INFORMATION COMPRESSION AND LOGIC

J Gerard Wolff

October 1995

School of Electronic Engineering and Computer Systems, University of Wales, Dean Street, Bangor, Gwynedd, LL57 1UT, UK. Tel: +44 1248 382691. E-mail: gerry@sees.bangor.ac.uk. Fax: +44 1248 361429. Web: http://www.sees.bangor.ac.uk/~gerry.

Abstract

The purpose of this article is to present and discuss some of the meta-theoretical considerations which provide some rationale for the idea that logic may usefully be understood as information compression (IC). This article may be seen as an extended introduction to the accompanying article [23] which presents some relatively concrete proposals in this area.

This article first reviews some concepts associated with information and the compression of information. Then it considers some aspects of human perception and cognition which may be understood as IC and which appear to be relevant to understanding the nature of logic. Next is considered some interconnections amongst concepts of randomness, redundancy, (inductive and deductive) inference and IC. Finally, some features of logic are considered which seem to point more directly to the idea that IC may provide a unifying theme for understanding logic.

1 Introduction

In every field of enquiry, there is a continuing need to integrate and rationalise concepts developed in the field or its sub-fields, to create and refine one or more theories aiming to maximise coherence and simplicity in the field as a whole.

In logic and mathematics this tradition is exemplified classically by work aiming to reduce mathematics to logic (Frege, Hilbert, Russell and Whitehead), to reduce logic to number theory (Skolem, Gödel) or to model logic within set theory (Tarski, Bourbaki school).

This article and the accompanying article [23] continue in this tradition, presenting the idea that information compression (IC) and associated concepts may provide a framework within which concepts in 'logic'[1] may usefully be understood. Given the close relationship between logic and mathematics, it appears that much of this thinking may also be applied to mathematics.

Although concepts of compression and simplification are recognised, for example, in connection with the need to avoid replication of steps in the proving of theorems [4, Section 4.5] and the need to simplify intermediate structures created as a proof proceeds [4, Section 4.4; 9, Section 4.1], the possibility that IC might have fundamental significance as a unifying concept in a general theory of logic seems to be not widely recognised.

The accompanying article presents some examples intended to illustrate the potential of IC as a unifying theme by showing in a relatively concrete way how some familiar concepts in logic may be understood in those terms.

The purpose of this article is to present and discuss some of the meta-theoretical considerations which provide some rationale for the proposals and some justification for pursuing what may otherwise seem an outlandish idea.

1.1 Background

This work is part of a programme of research [24-36] developing the 'SP' conjecture that:

> *All kinds of computing and formal reasoning may usefully be understood as information compression by pattern matching, unification*[2] *and search*

This research is motivated partly by the perceived need for rationalisation and simplification of concepts in computing and partly by the more practical consideration that this line of thinking may provide a foundation for developing a 'new generation' computing system with more flexibility

1. In these two articles, the term 'logic' is taken to cover the kinds of ideas discussed in [1, 8, 10].

2. Unless otherwise specified, the term *unification* in these two articles means a simple merging of matching patterns, close to the non-technical meaning of the word. Although there is some risk of confusion with the more complicated concept of 'unification' in logic, the term has been adopted in the SP programme because no other word captures the intended meaning so well.

and ‘intelligence’ than conventional systems.

Until recently, the main focus of the programme has been developing this conjecture in relation to aspects of computing such as unsupervised learning, best-match information retrieval, ‘fuzzy’ pattern recognition, and others. The possibility that ‘logic’ in some sense might fall within the scope of the conjecture has been kept clearly in view but only a few tentative steps had been taken in that area (see, for example, [35, Chapter 5]).

Recent insights [24-28] have now clarified a number of relevant issues and a much clearer path can now be seen towards the assimilation and integration of concepts from logic within the developing framework of ideas.

1.2 Why IC?

The term ‘information compression’ (or ‘data compression’) is commonly associated with computer programs designed to facilitate the transmission, processing or storage of information by reducing its size. But IC is significant in (at least) two other areas:

- The workings of brains and nervous systems.
- Inductive inference and the philosophy of science.

And of course, since the concept of IC includes the concept of ‘information’, there is a close relation between IC and theories of information.

Part of the motivation for exploring the idea that IC may provide a unifying theme in logic is that this theme suggests the additional possibility of integrating concepts from logic within a broader theoretical framework which includes theories of information, inductive inference, philosophy of science, and brain science. Some of these considerations are discussed in Sections 3, 4 and 5.

2 Information and Compression of Information

As an introduction to what follows, this section presents, in outline, some basic ideas related to concepts of information and IC. In this section and later, the implicit subject of much discussion is some given body of information which, for the sake of brevity, will be referred to as ‘*I*’.

2.1 Shannon’s Information Theory

In Shannon’s Information Theory [17], information is viewed as a pattern of discriminable differences distributed in one or more dimensions. One form of information which accords with this view is any kind of continuously-varying wave. But the advent of digital communications, CD-ROMS and so on has now made familiar the idea that any wave can be translated with arbitrary precision into a pattern of atomic symbols drawn with replacement from a finite ‘alphabet’ of symbol types (often ‘0’ and ‘1’), and *vice versa*. In this discussion, it will be convenient to consider cases where *I* is a one-dimensional sequence of symbols. Unless otherwise stated, the alphabet will be assumed to be ‘0’ and ‘1’.

If, at a given location in *I*, the symbol types in the alphabet are equi-probable, then the communicative value of the symbol which actually appears in that location is at a theoretical maximum. If symbol types are equi-probable at all locations in *I* then *I* is considered to be *random*. In terms of the theory, its communicative value is maximal.

If symbol types are not equi-probable in any part of *I* then *I* is not random and its communicative value is less than the theoretical maximum. The difference between the actual communicative value of *I* and the theoretical maximum is the *redundancy* in *I*. In an extreme case, all the symbols are the same (e.g., ‘000000000 ...’ or ‘1111111111 ...’) and, apart from the first symbol and the length of the sequence, *I* is wholely redundant.

2.1.1 Absolute and Contextual Probabilities

In the simplest analysis, the probability of each symbol type at any location in *I* can be estimated by counting the number of occurrences of each of the symbol types in *I*.

But it has been recognised from the beginning of work in this area that there are often contextual constraints amongst symbols in *I* which mean that the simplest analysis is likely to give an inaccurate measure of the relevant probabilities. In English, for example, it is well known that the probability of the letter ‘u’ to the right of the letter ‘q’ is almost 1. In that context, the letter ‘u’ is almost completely redundant.

Assessing contextual probabilities of symbols is more difficult than estimating absolute probabilities mainly because of uncertainty about what the ‘context’ of any symbol might be. However, as discussed elsewhere [29, 34], it is possible to get a handle on this problem by generalising the concept of ‘symbol’ to include sequences of one *or more* basic symbols and, in effect, assimilating the notion of ‘context’ within this generalised concept of ‘symbol’. To avoid confusion, the generalised concept will be referred to as a *pattern*. The term *symbol* will be reserved for basic or ‘atomic’ symbols as normally understood.

Given this generalisation, contextual probabilities for symbols give way to absolute probabilities of patterns. As with symbols, these probabilities may be estimated by counting the numbers of occurrences of patterns in *I*. A point which deserves emphasis in this connection is that, while a pattern is simply a sequence of symbols, the *instances* of any pattern as they appear in *I* include the many cases where the symbols in the given pattern are not contiguous in *I*.[3] Naturally, if one is counting the number of occurrences of any pattern, one would consider only *disjoint* instances which do not have any symbol instances in common.

The frequency of occurrence of any pattern within a random body of information depends on its size: small patterns repeat more frequently than large ones. When frequencies and probabilities are not balanced: *a pattern which repeats more frequently in I than other patterns of the same size*

3. In this article and the accompanying one [23], the term *substring* means a sequence of symbols within a larger sequence where the symbols are contiguous within the larger sequence. The term *subsequence* means a sequence of symbols within a larger sequence where the symbols are *not* necessarily contiguous within the larger sequence. The term *pattern* means any sequence, string, substring or subsequence of symbols.

represents redundancy in I. Equivalently, *a repeating pattern in I which is bigger than other patterns of the same frequency represents redundancy in I.*

A fuller discussion of these points may be found in [29].

2.2 IC by Pattern Matching, Unification and Search

IC is largely a matter of extracting redundancy from *I* although it can sometimes also be convenient to sacrifice some of the non-redundant information in *I* (to achieve more compaction or to speed up processing).

As indicated in Section 2.1, there is an intimate relation between Shannon's concept of redundancy and concepts of probability estimated from frequencies of occurrence and sizes of patterns of symbols. But 'frequency of occurrence' implies repetition of information: redundancy in *I* is information which repeats what is already present in *I* and is thus unnecessary in terms of its communicative value (although redundancy has well-known uses in error correction and in increasing the reliability of systems).

The idea that redundancy in *I* equates with the repetition of patterns in *I* provides a key to understanding how redundancy may be detected in *I* and then removed:

- To detect repetition of patterns it is necessary to make same-different comparisons between patterns - to match patterns against each other.

- Where a pattern repeats in *I* more often than other patterns of the same size, the corresponding redundancy may be reduced by merging or 'unifying' all instances of the pattern to make a single 'chunk'. To preserve information about the locations of instances of the chunk that have been deleted, it may be necessary to mark each such location with a relatively short 'code' or 'tag' pattern. An example from this article is the use of 'IC' as shorthand for 'information compression'.

- In any realistically large body of information, there is an astronomically large number of alternative ways in which patterns may be matched and, typically, many alternative ways in which patterns may be unified. Given the need to maximise compression for a given computational cost, some kind of *search* is required amongst the many alternatives. Given that exhaustive search of the entire abstract space is normally impossible, some kind of 'metrics-guided' search is normally required (e.g., 'hill climbing', 'beam search', 'genetic algorithm', 'simulated annealing'), or the search space must be constrained in some other way.

Given heavily constrained versions of the search space, these ideas form the basis of most of the simpler 'standard' techniques for information compression (see [19]). A working hypothesis in the SP programme is that, at some stage, *all* techniques for information compression may be seen to rest on these basic principles.

In this article and its companion, the terms 'pattern matching, unification and search' are abbreviated as 'PMUS'.

2.3 Algorithmic Information Theory

In Algorithmic Information Theory (AIT) (see, for example, [6, 14]), a key idea is that *I* is *random* if no computer program can be discovered or devised which can generate *I* and which is smaller than *I*. For most realistically large bodies of information, it is not possible to prove that no such computer program can be found. For any (reasonably large) *I* which is apparently random, there is always the possibility that, in the future, some computer program may be found which is smaller than *I* but which will generate it.

This view of the concept of randomness may, at first sight, appear to be radically different from the concept of randomness in Shannon's information theory. But, given the idea that redundancy in the Shannon view of information may be interpreted as 'relatively frequent repetition of patterns' and given the idea that unification of repeated patterns is the key to compression, then the two views may be seen to complement each other quite well: where *I* contains redundancy in the form of repeated patterns, it is compressible; and, where no such patterns can be found, *I* may be regarded as random.

Readers may notice an apparent contradiction between the idea of 'computing as compression' and the obvious fact (which forms part of the AIT concept of randomness, as just described) that computers may be used to decompress information and create redundancy. The way in which this apparent contradiction may be resolved is discussed in [26, 29, 33].

Within AIT, there is no suggestion that 'computing' may itself be understood as IC. This and other important differences between AIT and the SP theory are described in [29, 33].

3 Information Processing in Brains and Nervous Systems

Formal logic, especially when it is embedded in a computer system, seems to have an existence and a validity which is independent of how we think. Few people, these days, would see logic as part of human psychology but early workers in logic had little doubt that they were studying human thinking, witness the title of George Boole's book: *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. And there is no dispute that, for most of its history, logic has been exclusively a form of human thinking and that informal kinds of logic are a prominent part of everyday thinking.

Whilst it would be absurd to try to turn the clock back and make logic a mere branch of psychology, it seems that developments in psychology since that subject split away from logic may shed some light on the latter. If 'AI' is regarded as part of psychology, this is in keeping with Gabbay's suggestion that: "we should try and export from AI into logic, let logic evolve, and then reuse the new logic in AI." [10, p. vi]. More generally, IC may provide a unifying theme for a 'cognitive science' which embraces both cognitive psychology and logic.

3.1 IC and the Workings of Brains and Nervous Systems

Since at least as long ago as Ernst Mach's writings in the last century, it has been recognised that IC may be an important feature of the workings of brains and nervous systems. More specifically, it has been recognised that inhibitory processes in nervous tissue have a key role in compressing information by the extraction of redundancy from information (see, for example, [3, 15, 21]).

In any biological system for information processing which is exposed to the rigours of natural selection, IC makes good sense from the perspective of basic engineering because it should facilitate the transmission or storage of information. IC has the additional, and perhaps more significant advantage, that it is closely linked to inductive inference (as discussed in Sections 4.3, 4.4 and 4.5, below) which means an ability to anticipate or predict future events from past observations - and that means an ability to evade danger and find what is needed for living.

It is not appropriate here to review writings in this area in detail since reviews and summaries of the literature are published elsewhere [29, 34]. The purpose here is to highlight some aspects of this thinking which appear to be relevant to logic. Some of the ideas introduced here will be further developed in Sections 4 and 5.

3.2 Unsupervised Learning

Of the several meanings of the word ‘learning’, the one intended here is the kind of unconscious and automatic assimilation of knowledge without explicit teaching which a young child appears to do in learning his or her first language.

There is good evidence that this kind of ‘unsupervised’ learning may, to a large extent, be understood as IC: reducing large samples of language heard to compact descriptions in terms of ‘grammars’. There is empirical support for this view: computer models of language learning which perform IC by PMUS show remarkable correspondences with observed features of language development in children [37].

The connection between learning in this sense and logic may seem remote but unsupervised learning is closely related to inductive inference which, as suggested in Section 4, below, appears to be more closely related to deductive inference than has traditionally been thought. In another paper [28], I have made some tentative proposals about possible ways in which learning and logic might be integrated.

3.3 The Concept of an Object

The everyday concept of an ‘object’ - person, tree, car, house, and so on - is not part of the machinery of formal logic but logic is often applied to ‘real world’ objects, and symbolic entities such as the elements of a logical set are often equated with such objects.

It is often assumed that an object is some kind of conceptual primitive which is somehow immediately ‘given’ by the environment in which we live. But visual recognition of an object as an object is almost certainly the product of complex processing by our brains. The visual information which we receive from the environment is like a series of frames of cinema film each one of which is merely a pattern of light, shade and colour with no differentiation of the constituent objects from each other or from their backgrounds. Achieving that differentiation is a task which our brains perform so effortlessly and unconsciously that normally we perceive objects as simply ‘there’.

It appears that the process by which the brain distinguishes objects as objects is largely IC by PMUS. Some of the thinking behind this proposition is described in [34] and will not be repeated here. In brief, the argument is that matching and unification of patterns coupled with heuristic search through the many alternatives can differentiate the relatively constant pattern represented by an object from the relatively changeable patterns of the object's environment.

3.4 Classes and Inheritance

Psychologists, logicians and others concerned with the representation of knowledge in natural or artificial systems recognise that objects and other entities may form goupings which are variously called classes, sets, concepts, categories, and so on, depending on the context. It is generally recognised that any such grouping may be described extensionally by listing the members of the group or it may be described intensionally in terms of the attributes of each member of the group. In the latter case, it is recognised that an element of a group or a whole group may share or *inherit* the attributes of one or more groups within which they may be included, and this recursively through a hierarchy or heterarchy of groups and sub-groups. In addition to these characteristics, psychologists recognise that the kind of 'natural' categories which people seem to use in everyday thinking often have the property of *polythesis*: that no single attribute need be present in every member of the group.

In this connection, inheritance is a powerful mechanism for minimising redundancy and thus compressing information. For the kinds of concepts we ordinarily use, the impact of inheritance is enormous. It is, for example, hard to imagine that anyone would store in their brains a full description of a human being for every one of the people that they know. It is so much more economical to store the description of a human being once and then retrieve that description for each individual person as and when required.

This kind of thinking has led psychologists to believe, not unreasonably, that inheritance is part of our cognitive machinery. Attempts to demonstrate this experimentally (e.g., [7]) have not been entirely successful but this may say more about the limitations of the experimental method in cognitive psychology than about the role of inheritance in cognition. In general, it seems hard to resist the conclusion that inheritance and the associated mechanisms for IC are a key part of how we store and use information in our brains.

Inheritance is clearly a form of inference and, as suggested in [23], it seems to be closely related to *modus ponens* reasoning. The possible roles of inheritance in logic are considered in, for example, [12].

3.5 Pattern Recognition

One of the most prominent features of human (and animal) perception and cognition is the ability to recognise objects and, more generally, patterns despite distortions of various kinds: omission, addition or substitution of elements relative to the pattern that is recognised.

Establishing in broad terms what people can do in this area does not require any clever experimental method or sophisticated analysis. One needs only to look at a typical scene to understand that one object frequently occludes another and that we recognise many of the things in the scene each from only a subset of its parts. Consider, for example, how easily we may recognise a building as a building despite the intervening trunk and branches of a tree which may obscure quite large parts of the building.

That pattern recognition may be interpreted as IC has been recognised for some time [22]. It should in any case be clear that, since pattern recognition is largely a process of unifying a pattern in the incoming information with a pattern stored in memory, it can be seen as IC by PMUS (cf. Section 2.2).

As described in Section 4.5, this capability which we have for recognising patterns from subsets of their parts seems to provide a basis for both inductive and deductive reasoning of a 'naturalistic' kind. This thinking is further developed in [23, Sections 5 and 6].

4 Randomness, Redundancy, Inference and IC

This section traces some connections which suggest that inductive and probabilistic reasoning are more closely related to deductive reasoning than has traditionally been thought and that IC by PMUS may provide a unifying theme. In this area of thinking there seems to be a web of interconnections which suggests the possibility of integration into a coherent whole.

4.1 Deduction Rests on the Inductive Principle?

Concepts of randomness and redundancy in Shannon's information theory and AIT provide a useful perspective on logic and, indeed, on information processing of any kind.

Consider, for example, a world which is *totally* random. Such a world would consist entirely of variation, something like a combination of white light with 'white' hissing sound. There would be no persistent or repeating patterns, no regularities and no structure of any kind. That means that there would be none of the persistent structured entities that we are familiar with in everyday life and that includes those persistent structured entities - brains and computers - which we use to process information, including logic.

In a world that was totally random, the absence of brains and computers would be of little consequence because there would be nothing to be achieved by the use of such mechanisms for the storage or processing of information. Data stored at one moment would probably be invalid in the next nano-second. An algorithm which gives one result today would probably give a different result tomorrow. A logical deduction made on one occasion is unlikely to be valid on another occasion.

In short, concepts of calculation and logical deduction are predicated on the belief that we live in a world which is *not* random, in which certain entities and regularities are persistent from time to time and from place to place.

If this is accepted, then another conclusion, which is perhaps surprising, seems to follow: that logic and mathematics are founded on the inductive principle that the past is a guide to the future. Since inductive inferences can never be proved, we are led to the conclusion that the whole edifice of proof and calculation rests on an unprovable hypothesis or conjecture!

If logic has uncertain foundations, any inference made within logic can never be completely certain.

4.2 Uncertainty in Arithmetic

In relation to the inter-linked themes of compression and uncertainty in inference, it is pertinent to mention Chaitin's work in the framework of AIT: "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!" [5].

Number theory is not the same as logic but logic and mathematics are sufficiently close to suggest that results in one area may generalise to the other.

4.3 Inductive Inference and IC

The connection between inductive inference and IC has been recognised from at least as long ago as Solomonoff's classic papers [18] which considered the problem of estimating the probabilities of the strings of unobserved symbols which might follow the end of a long string of observed symbols.

In broad terms, the solutions which Solomonoff proposed amounted to finding an 'encoding' of the observed symbols which was as small as possible, where the size of the encoding was to include both the encoded sequence of symbols and the definition of the code. The required probabilities could then be derived from these encodings.

This and later research on 'Minimum Length Encoding', grammatical inference and language learning (see [29] for more references) is clearly related to the principle of simplicity in scientific explanation which is itself a version of Occam's Razor: that 'entities should not be multiplied without necessity'.

If it is accepted that the primary goal of science is to find compressed descriptions of natural phenomena then an implication is that there is no absolute truth or absolute proof in scientific explanation. This is because it is never possible to know whether the most economical description has been found (cf. Section 2.3). This was partially recognised by Popper [16] who suggested that scientific theories could never be proved, only disproved, and more fully recognised by Lakatos [13] who suggested that scientific theories could be neither proved nor disproved, only weighed as more or less probable.

In general, there is wide consensus that inductive inference is a probabilistic matter which is closely related to IC.

4.4 PMUS and the Discovery of Inductive Regularities

The inductive principle that the past is a guide to the future is, more precisely stated, that patterns which have occurred relatively frequently in the past are relatively likely to occur in the future: we expect the sun to rise each morning because it always has in the past; we discount the possibility that the next paving stone in a path will drop us into a hole because this has never happened to us in the past.

But, following the arguments in Section 2.2, relatively frequent occurrence of patterns in the past means redundancy. And we can only know that a pattern has occurred relatively frequently in the past by the matching and unification of the several instances of that pattern. Thus inductive reasoning depends on IC by PMUS.

4.5 Inductive and Deductive Reasoning as Partial Pattern Recognition

In order to discover an inductive regularity like the ones just considered, we need to observe several complete instances of the relevant pattern. But when we use such a pattern to make an inductive inference, our inference is based on observing only part of the pattern and predicting the rest. Inductive inference may be seen as a process of recognising a whole pattern from a subset of its parts. As discussed in Section 3.5, this kind of capability for partial pattern recognition is very well developed in people and in animals too.

Partial pattern recognition is itself yet another example of IC by PMUS, except that in this case only part of a given pattern is matched and unified with another pattern.

Most of the examples that we might give of inductive inference as partial pattern recognition may also be seen as deduction. If we see blue litmus paper being placed in acid, we expect that it will turn red. This is inductive reasoning based on the normal behaviour of litmus paper. But it may also be presented as a kind of *modus tollens* reasoning like this:

> *If blue litmus paper is placed in acid, then it will turn red.*
> *This piece of litmus paper is being placed in acid.*
> *Therefore, this piece of litmus paper will turn red.*

In a similar way, we may reason that, since Wednesday always follows Tuesday and since today is Tuesday, that tomorrow will be Wednesday. This reasoning may be seen as inductive reasoning based on the regular pattern of Wednesday always following Tuesday or it may be seen as simple syllogistic reasoning with two premises and a conclusion. Whichever way it is seen, any example like this may be interpreted as partial pattern recognition which itself may be interpreted as IC by PMUS.

The interpretation of deduction as partial pattern matching, just described, does not precisely capture the formal view of *modus ponens* deduction which incorporates the concept of ‘material implication’ which itself incorporates a concept of negation. The way in which formal concepts of deduction may be accommodated in a framework of IC by PMUS is discussed in [23, Section 5.1].

5 IC and the Elements of Logic

This section considers in relatively concrete terms how IC may be seen at work in some aspects of current formalisms and mechanisms of logic. Some of the ideas described in this section are developed more fully in the accompanying article [23].

5.1 Unification in Logic

Unification as it is being used in these two articles means a simple merging of identical patterns. This has an obvious role in the reduction of redundancy in any body of information and thus the compression of that information (as described in Section 2.2).

‘Unification’ in the sense of logic means making substitutions for variables in two structures so that the two structures become the same. If the two identical structures are merged, which is at least implicit in logical systems, this will also clearly have the effect of reducing redundancy in relevant structures and thus compressing them.

“All present day deduction systems - whether they are based on resolution or not - have a unification algorithm for first- or higher-order terms as their essential component: it is the ‘addition and multiplication’ of deduction work.” [2, p. 58]. The importance in deduction systems of unification in the sense of logic, and its obvious relevance to IC, is one of the strongest clues suggesting that logic in its entirety might usefully be understood as IC.

5.2 Sets and the Union and Intersection of Sets.

IC is implicit in the concept of a set because of the rule that no element of a set may appear more than once in the set. In other words, any redundancy represented by the appearance of two or more instances of any element in a collection of elements must have been eliminated before the collection can be called a set.

IC is also implicit in the concepts of ‘union’ and ‘intersection’ of sets. The ‘union’ of two sets means forming their elements into a single collection and then applying IC to eliminate replicated instances of any element. If each set of identical instances is unified in the sense that it is merged into one instance, then, within the union of the two original sets, the set of unified elements is the ‘intersection’ of those two sets.

In short, IC is intimately related to these three basic ideas in logic.

5.3 Chunking with Tags

As described in Section 2.2, a basic idea in most of the ‘standard’ methods for IC is to unify repeated instances of a pattern into a single ‘chunk’, to label that chunk with a relatively short ‘code’ or ‘tag’ and to use that tag as a marker for the positions of all instances of the chunk which have been deleted. As previously mentioned, this is a common device in ordinary language where a relatively long phrase may be shortened to a few letters like ‘IC’ or ‘PMUS’. Many of the ordinary words of natural language may be regarded as tags for the relatively complex bodies of information which are their meanings.

This idea can be seen at work in rewrite rules, at least when they are used so that there is always a single symbol on one side of each rule (usually the left-hand side) and two or more symbols on the other side. When rewrite rules are used in this way, the symbol on the ‘small’ side may be seen as a tag and the symbols on the ‘large’ side may be seen as a chunk of information which is labelled by the tag. Rewrite rules like this can provide an effective means of reducing a large body of information to a compressed version expressed in a ‘grammar’.

Another example of the chuning-with-tags idea is the relationship between a function and its name. The former is normally a relatively large body of information and the latter is a relatively short identifier which, with our without associated arguments, may be used to mark the use of a function in each context where it is required without the necessity of writing out the whole function.

Further discussion of the way formal notations provide mechanisms for compressing information may be found in [34, 32].

5.4 TRUE, FALSE and IC

In propositional logic and other standard forms of logic, TRUE and FALSE are symmetrical primitive concepts. This kind of symmetry is also found in databases where there is a roughly even distribution of some two-valued attribute such as whether seats in a train are facing forwards or backwards or whether a railway compartment is smoking or non-smoking.

But in systems like Prolog and in ordinary language, there is a striking asymmetry between TRUE and FALSE: everything is assumed to be true unless it is explicitly marked as false. The same asymmetry is found in databases where there is a very uneven distribution of propositions as, for example, a travel agent's database recording direct flights between pairs of cities: the database records the flights that are available and says nothing about the much greater number of possible flights which are not available (cf. [11]).

The asymmetry between TRUE and FALSE, where it exists, seems to be closely related to IC. In the case of the travel agent's database, it would be extremely uneconomical to record every possible flight which is available with an explicit marker as TRUE with a record of every possible flight which is not available with an explicit marker as FALSE. The default assumption that everything in the database is true, saves a lot of space.

This default assumption may be seen as a simple example of inheritance (cf. Section 3.4. See also [23, Section 6]). This saves much space but requires explicit denial where there are exceptions. When Nixon said “I am not a crook,” he was acknowledging the prior expectation that he *was* a crook.

6 Conclusions

This article has reviewed some of the considerations relating to the nature of information, human thinking, inductive inference, and aspects of logic which seem to support the SP conjecture and suggest that information compression may provide a unifying theme for understanding logic.

In the accompanying article [23], this thinking is developed into more concrete proposals about how logic may be understood as information compression by *multiple alignment* of patterns (with unification and search), where ‘multiple alignment’ has a meaning which is close to the meaning of that term in bio-informatics.

I hope that readers will feel that there is sufficient substance in these proposals to justify further investigation and development.

References

1 Abramsky, S., Gabbay, D. M. and Maibaum, T. S. E. (Eds.), *Handbook of Logic in Computer Science*, Oxford: Clarendon Press, Vols. 1-4, 1992.

2 Baader, F. and Siekmann, J. H. 'Unification theory,' in [10], Vol. 2, pp. 41-125, 1994.

3 Barlow, H. B., 'Trigger features, adaptation and economy of impulses.' In K. N. Leibovic (Ed.), *Information Processing in the Nervous System*. New York: Springer, pp 209-230, 1969.

4 Bibel, W. and Eder, E. 'Methods and calculi for deduction,' in [10], Vol. 1, pp. 68-169.

5 Chaitin, G. J. 'Randomness in Arithmetic,' *Scientific American* **259**(1), 80-85, 1988.

6 Chaitin, G. J. *Algorithmic Information Theory*, Cambridge: Cambridge University Press, 1987.

7 Collins, A. M. and Quillian, M. R. 'Experiments on semantic memory and language comprehension,' in L. W. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, pp. 117-147, 1972.

8 Copi, I. M. *Introduction to Logic*, New York: MacMillan, 1986.

9 Eisinger, N. and Ohlbach, H. J. 'Deduction systems based on resolution,' in [10], Vol. 1, pp. 183-271.

10 Gabbay, D. M., Hogger, C. J. and Robinson, J. A. (Eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vols. 1-3, Oxford: Clarendon Press, 1993-4.

11 Ginsberg, M. L. 'AI and nonmonotonic reasoning,' in [10], Vol. 3, pp. 1-33, 1994.

12 Horty, J. F. 'Some direct theories of nonmonotonic inheritance,' in [10], Vol. 3, pp. 111-187, 1994.

13 Lakatos, I. 'Falsification and the methodology of scientific research programmes,' in I. Lakatos and A. E. Musgrave (Eds.), *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, pp. 91-196, 1970.

14 Li, M. and Vitányi, P. *An Introduction to Kolmogorov Complexity and Its Applications*, New York: Springer-Verlag, 1993.

15 Mahowald, M. A. and Mead, C., 'The silicon retina,' *Scientific American* **264**(5), 40-47, 1991.

16 Popper, K. R. *Conjectures and Refutations: the Growth of Scientific Knowledge*, London: Routledge and Paul, 1962.

17 Shannon, C. E. and Weaver, W. *The Mathematical Theory of Communication*, Urbana: University of Illinois Press, 1949.

18 Solomonoff, R. J., 'A formal theory of inductive inference. Parts I and II,' *Information and Control* **7**, 1-22 and 224-254, 1964.

19 Storer, J. A. *Data Compression: Methods and Theory*, Rockville, Maryland: Computer

Science Press, 1988.

20 Uspensky, V. A. ‘Kolmogorov and mathematical logic,’ *Journal of Symbolic Logic* **57**(2): 385-412, 1992.

21 Von Békésy, G., *Sensory Inhibition*, Princeton, NJ: Princeton University Press, 1967.

22 Watanabe, S., ‘Pattern recognition as information compression,’ in S. Watanabe (Ed.), *Frontiers of Pattern Recognition*, New York: Academic Press, 1972.

23 Wolff, J. G. ‘Multiple alignment and logic,’ *Bulletin of the IGPL*, this volume.

24 Wolff, J. G. ‘Multiple alignment and computing’ (submitted).

25 Wolff, J. G. ‘Computing as compression by multiple alignment, unification and search (1),’ (submitted).

26 Wolff, J. G. ‘Computing as compression by multiple alignment, unification and search (2),’ (submitted).

27 Wolff, J. G. ‘Information compression by multiple alignment, unification and search as a general theory of computing,’ (in preparation).

28 Wolff, J. G. ‘Learning and reasoning as information compression by multiple alignment, unification and search,’ To appear in *Computational Learning and Probabilistic Reasoning*, A. Gammerman (Ed.), New York: Wiley. This is a revised version of ‘Learning and reasoning as compression by multiple alignment and unification’ presented at Applied Decision Technologies ‘95, Brunel University, April 1995. A copy of that paper is in the Proceedings for Stream 1 (Computational Learning and Probabilistic Reasoning), pp. 223-236.

29 Wolff, J. G. ‘Computing as compression: an overview of the SP theory and system,’ *New Generation Computing* **13**: 187-214, 1995.

30 Wolff, J. G. ‘Computing as compression: SP20’, *New Generation Computing* **13**: 215-241, 1995.

31 Wolff, J. G. ‘A scaleable technique for best-match retrieval of sequential information using metrics-guided search,’ *Journal of Information Science* **20** (1): 16-28, 1994.

32 Wolff, J. G. ‘Towards a new concept of software,’ *Software Engineering Journal* **9** (1): 27-38, 1994.

33 Wolff, J. G. ‘Computing and information compression: a reply,’ *AI Communications* **7** *(*3/4): 203-219, 1994.

34 Wolff, J. G. ‘Computing, cognition and information compression,’ *AI Communications* **6** *(*2): 107-127, 1993.

35 Wolff, J. G. *Towards a Theory of Cognition and Computing*, Chichester: Ellis Horwood, 1991.

36 Wolff, J. G. ‘Simplicity and power - some unifying ideas in computing*,’ Computer Journal* **33** *(*6): 518-534, 1990.

37 Wolff, J. G., 'Learning syntax and meanings through optimization and distributional analysis,' in Y. Levy, I. M. Schlesinger and M. D. S. Braine (Eds.), *Categories and Processes in Language Acquisition*, Hillsdale, N.J.: Lawrence, Erlbaum, pp. 179-215, 1988.