

# Information compression, intelligence, computing, and mathematics

J Gerard Wolff\*

August 12, 2016

## Abstract

This paper presents evidence for the idea that much of artificial intelligence, human perception and cognition, mainstream computing, and mathematics, may be understood as compression of information via the matching and unification of patterns. This is the basis for the *SP theory of intelligence*, outlined in the paper and fully described elsewhere. Relevant evidence may be seen: in empirical support for the SP theory; in some advantages of information compression (IC) in terms of biology and engineering; in our use of shorthands and ordinary words in language; in how we merge successive views of any one thing; in visual recognition; in binocular vision; in visual adaptation; in how we learn lexical and grammatical structures in language; and in perceptual constancies. IC via the matching and unification of patterns may be seen in both computing and mathematics: in IC via equations; in the matching and unification of names; in the reduction or removal of redundancy from unary numbers; in the workings of Post's Canonical System and the transition function in the Universal Turing Machine; in the way computers retrieve information from memory; in systems like Prolog; and in the query-by-example technique for information retrieval. The chunking-with-codes technique for IC may be seen in the use of named functions to avoid repetition of computer code. The schema-plus-correction technique may be seen in functions with parameters and in the use of classes in object-oriented programming. And the run-length coding technique may be seen in multiplication, in division, and in several other devices in mathematics and computing. The SP theory resolves the apparent paradox of

---

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

“decompression by compression”. And computing and cognition as IC is compatible with the uses of redundancy in such things as backup copies to safeguard data and understanding speech in a noisy environment.

*Keywords:* information compression, intelligence, computing, mathematics

## 1 Introduction

“Fascinating idea! All that mental work I’ve done over the years, and what have I got to show for it? A goddamned zipfile! Well, why not, after all?” (John Winston Bush, 1996).

This paper describes a range of observations and arguments in support of the idea that much of artificial intelligence, human perception and cognition, mainstream computing, and mathematics, may be understood as compression of information via the matching and unification of patterns. These observations and arguments provide the foundation for the *SP theory of intelligence* and its realisation in the *SP computer model*—outlined below and described more fully elsewhere—in which information compression is centre stage. The aim here is to review, update, and extend the discussion in [32], itself the basis for [34, Chapter 2].

### 1.1 Related ideas

Related ideas have been around from at least as far back as the 14th century when William of Ockham suggested that if something can be explained by two or more rival theories, we should choose the simplest. Later, Isaac Newton wrote that “Nature is pleased with simplicity” [13, p. 320]; Ernst Mach [2] and Karl Pearson [14] suggested independently 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.”<sup>1</sup>; cosmologist John Barrow has written that “Science is, at root, just the search for compression in the world” [5, p. 247]; and George Kingsley Zipf [44] developed the idea that human behaviour is governed by a “principle of least effort”.

Partly inspired by the publication of Claude Shannon’s “theory of communication” [22] (now called “information theory”), Fred Attneave [1], Horace Barlow [3, 4] and others examined the role of information compression (IC) in the workings of brains and nervous systems.

---

<sup>1</sup>Quoted in [8, p. 512]

The close connection between information compression and several other inter-related topics has been demonstrated by several researchers including Ray Solomonoff (inductive inference [23] and ‘algorithmic probability theory’ [25]), Chris Wallace (classification [27] and inference [28]), Jorma Rissanen (modelling by shortest description [18] and ‘stochastic complexity’ [19]), Andrey Kolmogorov and Gregory Chaitin (‘algorithmic information theory’ (see, for example, [10]), and Satoshi Watanabe (pattern recognition [29]). And Ray Solomonoff [24] has argued that the great majority of problems in science and mathematics may be seen as either ‘machine inversion’ problems or ‘time limited optimization’ problems, and that both kinds of problem can be solved by inductive inference using the principle of ‘minimum length encoding’.

In later research, information compression has featured as a guiding principle for artificial neural networks (see, for example, [21], Section 4.4) and in research on grammatical inference (see, for example, [20]).

## 1.2 Novelty and contribution

The ideas described in this paper provide a perspective on artificial intelligence, human perception and cognition, mainstream computing, and mathematics, which is not widely recognised. The main features distinguishing it from other research are:

- The scope is very much broader than it is, for example, in the previously-mentioned research on artificial neural networks or grammatical inference. The thrust of the paper is evidence pointing to information compression via the matching and unification of patterns as an organising principle across diverse aspects of artificial intelligence, human perception and cognition, mainstream computing, and mathematics.
- Most research relating to information compression and its applications makes extensive use of mathematics. By contrast, information compression in this paper and in the SP theory focusses on the simple primitive idea, dubbed “ICMUP” and described in Section 2, that redundancy in information may be reduced by finding patterns that match each other and merging or unifying patterns that are the same. Far from using mathematics as a basis for understanding information compression, the paper argues, in Section 10, that ICMUP may provide a basis for mathematics.
- Although this is not the main focus of the paper, it is pertinent to mention that ICMUP provides the foundation for the distinctive and

powerful concept of *multiple alignment*, a central part of the SP theory of intelligence and, on evidence to date, a key to versatility and adaptability in intelligent systems.

I believe this perspective is important for the field of artificial intelligence for three main reasons:

- It has things to say directly about the nature of perception, learning, and other aspects of artificial intelligence.
- It provides a foundation for the SP theory of intelligence which, via the SP computer model, has demonstrable capabilities in several aspects of artificial intelligence, as outlined in Section 3, and it has a range of potential benefits and applications (*ibid.*).
- It suggests how artificial intelligence may be developed within an encompassing theoretical framework that includes human perception and cognition, mainstream computing, and mathematics.

### **1.3 Apparent contradictions and their resolution**

Given that large amounts of information can be produced by people, by computers, and via mathematics, and given that ‘redundancy’ or repetition in information is often useful in the storage and processing of information, it may seem perverse to suggest that IC is fundamental in our thinking, or in computing and mathematics. But for reasons outlined in Section 11, these apparent contradictions can be resolved.

### **1.4 Presentation**

As an introduction to what follows, the next section describes some basic principles of IC. After that, the SP theory is described in outline, with pointers to further sources of information, and a summary of empirical support for the theory. This last is itself evidence for the importance of IC in computing and cognition. The sections that follow describe several other strands of evidence that point in the same direction.

## **2 Preliminaries: information compression via the matching and unification of patterns**

To cut through much of the mathematical complexity associated with information compression, I have found it useful to focus on a rather simple idea:

that we may identify repetition or ‘redundancy’ in information by searching for patterns that match each other, and that we may reduce that redundancy and thus compress information by merging or ‘unifying’ two or more copies to make one. For the sake of brevity, this idea may be shortened to “information compression via the matching and unification of patterns” or “ICMUP”.

As just described, ICMUP loses information about the positions of all but one of the original patterns. But this can be remedied with any of the three variants of the idea, described below.

## 2.1 Concepts of number

ICMUP may seem too trivial to deserve comment. But because it is the foundation on which the rest of the SP system is built, there are implications that may seem strange and may at first sight look like major shortcomings in the theory:

- The first of these is that the SP system, in itself, has no concepts of number and has no procedures for processing numbers. Unlike an ordinary operating system or programming language, there is no provision for integers or reals and no functions such as addition, subtraction, square roots, or the like.

As we shall see in Section 3, the SP system does use a concept of frequency and it does calculate probabilities. But these are part of the workings of the system and not available to users. In any case, they may be modelled via analogue signals, without using conventional concepts of number or arithmetic.

- Secondly, because the system has no concepts of number, it minimises the use of compression techniques that depend on numbers, such as arithmetic coding or wavelet compression, although it does use Shannon-Fano-Elias coding [34, Section 3.5.2.1].

Although the core of the SP system lacks any concept of number, there is potential for the system to represent numbers and process them, provided that it is supplied with knowledge about Peano’s axioms and related information about the structure and functioning of numbers, as outlined in [34, Chapter 10]. The potential advantage of starting with a clean slate, focussing on the simple ‘primitive’ concept of ICMUP, is that it can help us avoid old tramlines, and open doors to new ways of thinking.

## 2.2 Variants of ICMUP

With the first variant of ICMUP—a technique called *chunking-with-codes*—the unified pattern is given a relatively short name, identifier, or ‘code’ which is used as a shorthand for the pattern or ‘chunk’. If, for example, the words “Treaty on the Functioning of the European Union” appear in several different places in a document, we may save space by writing the expression once, giving it a short name such as “TFEU”, and then using that name as a code or shorthand for the expression wherever it occurs. Likewise for the abbreviations in this paper, “IC” and “ICMUP”.

Another variant, *schema-plus-correction*, is like chunking-with-codes but the unified chunk of information may have variations or ‘corrections’ on different occasions. For example, a six-course menu in a restaurant may have the general form ‘Menu1: Appetiser (S) sorbet (M) (P) coffee-and-mints’, with choices at the points marked ‘S’ (starter), ‘M’ (main course), and ‘P’ (pudding). Then a particular meal may be encoded economically as something like ‘Menu1: (3) (5) (1)’, where the digits determine the choices of starter, main course, and pudding.

A third variant, *run-length coding*, may be used where there is a sequence two or more copies of a pattern, each one except the first following immediately after its predecessor. In this case, the multiple copies may be reduced to one, as before, with something to say how many copies there are, or when the sequence begins and ends, or, more vaguely, that the pattern is repeated. For example, a sports coach might specify exercises as something like “touch toes ( $\times 15$ ), push-ups ( $\times 10$ ), skipping ( $\times 30$ ), ...” or “Start running on the spot when I say ‘start’ and keep going until I say ‘stop’”.

## 3 Outline of the SP theory of intelligence

The *SP theory of intelligence*, described most fully in [34] and more briefly in [36], aims to simplify and integrate observations and concepts across artificial intelligence, human perception and cognition, mainstream computing, and mathematics, with ICMUP as a unifying theme. Distinctive features of the SP theory and its advantages compared with several AI-related alternatives are described in [42].

The theory, as it stands now is the product of an extended programme of development and testing via the SP computer model. It is envisaged that that model will be the basis for a high-parallel, open-source version of the *SP machine*, hosted on an existing high-performance computer, and accessible via the web. This will be a means for researchers everywhere to explore what

can be done with the system and to create new versions of it [36, Section 3.2], [43].

The SP theory, via the SP computer model, has demonstrable capabilities in areas that include the representation of diverse forms of knowledge (including class hierarchies, part-whole hierarchies, and their seamless integration), unsupervised learning, natural language processing, fuzzy pattern recognition and recognition at multiple levels of abstraction, best-match and semantic forms of information retrieval, several kinds of reasoning (one-step ‘deductive reasoning’, abductive reasoning, probabilistic networks and trees, reasoning with ‘rules’, nonmonotonic reasoning, explaining away, causal reasoning, reasoning that is not supported by evidence, and inference via inheritance of attributes), planning, problem solving, and information compression [34, 36]. It also has useful things to say about aspects of neuroscience and of human perception and cognition (*ibid.*).

Several potential benefits and applications of the SP theory are described in [41], with more detail in [37] (understanding natural vision and the development of artificial vision), [39] (how the SP theory may help to solve nine problems associated with big data), [38] (the development of computational and energy efficiency, of versatility, and of adaptability in autonomous robots), [35] (the SP system as an intelligent database), and [33] (application of the SP system to medical diagnosis). An introduction to the theory may be seen in [40].

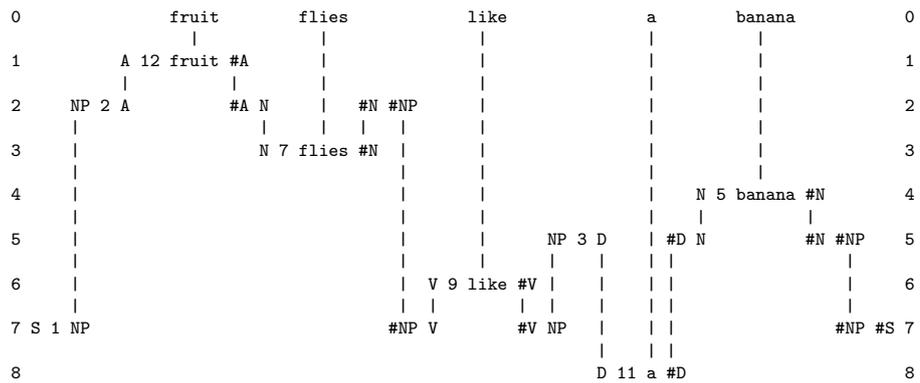
In broad terms, the SP theory has three main elements:

- All kinds of knowledge are represented with *patterns*: arrays of atomic symbols in one or two dimensions.
- At the heart of the system is compression of information via the matching and unification (merging) of patterns, and the building of *multiple alignments* like the two shown in Figure 1.<sup>2</sup> Here, the concept of multiple alignment has been borrowed and adapted from bioinformatics.
- The system learns by compressing *New* patterns to create *Old* patterns like those shown in rows 1 to 8 in each of the two multiple alignments in the figure.

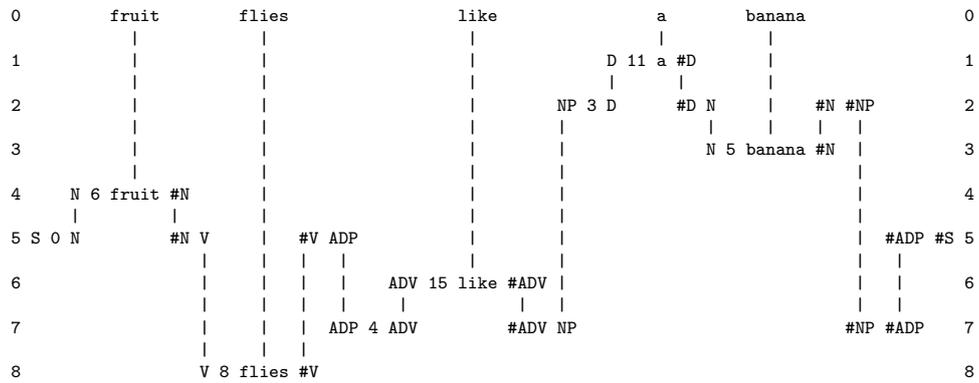
Because information compression is intimately related to concepts of prediction and probability [10], the SP system is fundamentally probabilistic. Each SP pattern has an associated frequency of occurrence, and probabilities may be calculated for multiple alignments and for inferences drawn from

---

<sup>2</sup>The example sentence is the second part of *Time flies like an arrow. Fruit flies like a banana.*, attributed to Groucho Marks.



(a)



(b)

Figure 1: Two multiple alignments created by the SP computer model showing two different parsings of the ambiguous sentence *Fruit flies like a banana*. Adapted from Figure 5.1 in [34], with permission.

multiple alignments [36, Section 4.4], [34, Section 3.7]. Although the system is fundamentally probabilistic, it may be constrained to deliver all-or-nothing results in the manner of conventional computing systems.

Since IC is central in the SP theory, the descriptive and explanatory range of the theory is itself evidence in support of the proposition that IC is a central principle in human perception and thinking, in computing and in mathematics.

## 4 Biology and engineering

This section and those that follow describe other evidence for the importance of IC in computing and cognition. First, let's take a bird's eye view of why IC might be important in people and other animals, and in computing.

In terms of biology:

- IC can confer a selective advantage to any creature by allowing it to store more information in a given storage space or use less space for a given amount of information, and by speeding up transmission of information along nerve fibres—thus speeding up reactions—or reducing the bandwidth needed for any given volume of information. In connection with the last point, Barlow [3, p. 548] draws attention to evidence that, in vertebrates at least, each optic nerve is far too small to carry reasonable amounts of the information emanating from the retina unless there is considerable compression of that information.
- Perhaps more important than the impact of IC on the storage or transmission of information is the close connection, already mentioned, between IC and inductive inference. Compression of information provides a means of predicting the future from the past and estimating probabilities so that, for example, an animal may get to know where food may be found or where there may be dangers.

Incidentally, the connection between IC and inductive prediction makes sense in terms of the matching and unification of patterns: any repeating pattern—such as the association between black clouds and rain—provides a basis for prediction—black clouds suggest that rain may be on the way—and probabilities may be derived from the number of repetitions.

- Being able to make predictions and estimate probabilities can mean large savings in the use of energy with consequent benefits in terms of survival.

As with living things, IC can be beneficial in computing—in terms of the storage and transmission of information and what is arguably the fundamental purpose of computers: to make predictions. It may also have a considerable impact in increasing the energy efficiency of computers [39, Section IX], [38, Section III].

As we shall see (Section 10), IC is more widespread in ordinary computers than may superficially appear.

## 5 Hiding in plain sight

Compression of information is so much embedded in our thinking, and seems so natural and obvious, that it is easily overlooked. Here are some examples.

### 5.1 Words as codes or shorthands

In the same way that “TFEU” is a convenient code or shorthand for “Treaty on the Functioning of the European Union”, a name like “New York” is a compact way of referring to the many things of which that renowned city is composed. Likewise for the many other names that we use: “Nelson Mandela”, “George Washington”, “Mount Everest”, and so on.

More generally, most words in our everyday language stand for *classes* of things and, as such, are powerful aids to economical description. Imagine how cumbersome things would be if, on each occasion that we wanted to refer to a “table”, we had to say something like “A horizontal platform, often made of wood, used as a support for things like food, normally with four legs but sometimes three, ...”, like the slow language of the Ents in Tolkien’s *The Lord of the Rings*. Likewise for verbs like “speak” or “dance”, adjectives like “artistic” or “exuberant”, and adverbs like “quickly” or “carefully”.<sup>3</sup>

### 5.2 Merging multiple views to make one

Here is another example. If, when we are looking at something, we close our eyes for a moment and open them again, what do we see? Normally, it is the same as what we saw before. But recognising that the before and after views are the same, means unifying the two patterns to make one and thus compressing the information, as shown schematically in Figure 2.

---

<sup>3</sup>Although natural language provides a very effective means of compressing information about the world, it is not free of redundancy. And that redundancy has a useful role to play in, for example, enabling us to understand speech in noisy conditions, and in learning the structure of language [36, Section 5.2].

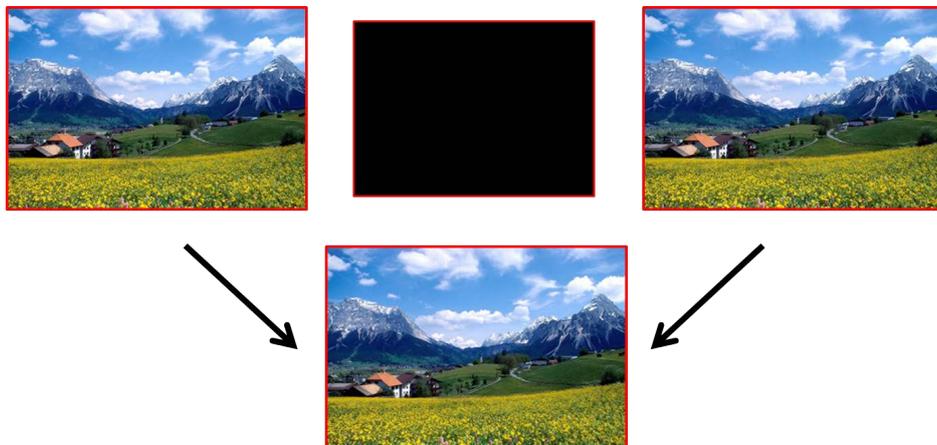


Figure 2: A schematic view of how, if we close our eyes for a moment and open them again, we normally merge the before and after views to make one. The landscape here and in Figure 3 is from Wallpapers Buzz ([www.wallpapersbuzz.com](http://www.wallpapersbuzz.com)), reproduced with permission.

It seems so simple and obvious that if we are looking at a landscape like the one in the figure, there is just one landscape even though we may look at it two, three, or more times. But if we did not unify successive views we would be like an old-style cine camera that simply records a sequence of frames, without any kind of analysis of understanding that, very often, successive frames are identical or nearly so.

### 5.3 Recognition

Of course, we can recognise something that we have seen before even if the interval between one view and the next is hours, months, or years. In cases like that, it is more obvious that we are relying on memory, as shown schematically in Figure 3. Notwithstanding the undoubted complexities and subtleties in how we recognise things, the process may be seen in broad terms as one of matching incoming information with stored knowledge, merging or unifying patterns that are the same, and thus compressing the information. If we did not compress information in that way, our brains would quickly become cluttered with millions of copies of things that we see around us—people, furniture, cups, trees, and so on—and likewise for sounds and other sensory inputs.

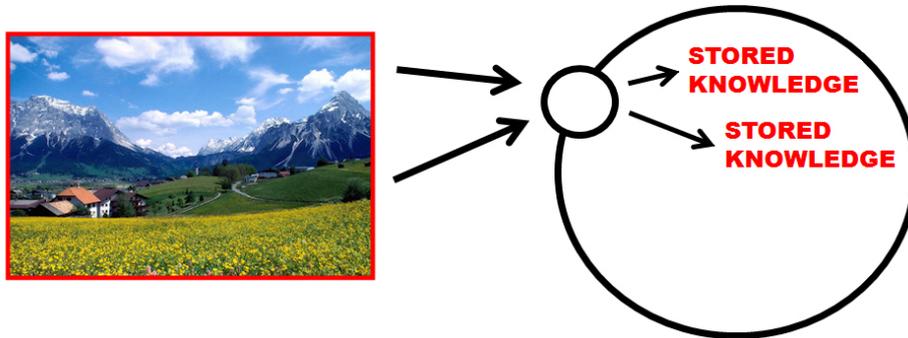


Figure 3: Schematic representation of how, in recognition, incoming visual information may be matched and unified with stored knowledge.

## 6 Binocular vision

IC may also be seen at work in binocular vision:

“In an animal in which the visual fields of the two eyes overlap extensively, as in the cat, monkey, and man, one obvious type of redundancy in the messages reaching the brain is the very nearly exact reduplication of one eye’s message by the other eye.” [4, p. 213].

In viewing a scene with two eyes, we normally see one view and not two. This suggests that there is a matching and unification of patterns, with a corresponding compression of information. Evidence in support of that conclusion comes from a demonstration with ‘random-dot stereograms’, as described in [37, Section 5.1].

In brief, each of the two images shown in Figure 4 is a random array of black and white pixels, with no discernable structure, but they are related to each other as shown in Figure 5: both images are the same except that a square area near the middle of the left image is further to the left in the right image.

When the images in Figure 4 are viewed with a stereoscope, projecting the left image to the left eye and the right image to the right eye, the central square appears gradually as a discrete object suspended above the background. Although this illustrates depth perception in stereoscopic vision—a subject of some interest in its own right—the main interest here is on how we see the central square as a discrete object. There is no such object in either of the two images individually. It exists purely in the *relationship* between the two images, and seeing it means matching one image with the other and unifying the parts which are the same.



Figure 4: A random-dot stereogram from [9, Figure 2.4-1], reproduced with permission of Alcatel-Lucent/Bell Labs.

1	0	1	0	1	0	0	1	0	1
1	0	0	1	0	1	0	1	0	0
0	0	1	1	0	1	1	0	1	0
0	1	0	Y	A	A	B	B	0	0
1	1	1	X	B	A	B	A	0	1
0	0	1	X	A	A	B	A	1	0
1	1	1	Y	B	B	A	B	0	1
1	0	0	1	1	0	1	1	0	1
1	1	0	0	1	1	0	1	1	1
0	1	0	0	0	1	1	1	1	0

1	0	1	0	1	0	0	1	0	1
1	0	0	1	0	1	0	1	0	0
0	0	1	1	0	1	1	0	1	0
0	1	0	A	A	B	B	X	0	0
1	1	1	B	A	B	A	Y	0	1
0	0	1	A	A	B	A	Y	1	0
1	1	1	B	B	A	B	X	0	1
1	0	0	1	1	0	1	1	0	1
1	1	0	0	1	1	0	1	1	1
0	1	0	0	0	1	1	1	1	0

Figure 5: Diagram to show the relationship between the left and right images in Figure 4. Reproduced from [9, Figure 2.4-3], with permission of Alcatel-Lucent/Bell Labs.

This example shows that, although the matching and unification of patterns is a usefully simple idea, there are interesting subtleties and complexities that arise when two patterns are similar but not identical.

Seeing the central object means finding a ‘good’ match between relevant pixels in the central area of the left and right images, and likewise for the background. Here, a good match is one that yields a relatively high level of IC. Since there is normally an astronomically large number of alternative ways in which combinations of pixels in one image may be aligned with combinations of pixels in the other image, it is not normally feasible to search through all the possibilities exhaustively.

As with many such problems in artificial intelligence, the best is the enemy of the good. Instead of looking for the perfect solution, we can do better by looking for solutions that are good enough for practical purposes. With this kind of problem, acceptably good solutions can often be found in a reasonable time with heuristic search: doing the search in stages and, at each stage, concentrating the search in the most promising areas and cutting out the rest, perhaps with backtracking or something equivalent to improve the robustness of the search. One such method for the analysis of random-dot stereograms has been described by Marr and Poggio [11].

It seems likely that the kinds of processes that enable us to see a hidden object in a random-dot stereogram also apply to how we see discrete objects in the world. The contrast between the relatively stable configuration of features in an object such as a car, compared with the variety of its surroundings as it travels around, seems to be an important part of what leads us to conceptualise the object as an object [37, Section 5.2]. Any creature that depends on camouflage for protection—by blending with its background—must normally stay still. As soon as it moves relative to its surroundings, it is likely to stand out as a discrete object.

The idea that IC may provide a means of discovering ‘natural’ structures in the world has been dubbed the ‘DONSVIC’ principle: *the discovery of natural structures via information compression* [36, Section 5.2].

## 7 Adaptation and run-length coding

IC may also be seen down in the works of vision. Figure 6 shows a recording from a single sensory cell (*ommatidium*) in the eye of a horseshoe crab (*Limulus polyphemus*), first when the background illumination is low, then when a light is switched on and kept on for a while, and later switched off—shown by the step function at the bottom of the figure.

As one might expect, the ommatidium fires at a relatively low rate of

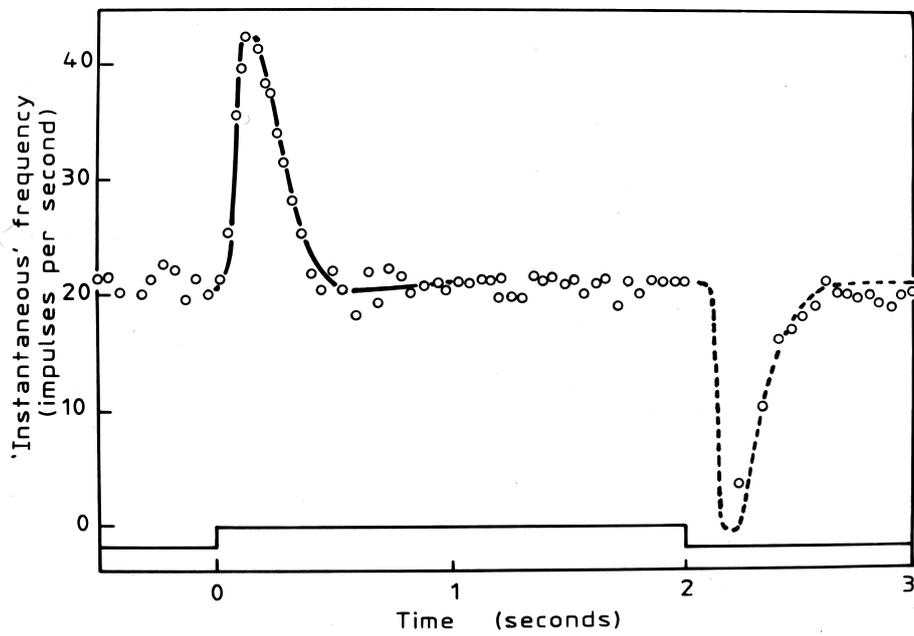


Figure 6: Variation in the rate of firing of a single ommatidium of the eye of a horseshoe crab in response to changing levels of illumination. Reproduced from [17, p. 118.], with permission from the Optical Society of America.

about 20 impulses per second even when the illumination is relatively low (shown at the left of the figure). When the light is switched on, the rate of firing increases sharply but instead of staying high while the light is on (as one might expect), it drops back almost immediately to the background rate. The rate of firing remains at that level until the light is switched off, at which point it drops sharply and then returns to the background level, a mirror image of what happened when the light was switched on.

This pattern of responding—adaptation to constant stimulation—can be explained via the action of inhibitory nerve fibres that bring the rate of firing back to the background rate when there is little or no variation in the sensory input [26]. But for the present discussion, the point of interest is that the positive spike when the light is switched on, and the negative spike when the light is switched off, have the effect of marking boundaries, first between dark and light, and later between light and dark.

In effect, this is a form of run-length coding. At the first boundary, the positive spike marks the fact of the light coming on. As long as the light stays on, there is no need for that information to be constantly repeated, so there is no need for the rate of firing to remain at a high level. Likewise, when the light is switched off, the negative spike marks the transition to darkness and, as before, there is no need for constant repetition of information about the new low level of illumination.

It is recognised that this kind of adaptation in eyes is a likely reason for small eye movements when we are looking at something, including sudden small shifts in position (‘microsaccades’), drift in the direction of gaze, and tremor [12]. Without those movements, there would be an unvarying image on the retina so that, via adaptation, what we are looking at would soon disappear.

Adaptation is also evident at the level of conscious awareness. If, for example, a fan starts working nearby, we may notice the hum at first but then adapt to the sound and cease to be aware of it. But when the fan stops, we are likely to notice the new quietness at first but adapt again and stop noticing it. Another example is the contrast between how we become aware if something or someone touches us but we are mostly unaware of how our clothes touch us in many places all day long. We are sensitive to something new and different and we are relatively insensitive to things that are repeated.

## 8 Other kinds of learning

As can be seen in Figure 7, people normally speak in ‘ribbons’ of sound, without gaps between words or other consistent markers of the boundaries

between words. In the figure—the waveform for a recording of the spoken phrase “on our website”—it is not obvious where the word “on” ends and the word “our” begins, and likewise for the words “our” and “website”. Just to confuse matters, there are three places within the word “website” that look as if they might be word boundaries.

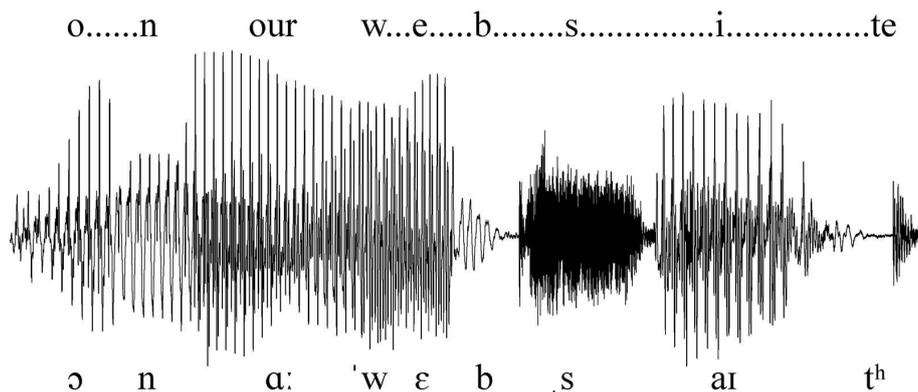


Figure 7: Waveform for the spoken phrase “On our website” with an alphabetic transcription above the waveform and a phonetic transcription below it. With thanks to Sidney Wood of SWPhonetics ([swphonetics.com](http://swphonetics.com)) for the figure and for permission to reproduce it.

Given that words are not clearly marked in the speech that young children hear, how do they get to know that language is composed of words? As before, it seems that IC and, more specifically, the DONSVIC principle, provide an answer. It has been shown that, via the matching and unification of patterns, the beginnings and ends of words can be discovered in an English-language text from which all spaces and punctuation has been removed, and this without the aid of any kind of dictionary or other information about the structure of English [36, Section 5.2]. It true that there are added complications with speech but it seems likely that similar principles apply.

The DONSVIC principle may also be applied to the process of learning the grammar of a language [31]. In addition to the learning of words, the process of grammar discovery or induction includes processes for learning grammatical classes of words (such as nouns, verbs and adjectives) and also syntactic forms such as phrases, clauses and sentences. Ultimately, grammar discovery should also include the learning of meanings and the association of meanings with syntax.

In connection with language learning, IC provides an elegant solution to

two problems: *generalisation*—how we generalise our knowledge of language without over-generalising; and *dirty data*—how we can learn a language despite errors in the examples we hear; with evidence that both these things can be achieved without the correction of errors by parents or teachers. In brief, a grammar that is good in terms of information compression is one that generalises without over-generalising; and such a grammar is also one that weeds out errors in the data. These things are described more fully in [36, Section 5.3].

## 9 Perceptual constancies

It has long been recognised that our perceptions are governed by *constancies*:

- *Size constancy.* To a large extent, we judge the size of an object to be constant despite wide variations in the size of its image on the retina [7, pp. 40-41].
- *Lightness constancy.* We judge the lightness of an object to be constant despite wide variations in the intensity of its illumination [7, p. 376].
- *Colour constancy.* We judge the colour of an object to be constant despite wide variations in the colour of its illumination [7, p. 402].

These kinds of constancy, and others such as shape constancy and location constancy, may each be seen as a means of encoding information economically. It is simpler to remember that a particular person is “about my height” than many different judgements of size, depending on how far away that person is. In a similar way, it is simpler to remember that a particular object is “black” or “red” than all the complexity of how its lightness or its colour changes in different lighting conditions.

## 10 Computing and mathematics

If, as seems to be the case, IC is fundamental in our thinking, then it should not be surprising to find that IC is also fundamental in things that we use to aid our thinking: computing in the modern sense where the work is done by machines, and mathematics, done by people or machines. Similar things can be said about logic but the main focus here will be on computing and mathematics, starting with the latter.

## 10.1 Mathematics and information compression

Roger Penrose [15] 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*.” (pp. 225–226, emphasis as in the original).

In a similar vein, John Barrow [5] 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. Mathematics works: as a result we have been tempted to equate understanding of the world with its mathematical encapsulation. ... Why is the world found to be so unerringly mathematical?” (Preface, p. vii).

These writings about the “mysterious” nature of mathematics, others such as Wigner’s [30] “The unreasonable effectiveness of mathematics in the natural sciences”, and schools of thought in the philosophy of mathematics—foundationism, logicism, intuitionism, formalism, Platonism, neo-Fregeanism, and more—have apparently overlooked an obvious point: *mathematics can be a very effective means of compressing information*. This apparent oversight is surprising since mathematics is indeed a useful tool in science and, as already mentioned, it is recognised that “Science is, at root, just the search for compression in the world.” [5, p. 247].

Here is an example of how ordinary mathematics—not some specialist algorithm for IC—can yield high levels of IC. Newton’s equation for his second law of motion,  $s = (gt^2)/2$ , is a very compact means of representing any realistically-large table of the distance travelled by a falling object ( $s$ ) in a given time since it started to fall ( $t$ ),<sup>4</sup> as illustrated in Table 1. That small equation would represent the table even if it was a 1000 times bigger, or more. Likewise for other equations such as  $E = mc^2$ ,  $a^2 + b^2 = c^2$ ,  $P = k/V$ , and so on.

In the subsections that follow, we shall dig a little deeper, looking at both mathematics and computing in terms of the ideas outlined earlier (Section 2): IC via the matching and unification of patterns, chunking-with-codes, schema-plus-correction, and run-length coding.

---

<sup>4</sup>Of course, the law does not work for something like a feather falling in air. The constant,  $g$ , is the acceleration due to gravity, about  $9.8m/s^2$ .

<i>Distance (m)</i>	<i>Time (sec)</i>
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
<i>Etc</i>	<i>Etc</i>

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

## 10.2 Information compression via the matching and unification of patterns

In mathematics, the matching and unification of patterns can be seen mainly in the matching and unification of names. 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  $x$  in the third equation with  $x$  in the first equation, and to unify the two so that the correct value is used for the calculation of  $z$ . Likewise for  $y$ .

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  $1 = 1$ ,  $2 = 11$ ,  $3 = 111$ , and so on. A numbering system like that is good enough for counting a few things but it is quite unmanageably cumbersome for large numbers. To be practical with numbers of all sizes, the obvious redundancies—in the repetitions of  $S$  and of 1—need to be reduced or eliminated. This can be done via the use of higher bases for numbers—binary, octal, decimal and the like [34, Section 10.3.2.2].

Emil Post's [16] "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. Much the same is true of the 'transition function' in a Universal Turing Machine.

The matching and unification of patterns may be seen in the way computers retrieve information from computer memory. This means finding a match between the address in the CPU and the address in memory, with implicit unification of the two. It is true that logic gates provide the mechanism for finding an address in computer memory but the process may also be seen as one of searching for a match between the address held in the CPU and the corresponding address in computer memory.

A system like Prolog—a computer-based version of logic—may be seen to function largely via the matching and unification of patterns. Much the same can be said about query-by-example, a popular technique for retrieving information from databases. Other examples will be seen in the subsections that follow.

## 10.3 Chunking-with-codes

If a set of statements is repeated in two or more parts of a computer program then it is natural to declare them once as a 'function', 'procedure' or 'sub-routine' within the program and to replace each sequence with a 'call' to the function from each part of the program where the sequence occurred. This may be seen as an example of the chunking-with-codes technique for IC: the

function may be regarded as a chunk, with the name of the function as its code or identifier.

In many cases but not all, a name or identifier in computing or in mathematics may be seen to achieve compression of information by serving as a relatively short code for a relatively large chunk of information.

Sometimes, the identifier can be larger than what it identifies but, normally, this can be seen to make sense in terms of IC via schema-plus-correction, next.

## 10.4 Schema-plus-correction

The schema-plus-correction idea may be seen in two main areas: functions with parameters, and object-oriented programming.

### 10.4.1 Functions with parameters

Normally, a function in a computer program, or a mathematical function, has one or more parameters, eg, `SQRT(number)` (to calculate a square root), `BIN2DEC(number)` (to convert a binary number into its decimal equivalent), and `COMBIN(count_1, count_2)` (to calculate the number of combinations of `count_1` things, taken `count_2` at a time).

Any such function may be seen as an example of schema-plus-correction: the function itself may be seen as a chunk of information that may be needed in many different places; the name of the function serves as a relatively short code; and the parameters provide for variations or ‘corrections’ for any given instance.

Imagine how inconvenient it would be if we were not able to specify functions in this way. Every time we wanted to calculate a square root, we would have to write out the entire procedure, and likewise for `BIN2DEC()`, `COMBIN()` and the many other functions that people use.

Here we can see why IC may be served, even if an identifier is bigger than what it identifies. Something that is small in terms of numbers of characters, such as the number 9, may be assigned to the relatively large identifier, “`number`”, in `SQRT(number)`, but that imbalance does little to outweigh the relatively large savings that arise from being able to call the function on many different occasions without having to write it out on each occasion. In any case, the processes of compiling or interpreting a computer program will normally convert long, human-friendly identifiers into short ones that can be processed more efficiently by computers.

### 10.4.2 Object-oriented programming

Apart from functions with parameters, the schema-plus-correction idea is prominent in object-oriented programming. From Simula, through Smalltalk to C++ and beyond, object-oriented languages allow programmers to create software ‘objects’, each one modelled on a ‘class’ or hierarchy of classes. Each such class, which normally represents some real-world category like ‘person’, ‘vehicle’, or ‘item for delivery’, may be seen as a schema. Like a function, each class normally has one or more parameters which may be seen as a means of applying ‘corrections’ to the schema. For example, when a ‘person’ object is created from the ‘person’ class, his or her gender and job title may be specified via parameters.

Classes in object-oriented languages are powerful aids to IC. If, for example, we have defined a class for ‘vehicle’, perhaps including information about the care and maintenance of vehicles, procedures to be followed if there is breakdown outside the depot, and variables for things like engine size and registration number, we avoid the need to repeat that information for each individual vehicle. Attributes of high-level classes are ‘inherited’ by lower-level classes, saving the need to repeat the information in each lower-level class.

## 10.5 Run-length coding

Run-length coding appears in various forms in mathematics, normally combined with other things. Here are some examples:

- Multiplication (eg,  $3 \times 4$ ) is repeated addition.
- Division of a larger number by a smaller one (eg,  $12/3$ ) is repeated subtraction.
- The power notation (eg,  $10^9$ ) is repeated multiplication.
- 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 a variable on each iteration, so these notations may be seen as a combination of run-length coding and schema-plus-correction.

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

Of course, things like multiplication and division are also provided in programming languages. In addition, there is more direct support for run-length coding with iteration statements like *repeat ... until*, *while ...*, and *for ....* For example,

```
s = 0;
for (i = 1; i <= 100; i++) s += i;
```

specifies 100 repetitions of adding  $i$  to  $s$ , with the addition of 1 to  $i$  on each iteration, without the need to write out each of the 100 repetitions explicitly.

Most programming languages also provide for run-length coding in the form of recursive functions like this:

```
int factorial(int x)
{
    if (x == 1) return(1) ;
    return(x * factorial(x - 1)) ;
}.
```

Here, the repeated multiplication and subtraction of the factorial function is achieved economically by calling the function from within itself.

## 11 Resolving apparent conflicts

As noted in the Introduction, the idea that IC is fundamental in artificial intelligence, human perception and cognition, and in mainstream computing and mathematics seems to be contradicted by the productivity of the human brain and the ways in which computers and mathematics may be used to create information as well as to compress it; and it seems to be contradicted by the fact that redundancy in information is often useful in both the storage and processing of information. These apparent contradictions and how they may be resolved are discussed briefly here.

### 11.1 Decompression by compression

An example of how computers may be used to create information is how the “hello, world” message of C-language fame may be printed 1000 times, with a correspondingly high level of redundancy, by a call to ‘`hello_world(1000)`’, defined as:

```

void hello_world(int x)
{
    printf("hello, world\n");
    if (x > 1) hello_world(x - 1) ;
}.

```

Here, the instruction ‘`printf("hello, world\n");`’ prints a copy of “hello, world”. Then, when the variable ‘`x`’ has the value 1000, the next line ensures that the whole process is repeated another 999 times.

The way in which IC may achieve this kind of productivity may be seen via the workings of the SP computer model. When that model [34, Sections 3.9, 3.10, and 9.2] is used to parse a sentence into its constituent parts and sub-parts, as shown in parts (a) and (b) of Figure 1, the model creates a relatively small code as a compressed representation of the sentence [34, Section 3.5]. But exactly the same computer model, using exactly the same processes of IC via the matching and unification of patterns, may reverse the process, reconstructing the original sentence from the code [34, Section 3.8]. This is similar to the way that a suitably-constructed Prolog program may not only be run ‘forwards’ to create ‘results’ from ‘data’ but may also be run ‘backwards’ to create ‘data’ from ‘results’. A very rough analogy is the way that a car can be driven backwards as well as forwards but the engine is working in exactly the same way in both cases.

Reduced to its essentials, the way that the SP model can be run ‘backwards’ works like this. Using our earlier example, a relatively large pattern like “Treaty on the Functioning of the European Union” is first assigned a relatively short code like “TFEU” to create the pattern “TFEU Treaty on the Functioning of the European Union” which combines the short code with the thing it represents. Then a copy of the short code, “TFEU”, may be used to retrieve the original pattern via matching and unification with “TFEU” within the combined pattern. The remainder of the combined pattern, “Treaty on the Functioning of the European Union”, may be regarded as the ‘output’ of the retrieval process. As such, it is a decompressed version of the short code. And that decompression has been achieved via a process of IC by the matching and unification of two copies of the short code.

Superficially, using one mechanism to run the model ‘forwards’ and ‘backwards’ has the flavour of a perpetual motion machine: something that looks promising but conflicts with fundamental principles. The critical issue is the size of the short code. It needs to be at least slightly bigger than the theoretical minimum for the process to work as described [34, Section 3.8.1]. If there is some residual redundancy in the code, the SP model has something to work on. With that proviso, “decompression by compression” is not as

illogical as it may sound.

## 11.2 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:

- With any kind of database, it is normal practice to maintain one or more backup copies as a safeguard against catastrophic loss of the data.
- With information on the internet, it is common practice to maintain two or more ‘mirror’ copies in different places to minimise transmission times and to reduce the chance of overload at any one site.
- The redundancy in natural language can be a very useful aid to comprehension of speech in noisy conditions.

These kinds of uses of redundancy may seem to conflict with the idea that IC—which means reducing redundancy—is fundamental in computing and cognition [6, p. 19]. However, the two things may be independent, or the usefulness of redundancy may actually be understood in terms of the SP theory itself.

An example of how the two things may be independent is the above-mentioned use of backup copies of databases: “... it is entirely possible for a database to be designed to minimise internal redundancies and, at the same time, for redundancies to be used in backup copies or mirror copies of the database ... Paradoxical as it may sound, knowledge can be compressed and redundant at the same time.” [34, Section 2.3.7].

An example of how the usefulness of redundancy may be understood in terms of the SP theory is how, in the retrieval of information from a database or other body of knowledge, there needs to be some redundancy between the search pattern and each matching pattern in the knowledge base (Section 11.1). Again, redundancy provides the key to how, in applications such as parsing natural language or pattern recognition, the SP system may achieve good results despite errors of omission, commission or substitution and thus, in effect, suggest interpolations for errors of omission and corrections for errors of commission or substitution [36, Sections 8 and 9], [35], [34, Section 6.2].

## 12 Conclusion

This paper presents evidence for the idea that much of artificial intelligence and of human perception and thinking, and much of computing and mathematics, may be understood as compression of information via the matching and unification of patterns.

This is the foundation for the SP theory of intelligence, outlined in Section 3, with pointers to where further information may be found. The explanatory range of the theory—in perception, reasoning, planning, problem solving, and more—provides indirect support for the idea that IC is an important principle in computing and cognition.

Information compression can mean advantages for creatures: in efficient storage and transmission of information; in being able to make predictions about sources of food, where there may be dangers, and so on; and in corresponding savings in energy. Likewise for artificial systems.

Some aspects of IC and its benefits are so much embedded in our everyday thinking that they are easily overlooked. Most nouns, verbs and adjectives may be seen as short codes for relatively complex concepts, and we frequently create shorthands for relatively long expressions. If we blink or otherwise close our eyes for a moment, we normally merge the before and after views into a single percept. In recognising something after a longer period, we are, in effect, merging the new perception with something that we remember. If we are viewing something with two eyes, we normally merge the two retinal images into a single percept.

IC may be seen in the phenomenon of adaptation in visual perception, in how we learn the structure of words and grammar in language, and in perceptual constancies.

IC via the matching and unification of patterns may be seen in both computing and mathematics. An equation can be a powerful aid to IC. In the processing of computer programs or mathematical equations, IC may be seen in the matching and unification of names. It may also be seen: in the reduction or removal of redundancy from unary numbers to create numbers with bases of 2 or more; in the workings of Post's Canonical System and the transition function in the Universal Turing Machine; in the way computers retrieve information from memory; in systems like Prolog; and in the query-by-example technique for information retrieval.

The chunking-with-codes technique for IC may be seen in the use of named functions to avoid repetition of computer code. The schema-plus-correction technique may be seen in functions with parameters and in the use of classes in object-oriented programming. And the run-length coding technique may be seen in multiplication, in division, and in several other

devices in mathematics and computing.

The SP theory resolves the apparent paradox of “decompression by compression”. And computing and cognition as IC is compatible with the uses of redundancy in such things as backup copies to safeguard data and understanding speech in a noisy environment.

This perspective can be fruitful in research into artificial intelligence, and human perception and cognition including neuroscience [34, 38, 37, 35] and in mainstream computing and its applications [41, 39, 33]. It may also prove useful in mathematics and its applications.

## References

- [1] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954.
- [2] E. C. Banks. The philosophical roots of Ernst Mach’s economy of thought. *Synthese*, 139:25–53, 2004.
- [3] H. B. Barlow. Sensory mechanisms, the reduction of redundancy, and intelligence. In HMSO, editor, *The Mechanisation of Thought Processes*, pages 535–559. Her Majesty’s Stationery Office, London, 1959.
- [4] H. B. Barlow. Trigger features, adaptation and economy of impulses. In K. N. Leibovic, editor, *Information Processes in the Nervous System*, pages 209–230. Springer, New York, 1969.
- [5] J. D. Barrow. *Pi in the Sky*. Penguin Books, Harmondsworth, 1992.
- [6] G. Dodig-Crnkovic. Rethinking knowledge. modelling the world as unfolding through info-computation for an embodied situated cognitive agent. *Litteratur och Språk*, 9:5–27, 2013.
- [7] J. P. Frisby and J. V. Stone. *Seeing: The Computational Approach to Biological Vision*. The MIT Press, London, England, 2010.
- [8] W. Isaacson. *Einstein: his life and universe*. Pocket Books, London, Kindle edition, 2007.
- [9] B. Julesz. *Foundations of Cyclopean Perception*. Chicago University Press, Chicago, 1971.
- [10] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, New York, 2009.

- [11] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B*, 204(1156):301–328, 1979.
- [12] S. Martinez-Conde, J. Otero-Millan, and S. L. Macknik. The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, 14:83–96, 2013.
- [13] 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*.
- [14] 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](http://bit.ly/1g2gNfk).
- [15] R. Penrose. *The Emperor’s New Mind*. Oxford University Press, Oxford, 1989.
- [16] E. L. Post. Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics*, 65:197–268, 1943.
- [17] F. Ratliff, H. K. Hartline, and W. H. Miller. Spatial and temporal aspects of retinal inhibitory interaction. *Journal of the Optical Society of America*, 53:110–120, 1963.
- [18] J. Rissanen. Modelling by the shortest data description. *Automatica*, 14(5):465–471, 1978.
- [19] J. Rissanen. Stochastic complexity. *Journal of the Royal Statistical Society B*, 49(3):223–239, 1987.
- [20] H. Sakamoto. Grammar compression: grammatical inference by compression and its application to real data. In *Proceedings of the 12th international conference on grammatical inference*, volume 34 of *JMLR: Workshop and Conference Proceedings*, pages 3–20. 2014.
- [21] J. Schmidhuber. Deep learning in neural networks: an overview. *Neural Networks*, 61:85–117, 2015.
- [22] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.
- [23] R. J. Solomonoff. A formal theory of inductive inference. Parts I and II. *Information and Control*, 7:1–22 and 224–254, 1964.

- [24] R. J. Solomonoff. The application of algorithmic probability to problems in artificial intelligence. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 473–491. Elsevier Science, North-Holland, 1986.
- [25] R. J. Solomonoff. The discovery of algorithmic probability. *Journal of Computer and System Sciences*, 55(1):73–88, 1997.
- [26] G. von Békésy. *Sensory Inhibition*. Princeton University Press, Princeton, NJ, 1967.
- [27] C. S. Wallace and D. M. Boulton. An information measure for classification. *Computer Journal*, 11(2):185–195, 1968.
- [28] C. S. Wallace and P. R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statistical Society B*, 49(3):240–252, 1987.
- [29] S. Watanabe, editor. *Frontiers of Pattern Recognition*. Academic Press, New York, 1972.
- [30] E. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13:1–14, 1960.
- [31] J. G. Wolff. Learning syntax and meanings through optimization and distributional analysis. In Y. Levy, I. M. Schlesinger, and M. D. S. Braine, editors, *Categories and Processes in Language Acquisition*, pages 179–215. Lawrence Erlbaum, Hillsdale, NJ, 1988. [bit.ly/ZIGjyc](http://bit.ly/ZIGjyc).
- [32] J. G. Wolff. Computing, cognition and information compression. *AI Communications*, 6(2):107–127, 1993. See [bit.ly/XL359b](http://bit.ly/XL359b).
- [33] J. G. Wolff. Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, unification and search. *Decision Support Systems*, 42:608–625, 2006. [bit.ly/XE7pRG](http://bit.ly/XE7pRG), [arXiv:1409.8053](https://arxiv.org/abs/1409.8053) [cs.AI].
- [34] 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](http://bit.ly/WmB1rs).
- [35] 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/Yg2onp](https://bit.ly/Yg2onp), arXiv:cs/0311031 [cs.DB]. A revised version of this paper is to be published in [?] as “Towards a NoSQL database system with intelligence founded on the SP theory of intelligence”.
- [36] J. G. Wolff. The SP theory of intelligence: an overview. *Information*, 4(3):283–341, 2013. [bit.ly/1hz0lFE](https://bit.ly/1hz0lFE).
- [37] J. G. Wolff. Application of the SP theory of intelligence to the understanding of natural vision and the development of computer vision. *SpringerPlus*, 3(1):552–570, 2014. [bit.ly/1scmpV9](https://bit.ly/1scmpV9).
- [38] J. G. Wolff. Autonomous robots and the SP theory of intelligence. *IEEE Access*, 2:1629–1651, 2014. [bit.ly/1zrSemu](https://bit.ly/1zrSemu).
- [39] J. G. Wolff. Big data and the SP theory of intelligence. *IEEE Access*, 2:301–315, 2014. [bit.ly/1jGWXDH](https://bit.ly/1jGWXDH). This article, with minor revisions, is to be reproduced in Fei Hu (Ed.), *Big Data: Storage, Sharing, and Security (3S)*, Taylor & Francis LLC, CRC Press, 2016, pp. 143–170.
- [40] J. G. Wolff. The SP theory of intelligence and the sp machine, in brief. Technical report, CognitionResearch.org, 2014. Unpublished document. See [bit.ly/1FvAMxf](https://bit.ly/1FvAMxf).
- [41] J. G. Wolff. The SP theory of intelligence: benefits and applications. *Information*, 5(1):1–27, 2014. [bit.ly/1lcquWF](https://bit.ly/1lcquWF).
- [42] J. G. Wolff. The SP theory of intelligence: its distinctive features and advantages. *IEEE Access*, 4:216–246, 2016. [bit.ly/21gv2jT](https://bit.ly/21gv2jT).
- [43] J. G. Wolff and V. Palade. Proposal for the creation of a research facility for the development of the SP machine. Technical report, CognitionResearch.org, 2015. [bit.ly/1zZjjIs](https://bit.ly/1zZjjIs), arXiv:1508.04570 [cs.AI].
- [44] G. K. Zipf. *Human Behaviour and the Principle of Least Effort*. Hafner, New York, 1949. Republished by Martino Publishing, Mansfield Centre, CT, 2012.