values; Bayesian reasoning with 'explaining away'; causal reasoning; reasoning that is not supported by evidence; the inheritance of attributes in class hierarchies; and inheritance of contexts in part-whole hierarchies [52, Chapter 7].

Because of the probabilistic nature of the SP system, these forms of reasoning are probabilistic, although some of them, such as one-step 'deductive' reasoning, have the all-or-nothing character of traditional forms of logic. Nevetheless, if it is accepted that logic, like mathematics, is probabilistic at a deep level—for reasons given in Section 6—then the above-mentioned strengths of the SP system in probabilistic reasoning may be seen as further evidence for the importance of ICMUP in logic.

4.2 Computing

As with logic, it seems likely that, since computing is closely related to mathematics, it may, like mathematics, be understood in terms of ICMUP.

Evidence in support of that view is presented in subsections that follow.

4.2.1 Matching and unification of patterns in definitions of 'computing'

Emil Post's [38] "Canonical System", which is recognised as a definition of 'computing' that is equivalent to a universal Turing machine, may be seen to work largely via the matching and unification of patterns [52, Chapter 4].

Much the same is true of the workings of the transition function in a universal Turing machine. This is essentially a look-up table like that shown in Table 2.

Input (1)	*Input (2)*	*Output (1)*	*Output (2)*
s_0	1	s_0	»
s_0	0	s_1	1
s_1	1	s_1	«
s_1	0	s_2	»

Table 2: An example of a transition function in a universal Turing machine, represented as a look-up table, as described in the text. *Key:* '»' means "move the read/write head one place to the right"; '«' means "move the read/write head one place to the left". Based on the example in [5, Section 2], with permission.

Much as with the examples described in Sections 3.3.2 and 4.1.1, ICMUP may be seen, for example, in the matching and unification of input values 's_1' and '1' with corresponding values in the input columns of the table. In this case, the effect will be to select the third row in the table, with the output values 's_1' and '«'—which mean "Set the state of the machine to 's_1' and move the read/write head of the machine one place to the left".

In a similar way, ICMUP may be seen in the workings of the NAND logical function which, as noted in Sections 3.2 and 4.1.1 may in principle provide the computational heart of any general-purpose digital computer.

4.2.2 Some other examples of ICMUP in computing

Here, in brief, are some other putative examples of ICMUP in computing:

- *Basic ICMUP*. As in mathematics (Section 3.3), basic ICMUP may be seen in computing in the matching of identifiers for variables and in calls to functions.

- *Chunking-With-Codes and Schema-Plus-Correction.* Again, as in Section 3.4, named functions in computing may be seen as examples of the chunking-with-codes version of ICMUP, and as in Section 3.5, functions with parameters may be seen as examples of the schema-plus-correction version of ICMUP.

- *Run-Length Coding.* As in mathematics (Section 3.6), run-length coding may be seen in computing in the basic arithmetic functions. It may also be seen in iteration statements like *while ...*, *do ... while ...*, *for ...*, or *repeat ... until ...*. It may also be seen in the use of recursion in functions such as *factorial(x)* for the calculations of the factorial of any number.

- *Class-Inclusion and Part-Whole Hierarchies.* In computing, the creation of classes and hierarchies of classes is supported in such object-oriented programming languages as Simula, Smalltalk, C++, and many more. Part-while hierarchies are also prominent in software. In both cases, ICMUP has a role to play, much as described in Sections 2.6.5 and 2.6.6.

- *Retrieving Data From Computer Memory.* It is true that electronic circuits provide the mechanism for finding an address in computer memory but, at a more abstract level, the process may be seen as one of searching for a match between the address held in the CPU and the corresponding address in computer memory. When a match has been found between the address in the CPU and the corresponding address in memory, there is implicit unification of the two.

- *Query-by-Example.* A popular technique for retrieving information from databases, query-by-example, is essentially a process of finding good matches between a query pattern and patterns in the database, with unification of the best matches.

5 An apparent solution to the mystery of why mathematics is so effective in science

In view of evidence that: 1) science is fundamentally a search for compression in the world (Section 2.5); and evidence that 2) mathematics may be seen to be largely a set of techniques for compressing information and their application (Section 3); and bearing in mind 3) the afore-mentioned intimate relation between information compression and concepts of prediction and probability

(Section 6); it seems reasonable to conclude that those three things may explain the effectiveness of mathematics as a means of representing scientific knowledge and in the making of scientific inferences.

An objection to the arguments above is that, even if we accept that mathematics provides an effective means of compressing information, we are still left with the problem of explaining why nature is so often compressible. Two possible answers are suggested in the following two subsections.

5.1 The anthropic principle

A possible answer to the problem just mentioned is that:

> "... maths is best thought of as the catalogue of all possible patterns and so it is inevitable that mathematics is effective in describing the world—it could not be otherwise because the world must have pattern to allow life to exist."[16]

And this appeal to the anthropic principle[17] may be adapted for information compression as something like: "the world must be at least partly compressible because otherwise everything, including ourselves, would be a soup of randomness."

5.2 The contingency principle

Another possible answer to the problem mentioned above is that, in view of the observable fact that at least some parts of the world (the observable universe) are compressible, we should be prepared to deal with those parts of the world, if or when we encounter them—in much the same way that a Swiss army knife can be a useful thing to carry around, even though we may never use such tools as the tweezers, the magnifying glass, and so on.

6 Mathematics, prediction, and probabilities

The main focus of this paper is on MICMUPA, but it is relevant to mention that it has been recognised for some time that there is an intimate connection between information compression and concepts of prediction and probability, as described in Ray Solomonoff's algorithmic probability theory (APT)

[16] John Barrow, personal communication, 2017-04-06, with permission.

[17] The anthropic principle, with several of its variants, and with associated controversies, is well described in "Anthropic principle", *Wikipedia*, bit.ly/2pVf1W8, retrieved 2017-11-24.

[45, 46], and in the closely-related AIT [28]. Information compression and concepts of prediction and probability may be seen as two sides of the same coin.

The close connection between those things makes sense in terms of ICMUP (Section 2.6):

- A pattern that repeats is one that, via inductive reasoning, we naturally regard as a guide to what may happen in the future.

- A pattern that repeats is one that, via the merging or unification of patterns, may yield compression of information.

- A partial match between one pattern and another can be the basis for predicting the occurrence of the unmatched parts, a form of inference that is sometimes called *prediction by partial matching* [49].

From Solomonoff's work and associated research, we may conclude that compression of information is important in science, partly as a means of representing scientific knowledge in a succinct form, but at least as important is how information compression can be a key to the making of scientific inferences and the calculation of probabilities of such inferences.

What has this got to do with mathematics? It would take us too far afield to discuss this issue in any depth. A few brief remarks are made here. The close connection between information compression and concepts of prediction and probability, and evidence for MICMUPA presented in this paper, suggests that:

- Notwithstanding the apparent certainty of equations like $2 + 2 = 4$, mathematics may be seen to be fundamentally probabilistic.

- In view of the important role that mathematics has in the making of inferences in science and elsewhere, and notwithstanding the apparent certainties of many of those inferences, MICMUPA may be seen as a driver for the making of 'exact' inferences.

Regarding the first point, a probabilistic foundation for mathematics is consistent with the discovery of randomness in number theory:

> "I have recently been able to take a further step along the path laid out by Gödel and Turing. By translating a particular computer program into an algebraic equation of a type that was familiar even to the ancient Greeks, I have shown that there is randomness in the branch of pure mathematics known as number theory. My work indicates that—to borrow Einstein's metaphor—God sometimes plays dice with whole numbers." [12, p. 80].

As indicated in this quotation, randomness in number theory is closely related to Gödel's incompleteness theorems. These are themselves closely related to the phenomenon of recursion, a feature of many formal systems, several of Escher's pictures, and much of Bach's music, as described in some detail by Douglas Hofstadter in *Gödel, Escher, Bach: An Eternal Golden Braid* [23].

With regard to the second point, it seems possible that, although mathematics may be fundamentally probabilistic, it may, with appropriate data or under appropriate conditions, deliver results where the associated probabilities are at or very close to 0 or 1. This kind of possibility is discussed briefly in [52, Sections 10.4.5 and 10.4.6].

7 so what?

While it may be accepted that this paper provides an explanation for the "mysterious" or "unreasonable" effectiveness of mathematics in science, readers may, nevertheless, wonder what implications there may be, if any, for the development of mathematics or science or both of them. Here are some possibilities:

- Since SP-multiple-alignment is a generalised version of ICMUP that encompasses the first six versions described in Section 2.6, and since SP-multiple-alignment is central in the SP theory of intelligence, evidence presented in this paper strengthens the already-strong position of the SP theory as a unifying principle across artificial intelligence [54, 52, 59], mainstream computing (*ibid.*), mathematics ([52, Chapter 10]), HLPC [60], and neuroscience [58].

- In connection with the previous bullet point, there are many potential benefits and applications of the SP theory, including [57, 56, 55], and others summarised in [62, Appendix A.7], with pointers to where fuller information may be found.

- With regard to mathematics and its effectiveness in science, there appears to be potential for the development of a "new mathematics for science" by treating the SP system as one of the techniques for the compression of information that are available within mathematics. Such a development, mentioned near the end of Section 2.3 above, may facilitate the automatic or semi-automatic development of new theories in science, and the qualitative and quantitative evaluation of scientific theories, both existing theories and new ones. Such a development may also broaden the scope of mathematics as a means of describing scientific observations and concepts, and as a means of drawing inferences about those observations and concepts across many areas of science.

8 Conclusion

This paper notes first that the effectiveness of mathematics in science appears to some writers to be “mysterious” or “unreasonable”. Then reasons are given for thinking that science is fundamentally a search for compression of empirical data. At more length, several reasons are given for believing that mathematics, and related disciplines, are, fundamentally, sets of techniques for information compression via the matching and unification of patterns (ICMUP), and their application. From there, it is argued that mathematics has proved to be effective in science because, with other modes of expression, it provides a means of achieving the compression of information which lies at the heart of science.

The anthropic principle provides an explanation of why we find the world—aspects of it at least—to be compressible.

ICMUP may be seen to be important in both science and mathematics, not only as a means of representing knowledge succinctly, but as a basis for scientific and mathematical inferences—because of the intimate relation that is known to exist between information compression and concepts of prediction and probability.

Since SP-multiple-alignment is a generalised version of ICMUP that encompasses six other versions described in the paper, and since SP-multiple-alignment is central in the SP theory of intelligence, evidence presented in this paper strengthens the already-strong position of the SP theory as a unifying principle across artificial intelligence, mainstream computing, mathematics, human learning, perception, and cognition, and neuroscience. Indirectly, it strengthens evidence for the many potential benefits and applications of the SP theory.

These ideas may also provide the basis for a “new mathematics for science” with potential benefits and applications in science and science-related areas.

Acknowledgements

I’m grateful to Roger Penrose, John Barrow, and anonymous reviewers for very helpful comments on earlier versions of this paper. For helpful comments on drafts of a related paper, I'm grateful to Robert Thomas, Michele Friend, and Alex Paseau. I'm also grateful for discussion from time to time of some of the ideas in this paper with Tim Porter and Chris Wensley.

A The SP theory of intelligence and the SP computer model

As noted in Section 2.2, much of the thinking in this paper derives from the *SP theory of intelligence* and its realisation in the *SP computer model*. This theory, which is described quite fully in [52] and more briefly in [54], aims to simplify and integrate observations and concepts across artificial intelligence, mainstream computing, mathematics, and HLPC.

Several other papers in the SP programme of research, most with download links, may be found via www.cognitionresearch.org/sp.htm.

The SP theory is conceived as a brain-like system that receives *New* information via its senses and stores some or all of it, in compressed form, as *Old* information.

In the SP system, all kinds of knowledge are stored as arrays of atomic *SP-symbols* called *SP-patterns*, mentioned in Section 2.1. At present, the SP computer model works only with one-dimensional SP-patterns but it is envisaged that it will be generalised to work with two-dimensional SP-patterns, in addition to 1D SP-patterns.

A.1 SP-multiple-alignment

A key idea in the SP system is the concept of *SP-multiple-alignment* borrowed and adapted from the concept of 'multiple alignment' in bioinformatics.

An example of a multiple alignment from bioinformatics is shown in Figure 4. Here, five DNA sequences have been arranged in rows and, by judicious "stretching" of sequences in a computer, matching symbols have brought into line. A 'good' multiple alignment is one with a relatively large number of matching symbols.

The key difference between the concept of multiple alignment in bioinformatics and the concept of SP-multiple-alignment is that, in the latter, a 'good' SP-multiple-alignment is one that allows one New SP-pattern (sometimes more than one) to be encoded economically in terms of one or more Old SP-patterns.

An example of an SP-multiple-alignment is shown in Figure 5. Here, the New SP-pattern is the sentence '`t w o k i t t e n s p l a y`' shown in row 0. Each of rows 1 to 8 shows an Old SP-pattern representing a

```
  G G A     G     C A G G G A G G A     T G     G   G G A
  | | |     |     | | | | | | | | |     | |     |   | | |
  G G | G   G C C C A G G G A G G A     | G G C G   G G A
  | | |     | | | | | | | | | | | |     | |     |   | | |
A | G A C T G C C C A G G G | G G | G C T G G A | G A
  | | |           | | | | | | | | |   |   |     |   | | |
  G G A A         | A G G G A G G A   | A G     G   G G A
  | |   |         | | | | | | | |     |   |     |   | | |
  G G C A         C A G G G A G G     C   G     G   G G A
```

Figure 4: A ‘good’ multiple alignment amongst five DNA sequences. Reproduced with permission from Figure 3.1 in [52].

grammatical structure, which in rows 1, 3 and 5 is a word. The overall effect of the SP-multiple-alignment is to analyse or parse the sentence into its constituent parts, each one marked with its grammatical category.

```
0                      t w o              k i t t e n     s                  p l a y           0
                       | | |              | | | | | |     |                  | | | |
1                      | | |         Nr 5 k i t t e n #Nr |                  | | | |           1
                       | | |         |                 |  |                  | | | |
2                      | | |    N Np Nr               #Nr s #N               | | | |           2
                       | | |    | |                         |                | | | |
3               D Dp 4 t w o #D | |                         |                | | | |           3
                |            |  | |                         |                | | | |
4            NP D            #D N |                         #N #NP           | | | |           4
             |                    |                             |            | | | |
5            |                    |                             |       Vr 1 p l a y #Vr       5
             |                    |                             |       |             |
6            |                    |                             |  V Vp Vr           #Vr #V    6
             |                    |                             |  | |                    |
7 S Num    ; NP                   |                            #NP V |                   #V #S 7
     |     |                      |                                  |
8   Num PL ;                      Np                                 Vp                        8
```

Figure 5: The best SP-multiple alignment created by the SP computer model with a store of Old SP-patterns like those in rows 1 to 8 (representing grammatical structures, including words) and a New SP-pattern, ‘ `t w o k i t t e n s p l a y`’, shown in row 0 (representing a sentence to be parsed). Adapted from Figures 1 in [53], with permission.

It turns out that the concept of SP-multiple-alignment within the SP system can do much more than the parsing of sentences, as indicated in the next subsection.

In connection with the concepts of pattern noted in Section 2.1, a New SP-pattern like ‘`t w o k i t t e n s p l a y`’ in row 0 of Figure 5 may be seen as an example of an ‘E-pattern’; an Old SP-pattern like ‘`Vr 1 p l a y #Vr`’ in row 5 of the figure may be seen as an ‘R-pattern’; and an Old SP-pattern like ‘`S Num ; NP #NP V #V #S`’ in row 7 of the figure may be seen as an example of an ‘A-pattern’. Old SP-patterns will normally include ‘code’ symbols, much like code symbols in Section 2.6.2 but, especially in the more abstract SP-patterns, with a wider range of functions.

A.2 Strengths of the SP system

Distinctive features and advantages of the SP system compared with other AI-related systems are described in [59].

Key strengths of the SP system, due mainly to the powerful concept of SP-multiple-alignment, are: versatility in aspects of intelligence (unsupervised learning, the analysis and production of natural language, pattern recognition that is robust in the face of errors, pattern recognition at multiple levels of abstraction, several kinds of reasoning, and more); versatility in the representation of knowledge (the syntax of natural languages; class-inclusion hierarchies (with or without cross classification); part-whole hierarchies; discrimination networks and trees; and more); and seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination. More detail may be found in [62] and in other sources cited there.

A.3 Potential applications

The SP system has potential in several areas of application including: helping to solve nine problems with big data; helping in the development of human-like intelligence in autonomous robots; helping in the understanding of human vision and in the development of computer vision; functioning as a database system with intelligence; and more. There is more detail in [62, A.7] with pointers to where fuller information may be found.

A.4 SP-neural

Key concepts in the SP theory may be mapped on to structures of neurons and their interconnections in a version of the SP theory called SP-neural [58].

B Related issues

This section considers some issues related to the idea that mathematics may be seen as a set of techniques for the compression of inforamtion, and their application.

B.1 The apparent paradox of creating redundancy via information compression

The idea that mathematics or computing is largely, perhaps entirely, about compression of information may seem to conflict with the undoubted fact that, with some simple mathematics or a simple computer program, it is possible to create data containing large amounts of repetition or redundancy.

This issue and how it may be resolved is discussed in [60, Appendix C.1].

B.2 Redundancy is often useful in the storage and processing of information

There is no doubt that informational redundancy—repetition of information—is often useful. For example, it is standard practice in computing to maintain two or more copies of important data, and redundancy in messages can provide a useful means of correcting errors. These kinds of uses of redundancy may seem to conflict with the idea that information compression—which means reducing redundancy—is fundamental in mathematics, computing and cognition.

This issue and how it may be resolved is discussed in [60, Appendix C.2].

References

[1] A. Baker. Are there genuine mathematical explanations of physical phenomena? *Mind*, 114(454):223–238, 2005.

[2] A. Baker. Mathematical explanation in science. *British Journal for the Philosophy of Science*, 60(0):611–633, 2009.

[3] A. Baker. Simplicity. In E. N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Winter 2016 edition edition, 2016. First published 2004-10-29; substantive revision 2016-12-20.

[4] E. C. Banks. The philosophical roots of Ernst Mach's economy of thought. *Synthese*, 139:25–53, 2004.

[5] D. Barker-Plummer. Turing machines. In E. N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Winter 2012 edition edition, 2012. First published 1995-09-14; substantive revision 2012-06-26.

[6] J. D. Barrow. *Pi in the Sky*. Penguin Books, Harmondsworth, 1992.

[7] R.W. Batterman. On the explanatory role of mathematics in empirical science. *The British Journal of Philosophy of Science*, 61(1):1–25, 2010.

[8] E. T. Bell. *The Handmaiden of the Sciences.* The Williams & Wilkins Company, Baltimore, 1937.

[9] A. Bird. *Philosophy of Science.* Routledge, London, Kindle edition, 1998.

[10] G. M. Birtwistle, O-J Dahl, B. Myhrhaug, and K. Nygaard. *Simula Begin.* Studentlitteratur, Lund, 1973.

[11] O. Bueno and M. Colyvan. An inferential conception of the application of mathematics. *Noûs*, 45(2):345–374, 2011.

[12] G. J. Chaitin. Randomness in arithmetic. *Scientific American*, 259(1):80–85, 1988.

[13] J. Cohen and C. Callender. A better best system account of lawhood. *Philosophical Studies*, 145:1–34, 2009.

[14] M. Colyvan. *The Indispensibility of Mathematics.* Oxford University Press, Oxford, 2001.

[15] M. Colyvan. The undeniable effectiveness of mathematics in the special sciences. In M. C. Galavotti, D. Dieks, W.J. Gonzalez, S. Hartman, T. Uebel, and M. Weber, editors, *New Directions in the Philosophy of Science*, pages 63–73. Springer, Heidelberg, 2014.

[16] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley, New York, 1991.

[17] D. C. Dennett. Real patterns. *The Journal of Philosophy*, 88(1):27–51, 1991.

[18] K. Devlin. *Logic and Information.* Cambridge University Press, Cambridge, 1991.

[19] K. Devlin. *Mathematics: The Science of Patterns.* Scientific American Library, New York, 1997.

[20] P. Godfrey-Smith. *Theory and Reality: An Introduction to the Philosophy of Science.* The University of Chicago Press, Chicago, Kindle edition, 2003.

[21] I. Grattan-Guinness. Solving wigner's mystery: the reasonable (though perhaps limited) effectiveness of mathematics in the natural sciences. *The Mathematical Intelligencer*, 30(3):7–17, 2008.

[22] R.W. Hamming. The unreasonable effectiveness of mathematics. *American Mathematics Monthly*, 87(2):81–90, 1980.

[23] D. Hofstadter. *Gödel, Escher, Bach: an Eternal Golden Braid.* Penguin Books, Harmondsworth, 1980.

[24] W. Isaacson. *Einstein: His Life and Universe.* Pocket Books, London, Kindle edition, 2007.

[25] J. Ladyman. *Understanding Philosophy of Science.* Routledge, London, Kindle edition, 2002.

[26] J. Ladyman and D. Ross. The world in the data. In D. Ross, J. Ladyman, and H. Kincaid, editors, *Scientific Metaphysics*, pages 108–150. Oxford University Press, Oxford, 2013.

[27] D. Lewis. Humean supervenience debugged. *Mind, New Series*, 103(412):473–490, 1994.

[28] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer, New York, 3rd edition, 2014.

[29] Ø. Linnebo. *Philosophy of Mathematics.* Princeton University Press, Princeton and Oxford, Kindle edition, 2017.

[30] A. Lyon and M. Colyvan. The explanatory power of phase spaces. *Philosophia Mathematica (3)*, 16(3):227–243, 2008.

[31] R. Marcus and M. McEvoy, editors. *An Historical Introduction to the Philosophy of Mathematics.* Bloomsbury Academic, London, Kindle edition, 2016.

[32] M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry. ISBN 0-262-63022-2.* MIT Press, Cambridge, Mass., 1969.

[33] I. Newton. *The Mathematical Principles of Natural Philosophy.* The Perfect Library, Kindle edition, 2014. First published 1687. Illustrated and bundled with *Life of Sir Isaac Newton.*

[34] N. Nisan and S. Schocken. *The Elements of Computing Systems: Building a Modern Computer from First Principles.* The MIT Press, Cambridge, Mass., Kindle edition, 2005.

[35] C. Parsons. *Philosophy of Mathematics in the Twentieth Century.* Harvard University Press, Cambridge, Mass., Kindle edition, 2014.

[36] K. Pearson. *The Grammar of Science.* Walter Scott, London, 1892. Republished by Dover Publications, 2004, ISBN 0-486-49581-7. Internet archive: bit.ly/1g2gNfk.

[37] R. Penrose. *The Emperor's New Mind.* Oxford University Press, Oxford, 1989.

[38] E. L. Post. Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics*, 65:197–268, 1943.

[39] D. Ross, J. Ladyman, and J. Collier. Rainforest realism and the unity of science. In J. Ladyman, D. Ross, D. Spurrett, and J. Collier, editors, *Everything Must Go: Metaphysics Naturalized*, pages 190–257. Oxford University Press, Oxford, 2007.

[40] C. Rovelli. *Reality Is Not What It Seems: The Journey to Quantum Gravity.* Penguin Books, London, Kindle edition, 2016.

[41] K. Sayood. *Introduction to Data Compression.* Morgan Kaufmann, Amsterdam, 2012.

[42] D. B. Schulz. *Simplicity in Science.* Iowa Research Online, Iowa City, 2012. bit.ly/2ijOG59.

[43] Gerhard Schurz. *Philosophy of Science: A Unified Approach.* Routledge, Abingdon, Kindle edition, 2014.

[44] L. Scorzato. On the role of simplicity in science. *Synthese*, 190:2867–2895, 2013.

[45] R. J. Solomonoff. A formal theory of inductive inference. Parts I and II. *Information and Control*, 7:1–22 and 224–254, 1964.

[46] R. J. Solomonoff. The discovery of algorithmic probability. *Journal of Computer and System Sciences*, 55(1):73–88, 1997.

[47] K. W. Staley. *An Introduction to the Philosophy of Science.* Cambridge University Press, Cambridge, Kindle edition, 2014.

[48] M. Steiner. *The Applicability of Mathematics as a Philosophical Problem.* Harvard University Press, Cambridge, Mass., 1998.

[49] W. J. Teahan and K. M. Alhawiti. Preprocessing for ppm: compressing utf-8 encoded natural language text. *International Journal of Computer Science & Information Technology*, 7(2):41–51, 2015.

[50] E. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13:1–14, 1960.

[51] M. Wilson. The unreasonable uncooperativeness of mathematics in the natural sciences. *Monist*, 83(2):296–314, 2000.

[52] J. G. Wolff. *Unifying Computing and Cognition: the SP Theory and Its Applications.* CognitionResearch.org, Menai Bridge, 2006. ISBNs: 0-9550726-0-3 (ebook edition), 0-9550726-1-1 (print edition). Distributors, including Amazon.com, are detailed on bit.ly/WmB1rs.

[53] J. G. Wolff. Towards an intelligent database system founded on the SP theory of computing and cognition. *Data & Knowledge Engineering*, 60:596–624, 2007. bit.ly/1CUldR6, arXiv:cs/0311031.

[54] J. G. Wolff. The SP theory of intelligence: an overview. *Information*, 4(3):283–341, 2013. bit.ly/1NOMJ6l, arXiv:1306.3888.

[55] J. G. Wolff. Autonomous robots and the SP theory of intelligence. *IEEE Access*, 2:1629–1651, 2014. bit.ly/18DxU5K, arXiv:1409.8027.

[56] J. G. Wolff. Big data and the SP theory of intelligence. *IEEE Access*, 2:301–315, 2014. bit.ly/2qfSR3G, arXiv:1306.3890. This paper, with minor revisions, is reproduced in Fei Hu (Ed.), *Big Data: Storage, Sharing, and Security (3S)*, Taylor & Francis LLC, CRC Press, 2016, pp. 143–170.

[57] J. G. Wolff. The SP theory of intelligence: benefits and applications. *Information*, 5(1):1–27, 2014. bit.ly/1FRYwew, arXiv:1307.0845.

[58] J. G. Wolff. Information compression, multiple alignment, and the representation and processing of knowledge in the brain. *Frontiers in Psychology*, 7:1584, 2016. bit.ly/2esmYyt, arXiv:1604.05535.

[59] J. G. Wolff. The SP theory of intelligence: its distinctive features and advantages. *IEEE Access*, 4:216–246, 2016. bit.ly/2qgq5QF, arXiv:1508.04087.

[60] J. G. Wolff. Evidence for information compression via the matching and unification of patterns in human leaning, perception, and cognition. Technical report, CognitionResearch.org, 2017. bit.ly/2ruLnrV, viXra:1707.0161v2, hal-01624595, v1.

[61] J. G. Wolff. Software engineering and the SP theory of intelligence. Technical report, CognitionResearch.org, 2017. Submitted for publication. bit.ly/2w99Wzq, arXiv:1708.06665.

[62] J. G. Wolff. Strengths and potential of the sp theory of intelligence in general, human-like artificial intelligence. Technical report, CognitionResearch.org, 2017. bit.ly/2z3rM8b, viXra:1711.0292, hal.archives-ouvertes.fr/hal-01633376.

[63] J. Woodward. Simplicity in the best systems account of laws of nature. *The British Journal for the Philosophy of Science*, 65(1):91–123, 2014.

[64] G. K. Zipf. *Human Behaviour and the Principle of Least Effort.* Hafner, New York, 1949. Republished by Martino Publishing, Mansfield Centre, CT, 2012.