Neural Mechanisms for Information Compression by Multiple Alignment, Unification and Search

J Gerard Wolff

CognitionResearch.org.uk,
Telephone: +44(0)1248 712962.
E-mail: gerry@cognitionresearch.org.uk.

March 3, 2003

Abstract

This article describes how an abstract framework for perception and cognition may be realised in terms of neural mechanisms and neural processing.

This framework—called *information compression by multiple alignment, unification and search* (ICMAUS)—has been developed in previous research as a generalized model of *any* system for processing information, either natural or artificial. Applications of the framework include the representation and integration of diverse kinds of knowledge, analysis and production of natural language, unsupervised inductive learning, fuzzy pattern recognition, recognition through multiple levels of abstraction, probabilistic 'deduction' and abduction, chains of reasoning, nonmonotonic reasoning, 'explaining away', solving geometric analogy problems, and others.

A key idea in the ICMAUS framework is that information compression is important both as a means of economising on the transmission or storage of information and also as the basis of probabilistic reasoning.

The proposals in this article may be seen as an extension and development of Hebb's [1949] concept of a 'cell assembly'. The article describes how the concept of 'pattern' in the ICMAUS framework may be mapped onto a version of the cell assembly concept and the way in which neural mechanisms may achieve the effect of 'multiple alignment' in the ICMAUS framework.

By contrast with the Hebbian concept of a cell assembly, it is proposed here that any one neuron can belong in one assembly and *only* one assembly. A key feature of present proposals, which is not part of the Hebbian concept, is that any cell assembly may contain 'references' or 'codes' that serve to identify one or more other cell assemblies. This mechanism allows information to be stored in a compressed form, it provides a robust mechanism by which assemblies may be connected to form hierarchies and other kinds of structure, it means that assemblies can express abstract concepts, and it provides solutions to some of the other problems associated with cell assemblies.

Drawing on insights derived from the ICMAUS framework, the article also describes how learning may be achieved with neural mechanisms. This concept of learning is significantly different from the Hebbian concept and appears to provide a better account of what we know about human learning.

1 Introduction

In the last few years, I have been developing a conceptual framework—*information compression by multiple alignment, unification and search* (ICMAUS)—that aims to integrate concepts in computing, AI and cognitive science.

At its most abstract level, the framework is intended to model any kind of system for processing information, either natural or artificial. However, much of the inspiration for this work came from established ideas about perception and cognition in humans and other animals. And the framework may be viewed as a model for natural kinds of information processing such as the analysis and production of natural language, recognition of patterns and objects despite distortion or omission of information, probabilistic kinds of reasoning, unsupervised learning, and others.

A key idea in the ICMAUS framework is that information compression is important both as a means of reducing the volume of information (leading to economies in the transmission or storage of information) and also as the basis of probabilistic reasoning. With respect to volume reductions, this idea contrasts with Barlow's recent views [2001a, 2001b], discussed briefly in Section 8.4 towards the end of the article.

In work to date, the ICMAUS framework has been developed in purely abstract terms without reference to the anatomy or physiology of neural tissue. The main purpose of this article is to consider, as far as current knowledge allows, possible ways in which the abstract concepts that have been developed within the ICMAUS framework may be mapped on to structures and mechanisms in the brain. It will be convenient to refer to these proposals as 'ICMAUS-neural' or 'ICMAUS-N'.

The main focus of this article is on the storage of knowledge and the way in which sensory data may connect with previously-learned knowledge in perception and learning. These are areas where there is a relative paucity of neurophysiological evidence compared with what is now known about the processing of sensory data in sense organs, thalamus and cortex. The ICMAUS framework, together with existing knowledge about the way in which neural tissue normally works, is a source of hypotheses about the organisation and workings of the brain.

To anticipate a little, it is proposed that 'patterns' in the ICMAUS framework are realised with structures resembling Hebb's [1949] concept of a 'cell assembly'. By contrast with that concept, it is proposed here that any one neuron can belong in one assembly and *only* one assembly. However, any assembly may contain neurons that serve as 'references', 'codes' or 'identifiers' for one or more other assemblies. This mechanism allows information to be stored in a compressed form, it provides a robust mechanism by which assemblies may be connected to form hierarchies and other kinds of structure, it means that assemblies can express abstract concepts, and it provides solutions to some of the other problems associated with cell assemblies.

The next section describes the ICMAUS framework and associated ideas in outline. The sections after that consider how the organisation and operation of the ICMAUS framework

2

may be achieved with neural mechanisms.

2 The ICMAUS framework

The outline of the ICMAUS framework presented here is intended to provide sufficient detail for the purpose of discussing neural mechanisms without swamping the reader. A more comprehensive overview may be found in Wolff [to appear] and even more detail in articles discussing how the framework relates to mathematics and logic [Wolff, 2002a], analysis and production of language [Wolff, 2000], the nature of 'computing' [Wolff, 1999a], probabilistic kinds of reasoning [Wolff, 1999b, 2001a], and unsupervised learning [Wolff, 2002b].

2.1 Information compression

Given that IC is a central plank of the ICMAUS framework, a few words may be helpful about the nature of IC. The essence of IC is the extraction of *redundancy* from information. In terms of communication, 'redundant' information is 'surplus to requirements' because it repeats what has already been expressed. However, redundant information can be useful for other reasons (see Section 3.3, below). There may, on occasion, be a case for discarding non-redundant information as well as redundant information—leading to 'lossy' IC. However, the main focus in this article will be IC by the extraction of redundancy.

Mathematical treatments of IC sometimes have the effect of obscuring a very simple idea that can be seen most clearly in the simpler 'standard' methods for IC [see, for example, Storer, 1988] and, arguably, lies at the heart of *all* methods for IC:

> *If a pattern repeats two or more times in a body of information then the information may be compressed by merging or 'unifying' the repeated instances of the pattern to make one.*

A repeated pattern is an expression of redundancy in information and unification is the means by which that redundancy is extracted. Clearly, large patterns yield more compression than small ones and frequent patterns yield more compression than rarer ones. For any given size of pattern, there is a minimum frequency for achieving compression. Large patterns can yield compression when frequencies are low (even as small as 2) but with smaller patterns, the minimum frequency is larger.

Since unification destroys information about the positions of all but one of the patterns that are unified, these positions are normally marked by some kind of relatively short 'reference', 'identifier' or 'code'. As a general rule, patterns that occur frequently should be given shorter codes than patterns that are rare. This is the central idea in Huffman coding and related techniques [see Cover and Thomas, 1991].

These kinds of ideas are widespread in our everyday experience, not confined to computer-based methods for IC. We make frequent use of abbreviations like 'EU' for 'European Union' or 'TV' for 'television'. In this article, 'IC' is short for the expression 'information compression'. Any kind of reference to a book or article may be seen as a code for the details given in the bibliography. And these details may be seen as a relatively short code for the entire document. Any 'content' word in a natural language (e.g., 'house' or 'cooking') and any kind of name of a person or thing, may be seen as a short code for the relatively complex

concept that it represents. Broadly in keeping with Huffman coding and related techniques, frequencies of words in most natural languages are inversely related to their sizes [Zipf, 1935].[1]

2.1.1 Information compression, frequency and counting

As we have seen, frequency is important for IC, both in terms of the number of patterns that are unified and also because frequent patterns can be given shorter codes than rare ones.

But there is another connection between IC and frequency that is less widely recognised: frequency implies counting, counting implies recognition, and recognition implies IC by the unification of matching patterns.

In case the last two points seem obscure, consider what is entailed in counting the number of apples in a bowl of fruit. It is not possible to do this without recognising that one apple is the 'same' (at some level of abstraction) as the next one or that they are all the 'same' as some abstract representation of the concept 'apple'. In itself, the process of counting implies the merging or unification of several instances into a single concept. Thus counting, in itself, implies IC.

2.2 Origin of the ICMAUS proposals

From the writings of William of Ockham in the 14th century and Ernst Mach and Karl Pearson in the 19th century, a principle of parsimony has been recognised as relevant to an understanding of thinking, perception and the workings of the brain and other parts of the nervous system. Other publications in this tradition include Atick [1992], Attneave [1954], Baddeley et al. [2000], Barlow [1959, 1961, 1969, 1972, 1983, 1997], Garner [1974], Oldfield [1954], Sakitt and Barlow [1982], Snyder and Barlow [1988], von Békésy [1967], Watanabe [1972], Wolff [1993, 1988], Zipf [1949]. Nice reviews of some of the issues and pointers to related research are provided by Chater [1996, 1999]. Although the terminology varies, a theme of these writings is that neural tissue exploits *information compression* (IC) to make economies in the transmission or storage information. We shall return to this issue in Sections 3 and 8.4, below.

A related idea, developed by Solomonoff [1964], Rissanen [1978], Wallace and Boulton [1968] and others, is that there is an intimate connection between information compression and the inductive prediction of the future from the past [see also Li and Vitányi, 1997, Solomonoff, 1997]. The connection between these two things—which may at first sight seem obscure—lies in the detection of repeating patterns: standard methods for IC exploit repetition of patterns in data (as noted above) and inductive predictions are derived from recurrent patterns such as night following day or the annual repetition of the four seasons.

Although IC is the key in this area of thinking, the field is often referred to as Minimum Message Length encoding (MML), Minimum Description Length encoding (MDL) or, more

---

[1]It is perfectly true that natural languages like English, in both spoken and written forms, contain quite a lot of redundancy (see Section 5.4, below). The existence of this redundancy should not disguise the huge compression of information that can be achieved by the use of a relatively short word to refer to something that is complex. Some residual redundancy is needed so that the retrieval process can function correctly. And the residual redundancy in natural language has a rôle to play in reducing errors when we are listening to speech in noisy conditions or reading text with smudges or other distortions.

generally, Minimum Length Encoding (MLE). The last of these terms will be used in the rest of this paper.

The significance of these kinds of ideas became increasingly evident in work I did some time ago developing computer models of language learning (see Wolff [1988] and earlier papers cited there). To a large extent, this kind of learning may be understood as a process of compressing linguistic input by looking for patterns that match each other and merging or 'unifying' patterns that are the same. Because there are typically many alternative ways in which patterns may be unified, some better than others, there is also a need to *search* amongst the many possibilities for unifications yielding relatively large amounts of compression.

Development of the ICMAUS framework was inspired originally by the parallels that seemed to exist between the Prolog system, designed originally for theorem proving, and my models of language learning. Although IC is not a recognised feature of Prolog, a process of searching for patterns that match each other is fundamental in that system and the merging of matching patterns is an important part of 'unification' as that term is understood in logic. It seemed possible that IC might have the same fundamental role in logic—and perhaps also in 'computing' and mathematics—as it does in the learning of a language [Wolff, 1990, 1993].[2]

These observations led to the thought that it might be possible to integrate unsupervised learning and logical inference within a single system, dedicated to IC by pattern matching, unification and search. Further thinking suggested that the scope of this integrated system might be expanded to include such things as the retrieval of information from memory, fuzzy pattern recognition, analysis and production of language, probabilistic inference, and other aspects of perception and cognition. It turned out later that the explanatory 'power' of the system could be greatly increased if the relatively general concept of 'pattern matching' was replaced by a more specific concept of 'multiple alignment', as described below.

2.3 Organisation of the ICMAUS framework

In broad terms, the ICMAUS framework works like this. It receives data (designated 'New') from its environment and adds these data to a body of stored knowledge (designated 'Old'). At the same time, it tries to compress the information as much as possible by searching for full or partial matches between patterns and unifying patterns or sub-patterns that are the same. In the course of trying to compress information, the system builds multiple alignments of the kind that will be seen below.

Generally speaking, 'New' information may be equated with sensory information but information within the system may sometimes play the same role (see Sections 2.5.1 and 7, below).

---

[2]In a similar vein, Barlow [1959] wrote that "The tools of logical reasoning appear to be the same as those needed for optimal coding, so perhaps they can be regarded as the verbal expression of rules for handling facts which our nervous system uses constantly and automatically to reduce the barrage of sensory impulses to useable proportions." (p. 555).

2.3.1 A 'universal' format for knowledge

All information in the system is expressed as arrays or *patterns* of symbols. In work done to date, the focus has been on one-dimensional strings or sequences of symbols. But it is envisaged that, at some stage, the ideas will be generalized to patterns in two dimensions and possibly more.

This very simple 'universal' format for knowledge has been adopted with the expectation that it would facilitate the representation of diverse kinds of knowledge and their integration. Notwithstanding the simplicity of the format, the processing mechanisms provided within the ICMAUS framework mean that it is possible to model such things as grammatical rules (with 'context-sensitive' power), if-then rules, networks, class-inclusion hierarchies and heterarchies (with inheritance of attributes), and part-whole hierarchies[3]. Examples will be seen in Sections 2.4 and 2.5.3, below, and others may be found in the sources cited above.

This format for knowledge is also motivated by the thought that much of our knowledge derives ultimately from sensory inputs, especially vision, and most sensory inputs map naturally onto arrays of symbols in one or two dimensions. This idea sits comfortably with the observation that there is a relatively direct mapping between the retina and each of the areas of the visual cortex, between the skin and the somatosensory cortex, and likewise for other senses. It seems reasonable to suppose that this kind of mapping is a general feature of the way the brain handles information.[4]

A *symbol* in the system is an atomic 'mark' that can be compared with any other symbol to determine whether it is the same or different. Any meaning attaching to a symbol must be expressed in the form of other symbols: there is no place for 'hidden' meanings like 'multiply' for the arithmetic symbol '×' or 'add' for the symbol '+'.[5]

The example presented in Section 2.4, below, will put some flesh on this rather bare description.

2.3.2 Minimum length encoding

The whole system is founded on MLE principles mentioned in Section 2.2, above. The key idea here is that in grammar induction and related kinds of learning (deriving a grammar from a 'corpus' of raw data), one should seek to minimise $(G + E)$, where G is the size of the grammar (in bits or equivalent measure) and E is the size of the corpus after it has been encoded in terms of the grammar. An example of this kind of encoding is presented in Section 2.4, below.

---

[3]In a class 'heterarchy', any one class may belong in two or more higher-level classes and may inherit attributes from those classes. In object-oriented software systems, this is termed 'multiple inheritance'.

[4]No attempt will be made in this article to consider at any length how we may encode our knowledge of the three-dimensional shapes of objects or the arrangement of objects in three-dimensional space. Given that the ICMAUS concept of 'pattern' can, in principle, be generalised to three or more dimension, one could envisage a scheme in which 3D information is encoded with 3D patterns. However, in biological terms, this may not make such good sense. Given the predominantly two-dimensional nature of the retina and the cortex, it may be more plausible to suppose that we encode our knowledge of three dimensions using two-dimensional patterns in the manner of plans and elevations used by architects and engineers.

[5]An apparent exception to the rule of 'no hidden meanings' is that there is a distinction, described below, between 'identification' symbols and 'contents' symbols. But this is really part of the mechanism by which the system organises its knowledge, not part of the knowledge itself.

Trying to minimise G by itself may well yield a very small grammar but it is likely to be one that is relatively inefficient for the encoding of data. Likewise, we may well succeed in finding a grammar with a small value for E but it is likely to be rather large. By seeking to minimise $(G + E)$, we guard against these two extremes [see Solomonoff, 1997].

2.3.3 Computer models

Two computer models of the ICMAUS framework have been developed:

- SP61 is a partial realisation of the framework that does not transfer any information from New to Old. Given one pattern in New and one or more patterns in Old, this model tries to compress the New pattern as much as possible by building multiple alignments of the kind that will be seen below. SP61 also contains procedures for calculating the probabilities of inferences that may be drawn from alignments (see Section 2.5.4, below). This model, which is relatively robust and stable, is described quite fully in Wolff [2000].

- SP70 realises all the main elements of the framework, including the transfer of information from New to Old. In addition to building multiple alignments like SP61, the model compiles one or more alternative 'grammars' for the information in New, using MLE principles. This model, and its application to unsupervised learning, is described quite fully in Wolff [2002b]. More work is required to realise the full potential of this model.

Although SP61 is largely a subset of SP70, it is convenient to run it as a separate model for many applications. Most of the examples in this paper are output from the model.

The process for building multiple alignments which is central to both models incorporates a version of 'dynamic programming' [see, for example, Sankoff and Kruskall, 1983] for finding full matches between patterns and good partial matches. A detailed description of this mechanism may be found in Wolff [1994].

Other aspects of the way these models work will be explained at relevant points below.

2.4 An example

The main elements of the multiple alignment concept as it has been developed in this research are illustrated in the example presented here showing how parsing of natural language may be understood in terms of multiple alignment.

For the sake of clarity and to save space, this example and ones presented later are relatively simple. However, this should not be taken to represent the limits of what the models can do. More complex examples may be found in Wolff [2000] and the other sources cited above.

Figure 1 shows the best alignment found by SP61 when New contains the sentence 't h e c a t s l e e p s' and Old contains patterns representing grammatical rules. The meaning of 'best' is described in Section 2.4.4, below.

By convention, the New pattern is always in the top row of the alignment (row 0) with patterns from Old in the rows underneath, one pattern per row. The order of the patterns

```
0                       t h e           c a t             s l e e p   s   0
                        | | |           | | |             | | | | |     |
1                       | | |           | | |           < Vstem 2 s l e e p > |   1
                        | | |           | | |             |   |           | |
2                       | | |           | | |         < V Vs < Vstem         > s > 2
                        | | |           | | |           | | |                 |
3 S Num    ; < NP       | | |           | | |       > < V |                   > 3
      |      | | |      | | |           | | |         |   |
4     |      | | |      | | |         < Nstem 3 c a t >   |                       4
      |      | | |      | | |           |   |       |     |
5     |      | | |      | | |     < N Ns < Nstem       > > |                       5
      |      | | |      | | |       | | |               | |
6     |      | | | < D 0 t h e > | | |                 | |                       6
      |      | | | | |     | | | |                     | |
7     |      | < NP < D       > < N |                 > >                         7
      |        |           |       |
8     Num SNG ;                     Ns                  Vs                        8
```

Figure 1: The best alignment found by SP61 with 't h e c a t s l e e p s' in New and patterns representing grammatical rules in Old.

below the top row is arbitrary. It is the result of how alignments were built up by the program and has no special significance.

A pattern like '< NP < D > < N > >' in row 7 of the alignment expresses the idea that a noun phrase ('NP') is composed of a determiner ('D') followed by a noun ('N'). In a context-free phrase-structure grammar (CF-PSG), this would be expressed with a rule like 'NP → D N'.

If we ignore row 8, the whole alignment may be seen to achieve the effect of a context-free parsing, dividing the sentence into its constituent words, identifying 't h e c a t' as a noun phrase, marking each word with its grammatical class, and, within the verb 's l e e p s', marking the distinction between the stem and the suffix.

2.4.1 Context-sensitive power

Although some of the patterns in the alignment are similar to rules in a CF-PSG, the whole system has the expressive power of a context-sensitive system. This can be seen in row 8 of the alignment where the pattern 'Num SNG ; Ns Vs' marks the 'number' dependency between the singular noun in the subject of the sentence ('c a t') and the singular verb ('s l e e p s'). A basic CF-PSG, without augmentation, cannot handle this kind of dependency.

2.4.2 'Identification' symbols and 'contents' symbols

Within each pattern in Old, there is a distinction between 'identification' (ID) symbols and 'contents' (C) symbols. The former serve to identify the pattern or otherwise define its relationship with other patterns, while the latter represent the contents or substance of the pattern.

For example, in the pattern '< NP < D > < N > >' in Figure 1, the ID-symbols are the initial left bracket ('<') and the terminating right bracket ('>') together with the symbol 'NP' following the initial left bracket. All other symbols in the pattern are C-symbols. In the pattern '< Nstem 3 c a t >', the ID-symbols are the initial and terminal brackets, as

before, plus the symbols 'Nstem' and '3', while the C-symbols in that pattern are 'c', 'a' and 't'. In the pattern 'Num SNG ; Ns Vs', the first three symbols are ID-symbols and the last two are C-symbols.

In SP61—which does not attempt any learning—each symbol within each pattern for Old (supplied to the program by the user) is marked to show whether it is an ID-symbol or a C-symbol. In SP70—which creates patterns and adds them to Old in the course of learning— the distinction between ID-symbols and C-symbols within any one pattern is marked by the system when the pattern is created.

2.4.3 Compression, references, hierarchies, heterarchies, recursion and abstraction

If the patterns in Old have been compiled in accordance with MLE principles (as in SP70), then they may be seen as a reflection or distillation of recurrent—and thus redundant— patterns in the world. A pattern like '< Nstem 3 c a t >' expresses a relatively concrete and relatively localised pattern of redundancy while others, such as the 'sentence' pattern 'S Num ; < NP > < V >', express patterns of redundancy with a broader sweep and a higher level of abstraction.

The ID-symbols in each pattern serve as an abbreviated code by which the pattern may be referenced. As we saw in Section 2.1, this is a key feature of standard methods for achieving IC. The references themselves are copies of these symbols that appear as C-symbols in other patterns. For example, the C-symbols '< N >' in the 'noun phrase' pattern '< NP < D > < N > >' may be seen as a reference to any one of the set of patterns belonging to the class 'noun'.

This reference mechanism provides the key to the representation of part-whole hierarchies (e.g., the hierarchy of structures within a typical sentence), class hierarchies or heterarchies with inheritance of attributes (Section 2.5.3) and other kinds of knowledge structure. It also provides the key to the representation of abstract concepts like 'sentence' and the representation of recursive (self-referential) structures [see Wolff, 1999a].

2.4.4 The evaluation of alignments

A combination of ID-symbols, derived from an alignment like the one shown in Figure 1, can provide an abbreviated code for the entire sentence pattern (or other pattern) held in New. A 'good' alignment is one where this code is relatively small in terms of the number of bits of information that it contains. The 'compression score' for an alignment is the difference between the size (in bits) of the New pattern in its original form and its size after it has been encoded.

The procedure for deriving an encoding from an alignment is quite simple: scan the alignment from left to right looking for columns containing a single instance of an ID-symbol, not matched to any other symbol. The encoding is simply the symbols that have been found by this procedure, in the same order as they appear in the alignment. The encoding derived in this way from the alignment in Figure 1 is 'S SNG 0 3 2'. This is smaller than the original sentence in terms of the number of symbols it contains. It is also smaller in terms of the number of bits it contains, calculated as described in Wolff [2000].

9

2.4.5 Ambiguity

For any given combination of sentence and grammatical patterns, SP61 normally produces a set of alternative alignments, each one with a compression score as just described. This framework naturally accommodates ambiguities in parsing as, for example, with the second sentence in Groucho Marx's "Time flies like an arrow. Fruit flies like a banana". In cases like this, the system normally delivers an alignment for each of the alternative analyses, each one with a compression score that is relatively high.

If disambiguating context is provided, the system may still produce alternative analyses but one of them will normally have a much higher compression score than any of the alternatives.

2.5 Applications

This subsection outlines the range of applications of the ICMAUS framework, with an emphasis on those that reflect aspects of human cognition. Much more detail may be found in the sources that are cited.

2.5.1 Analysis and production of natural language

We have seen already how the ICMAUS framework can model the parsing of natural language, including ambiguities in parsing, in purely syntactic terms.

Although the ICMAUS framework can model non-syntactic forms such as class inclusion with inheritance of attributes (Section 2.5.3, below), no serious attempt has yet been made to examine how the framework may support the integration of syntax with semantics and the derivation of meanings from surface forms. However, the framework has been designed with the intention that it should allow the close interaction of syntax and semantics. As noted in Section 2.3.1, a simple 'universal' format has been chosen for representing knowledge, to accommodate diverse kinds of knowledge and their seamless integration.

An interesting feature of the ICMAUS framework is that, *without any modification*, it can support the production of language as well as its analysis. If SP61 is run again, with the sentence in New replaced by the encoded form of the sentence ('S SNG 0 3 2'), the best alignment found by the system is exactly the same as before except that row 0 contains the encoded pattern, with each symbol in that pattern aligned with matching symbols in the rows below. The original sentence has, in effect, been recreated because the alignment contains the words of the sentence in their correct order. This is an example of the possibility noted earlier (Section 2.3) where the role of 'New' information is played by information within the system rather than by sensory data.

It is envisaged that the production of sentences from meanings may be modelled in a similar way. Instead of using a code pattern like 'S SNG 0 3 2' to drive the production process, some kind of semantic structure may be used instead.

More detail about the way in which the ICMAUS framework may be applied to the analysis and production of natural language, with a range of examples, may be found in Wolff [2000] [see also Wolff, to appear].

10

2.5.2   Fuzzy pattern recognition and best-match information retrieval

The dynamic programming built into both computer models [Wolff, 1994] gives them a human-like capability for recognising patterns despite errors of omission, commission or substitution.

In a similar way, they can model our ability to recall information from fragmentary cues. The ICMAUS system serves as a very flexible form of 'content-addressable' memory in which any sub-set of the symbols in a pattern may serve as a retrieval key for the whole pattern.

With multiple alignments, it is also possible to model 'semantic' or 'indirect' retrieval of information, e.g., 'Jumbo' retrieves 'plane' and 'plane' retrieves 'flight'.

As with parsing of natural language, the ICMAUS framework naturally accommodates ambiguities in our perceptions and the way in which ambiguities can be resolved by appropriate context.

Further information may be found in Wolff [1999b].

2.5.3   Class hierarchies and heterarchies with inheritance of attributes

Without any *ad hoc* additions, the ICMAUS framework can also support the recognition of patterns and objects at varying levels of abstraction, with 'inheritance' of attributes from high levels of abstraction to lower levels. A very simple example is shown in Figure 2 where an individual ('Tibs') with the attribute 'tabby' is recognised as belonging in the class 'cat' (with the attribute 'purrs') which itself belongs in the class 'mammal' (with the attribute 'furry'). Although issues in this area have been the subject of some debate, it does seem that mechanisms of this sort are an important part of human perception and cognition.

```
0                              furry   purrs   tabby   0
                                 |       |       |
1 < Tibs < cat                   |       |     > tabby > 1
          |  |                   |       |     |
2         < cat < mammal  |    > purrs >          2
             |    |       |    |
3                < mammal furry >                 3
```
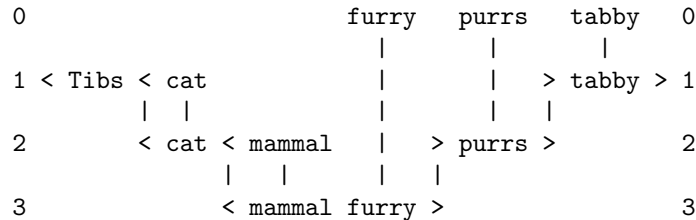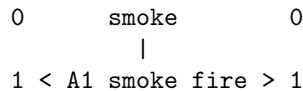
Figure 2: A very simple example of the way in which a multiple alignment can model recognition at multiple levels of abstraction, with inheritance of attributes.

More information about these aspects of the ICMAUS framework, with examples, may be found in Wolff [1999b] [see also Wolff, to appear, 2001a,b,c].

2.5.4   Probabilistic reasoning

Any column within an alignment that does *not* contain a symbol from the New pattern in the alignment (row 0) represents a prediction or inference that may be drawn from the alignment. For example, in an alignment like this:

```
0       smoke        0
          |
1 < A1 smoke fire > 1
```

11

the pattern from New (in row 0) contains the single symbol 'smoke' which is aligned with the matching symbol in a pattern in Old representing the association between smoke and fire. The unmatched C-symbol 'fire' represents an inference that may be drawn from the alignment (we may ignore the ID-symbols in the alignment, '<', 'A1' and '>').

This kind of inference may be drawn from any alignment, no matter how complex, provided it contains C-symbols that are not matched to corresponding symbols in New. In addition to simple one-step inferences like the one shown here, it is possible to model chains of inference and other kinds of composite reasoning including reasoning with default values, nonmonotonic reasoning, 'explaining away' [see Pearl, 1988] and the solution of geometric analogy problems.

The whole system is inherently probabilistic. With SP61, it is possible to calculate absolute and relative probabilities for any inference or set of alternative inferences that may drawn from an alignment or set of alternative alignments.

More detail and a range of examples may be found in Wolff [1999b] [see also Wolff, 1998].

2.5.5 Unsupervised learning

In the broad outline sketched in Section 2.3, the ICMAUS framework may be seen as a system that learns by assimilating 'raw' information from its environment and distilling the essence of that information by a process of information compression.

Although there is work still to be done in refinement and development, the whole framework is now realised in the SP70 computer model. This model is able to abstract simple grammars from appropriate data without any kind of external 'teacher' or the provision of 'negative' samples or the grading of samples from simple to complex [cf. Gold, 1967]. In short, it is an *unsupervised* model of learning.

A relatively full account of the SP70 model may be found in Wolff [2002b].

2.5.6 Computing, mathematics and logic

The ICMAUS framework provides an interpretation for models of 'computing' such as the Universal Turing Machine and the Post Canonical System and may be seen as alternative to those theories [Wolff, 1999a]. The framework also provides an interpretation for a range of concepts in mathematics and logic [Wolff, 2002a].

Although these aspects of the ICMAUS framework may be seen to be largely independent of human psychology, it should not be forgotten that until artificial systems came on the scene, a 'computer' was a person who did computations. Notwithstanding the existence of automated systems, much of the work of discovery and proof in mathematics and logic is still done by human brains. In short, the ICMAUS framework may be viewed as a model of abstract concepts in computing, mathematics and logic but it may also be viewed as a conjectural model of human thought processes in those domains.

3 Global considerations: arguments from biology and engineering

The immense difficulties in studying what is going on inside a living brain means that there are enormous gaps in our knowledge. Trying to piece together a coherent theory is a bit like reconstructing an ancient document from a few charred fragments of parchment or papyrus. Various kinds of inference are needed to build understanding in the sea of ignorance that surrounds the scattered islands of knowledge.

To the extent that the ICMAUS theory provides a coherent account of cognitive phenomena, it may itself suggest predictions about what is going on in the brain. If, as a working hypothesis, we assume that the theory is true, we may expect to find correspondences between elements of the theory and structures or processes in the brain.

The main aim in this and subsequent sections is to consider how the elements of the ICMAUS framework outlined in the previous section may be realised in neural tissue. But before we examine the nuts and bolts of the theory, there are some global considerations that have a bearing on how the ICMAUS theory might relate to concepts in neurophysiology.

Given that brains and nervous systems are the product of evolutionary processes of natural selection, we may evaluate the theory in these terms. Given that biological systems must survive the rigours of the world, we need to consider how the theory may relate to principles of engineering that govern all physical systems, both natural and artificial.

3.1 IC and natural selection

We would expect natural selection to have favoured organisms that process information in an efficient manner, minimising an organism's requirements for information storage (or maximising the amount of information that can be stored for a given 'cost') and maximising the speed with which any given volume of information can be transmitted from one place to another within the nervous system (or minimising the biological cost of transmitting information at a given speed). These considerations may increase our confidence in the long-standing idea—incorporated in the ICMAUS theory—that IC has a central role in perception and cognition (but see Section 8.4, below).

Another reason—perhaps even stronger—for believing that IC is a significant factor in the way nervous systems have evolved is the intimate connection that exists between IC and the inductive prediction of the future from the past (Section 2.2). Arguably, this kind of prediction—in the processes of finding food, avoiding predators and so on—provides the entire *raison d'être* in biological terms for any kind of nervous system. If that is accepted, then we should expect IC to have a central role in the way neural tissues operate.

3.2 Is the 'cost' of IC significant?

It may be argued that the benefits of IC may not always justify the resources required to achieve it. May we not expect to find situations where organisms use information without compressing it?

If storage costs or speed of transmission were the only consideration, this kind of expectation would seem to be justified. But the very close connection between inductive prediction and IC seems to imply that we cannot have the former without the latter. If it is accepted

that prediction is indeed a fundamental function of nervous systems, it seems that that the costs of IC cannot be avoided.

Compression of information may occur in ways that are not obvious. For example, it may on occasions be useful to store information from the environment in 'raw', uncompressed form and this may suggest that IC has no role. But this stored information is of no use unless it can be related to information that arrives later. And the process of relating the incoming information to information that has been stored means matching and unification of patterns—which are the elements of IC.

3.3   IC and the benefits of redundancy in information

A key part of IC is the removal of redundancy from information and this seems to conflict with another principle of information engineering: that redundancy in information can be exploited to reduce errors in processing, to speed up processing and to guard against the risk of losing information.

Any user of a computer system knows that it is very unwise to run the system without keeping backup copies of critical information. By the same token, it seems unlikely that evolution would have endowed us with information processing systems that allow us to keep only a single copy of our hard-won knowledge. And it seems possible that multiple copies of stored information—a form of redundancy—may also have a role to play in reducing errors in processing or speeding up processing or both.

Despite the apparent conflict between the reduction of redundancy resulting from IC and the need for redundancy for the purposes just described, the two requirements can be reconciled very easily. For the reasons given earlier, we would expect a given body of knowledge to be stored or transmitted in a compressed form, with as little redundancy as possible. But for the reasons described in the preceding paragraphs, we may also expect biological systems to keep multiple copies of that compressed body of knowledge. Paradoxical as it may sound, knowledge can be compressed and redundant at the same time.

The number of copies required to safeguard information against accidental loss is not great. Provided they are independent, two copies are very much safer than one—and three copies are sufficient for most purposes.

3.4   IC in nervous systems

As background to the main proposals of this article, this subsection briefly reviews some of the ways in which the operation of nervous systems may be understood in terms of IC:

- The widespread phenomenon of adaptation in neural tissue may be seen as a manifestation of IC [Barlow, 1959, 1969, von Békésy, 1967]. Adaptation has the effect of filtering out the informational redundancy in steady stimulation—thus compressing the data. Typically, there is an enhanced response when stimulation changes which has the effect of capturing the non-redundant parts of the information stream.

- Perceptual recognition of objects and patterns may be understood in terms of information compression [Watanabe, 1972]. By assimilating new information to previously-stored information (with or without variations) we can drastically reduce the huge quantities of redundancy which exists in normal sensory inputs. It is hard to imagine

14

how we could survive if we did not merge new percepts with old ones in this kind of way. Shutting one's eyes and opening them again would result in a complete new record of what we had been looking at, without any relation to the previous record. Merging successive percepts is essential if we are to maintain a reasonably coherent view of the world.

- A closely-related idea is that the concept of motion implies compression of information [Barlow, 1961]. Encoding a sequence of images in terms of entities that move requires much less information than simply storing the raw images. The waterfall illusion shows that one's perception of motion can conflict with one's perception of position and this suggests that there are mechanisms dedicated to the perception of motion that are distinct from mechanisms that detect position. There is evidence for such motion detectors in the neurons of the 'magnocellular' layers of the lateral geniculate body and in neurons to which they project in areas V_1, V_2 and V_5 of the cortex [Nicholls et al., 2001, 434–435].

- Normally, we perceive the brightness or colour of an object as constant, regardless of the brightness or colour of the incident light. This kind of 'perceptual constancy' greatly simplifies and thus compresses what we need to remember about the things that we perceive.

- Lateral inhibition in the mammalian retina allows visual images to be encoded in terms of transitions (edges) between relatively uniform areas. This greatly reduces the volume of information that needs to be transmitted along the relatively low bandwidth of the optic nerve [Barlow, 1959, 1969, von Békésy, 1967] (but see Barlow [2001b, p. 245] for further thoughts on that idea). Similar effects occur in the innervation of the skin and the Basilar membrane [von Békésy, 1967] and in the visual systems of invertebrates such as the horseshoe crab [Ratliff and Hartline, 1959, Ratliff et al., 1963].

- "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." [Barlow, 1969, p. 213]. We exploit the differences between the two messages to gain information about visual depth but we also merge the near-duplicated messages into a single view—and this means information compression.

- Compression of information or, more specifically, principles of Minimum Length Encoding, can explain several aspects of language learning that cannot be satisfactorily explained in any other way [Wolff, 1982, 1988]. As noted in Section 2.1, natural language may itself be seen as an embodiment of the basic principles of IC.

3.5   ICMAUS-N and the modular structure of the brain

A last issue to consider briefly before we proceed to the main proposals is the way in which the ICMAUS theory may relate to the structure of the brain.

The mammalian brain (and other parts of the nervous system) is divided into clearly-defined modules, many of which have a clear association with particular functions—Broca's

areas is concerned with language, the visual cortex deals with vision, and so on. Furthermore, there are clear differences in anatomy and physiology from one module to another.

It might be thought that this modular structure—and the differences amongst modules—rules out any global theory of brain function such as the ICMAUS-N proposals. An alternative view—which is the one adopted in this research—is that the ICMAUS-N theory describes general principles that may seen to apply in a variety of different modules in different parts of the brain. It is entirely possible that the same principles may apply in different modules and, at the same time, there may be differences in how they are applied. It is clear for example, that two-dimensional visual images need to be handled in a way that is rather different from the time-ordered and predominantly one-dimensional patterns of hearing. And yet the matching and unification of patterns—and the building of multiple alignments—are ideas that may be applied in both areas.

4   Neural mechanisms for ICMAUS

This section presents an outline of how the ICMAUS concepts may be realised in neural tissue, while the next section considers aspects of the outline in more detail in the light of available evidence, highlighting areas of uncertainty and issues that need to be resolved.

In broad-brush terms, the proposals are these:

1. An ICMAUS *pattern* in Old may be realised by a network of interconnected cells, similar to Donald Hebb's [1949] concept of a *cell assembly* [see also Hebb, 1958, pp. 103–107]. Because there are significant differences between the ICMAUS-N proposals and Hebb's original concept, the neural realisation of an ICMAUS pattern in Old will be referred to as a *pattern assembly* (PA).

2. An ICMAUS *pattern* in New may be represented in arrays of cells in the sensory systems or elsewhere. In the case of vision for example, these may be in the layers of the visual cortex concerned with early stages of processing (layers 3 and 6) or, perhaps, in the lateral geniculate body. Comparable structures may be found in other parts of the sensory cortex or thalamus.

   Each such structure must be able to respond to a succession of different patterns, as discussed below. In addition to its rôle in processing information, each such structure functions as a buffer within which a pattern from New may be stored pending further processing. Any structure like this that can receive a pattern and store it temporarily will be referred to as a *neural array* (NA). Anything that is either a PA or an NA may be referred to as a 'neural pattern'.

3. Much of the discussion below will focus on the analysis of sensory data. But it should not be forgotten that the role of 'New' may, on occasion, also be played by patterns that are in some sense 'internal' to the system. As we saw in Section 2.5.1, an internally-generated code pattern for a sentence or, perhaps, a semantic structure, may take the role of New. We shall suppose that there is some kind of 'internal' NA to store patterns of that kind, distinct from NAs for sensory data that will be referred to as 'sensory' NAs. It is conceivable, of course, that sensory NAs may also serve as internal NAs.

4. Where an ICMAUS *symbol* represents a fine-grained detail, it may be realised as a single neuron. Where a symbol represents something a bit larger, it may be realised with a PA. It is envisaged that each such 'neural symbol'—a single cell or a small PA—will respond preferentially to a particular pattern of stimulation received from elsewhere within the nervous system.

5. A neuron within a PA that corresponds to a C-symbol within an ICMAUS pattern will be referred to as a C-neuron. Likewise, a neuron that represents an ID-symbol will be called an ID-neuron.

6. Although sensory information is received initially in analogue form, it is widely accepted that the information is converted at an early stage into something like David Marr's [1982] concept of a 'primal sketch' in the domain of vision or something equivalent in other domains. In vision, it seems that sensory data is converted into the digital language of on-centre cells, off-centre cells, cells that are sensitive to lines at particular angles, cells that detect motion, and so on [see, for example, Nicholls et al., 2001, Hubel, 2000]. Other sensory modalities seem to be similar.

7. It is envisaged that, when the cells of a sensory NA become active as a result of sensory input, the pattern of excitation will be transmitted to the C-neurons of a range of PAs. Through an interplay of excitatory and inhibitory processes to be described below, activity will spread to other PAs until sets of PAs have become active in a manner comparable with multiple alignments in the ICMAUS framework.

8. In a similar way, cells of an internal NA may broadcast excitation to ID-neurons of a range of PAs. From that point on, the spread of activity should achieve the formation of multiple alignments in precisely the same way as if the excitation had originated from sensory input.

9. In SP70, learning occurs by adding newly-created patterns to Old, including some derived relatively directly from New patterns and others derived from multiple alignments in which there is partial matching of patterns. When a range of patterns have been added, there is a process of evaluation to differentiate those that are 'good' in terms of the economical encoding of patterns from New from those that are 'bad'. It is envisaged that, in future versions of the model, bad patterns will be removed from Old, leaving only the relatively good ones. This kind of evaluation and purging may occur periodically as patterns from New are processed.

   There seems no reason in principle why a comparable kind of process should not operate with PAs. It seems unlikely that cells would be literally added or removed from within a brain but it seems possible that groups of pre-existing cells could be brought into play as PAs, or the reverse, according to need.

4.1   A schematic representation

In order to summarise the proposals as they have been developed so far and to provide a focus for the discussion that follows, Figure 3 provides a diagrammatic representation of the way in which the neural substrate for ICMAUS concepts may be organised.

17

This example has a 'linguistic' flavour but it should be emphasised that the proposals are intended to apply to all sensory modalities—vision, hearing, touch etc—both individually and in concert. Thus each of the patterns shown in the figure should be interpreted broadly as an analogue of information in *any* modality (or combination of modalities), not merely as a textual representation of some aspect of linguistic structure.

4.1.1   Neural arrays

The pattern 'j o h n r u n s' at the bottom of the figure represents raw sensory data in analogue form. Each letter is intended to represent a low-level feature of the sensory input that will be recognised by the system.

The pattern shown immediately above, with the same letters but with dots in between them, is intended to represent a sensory NA containing detectors (neurons) for low-level features of the image—lines at various angles, colours, luminances, elementary movements and the like. Physiologically, it seems that different kinds of features are processed in different areas of the visual cortex but conceptually they may be seen to belong to one NA.

The entire alphabet of detectors is repeated many times across the NA so that it can respond to any feature at any position within the NA. The schematic representation in the figure is intended to convey the idea that, at each position, an appropriate detector is responding to the corresponding small portion of the sensory input.

4.1.2   Pattern assemblies

Each of the patterns shown in the figure above the NA are intended to represent PAs located somewhere in the brain, most probably in the cortex. Each character within each PA represents a neuron. Angle brackets ('<' and '>') at the beginning and end of each pattern and characters with an oblique angle represent ID-neurons. All other characters represent C-neurons.

Although the PAs shown in the figure are all one dimensional, it seems likely that many PAs (e.g., in the domain of vision) will have two dimensions.

Contrary to what the figure may suggest, it is likely that the lowest-level PAs correspond to features of the sensory input that are relatively small (but larger than the features registered by detectors in the sensory NA). In vision, these might be corners, angles or other small motifs. In speech, they may be structures such as vowels, consonants or dipthongs.

It seems likely that there will be many more 'levels' of PAs than the few levels shown in Figure 3. As indicated above, it seems likely that the higher-level PAs will represent associations bridging several different sensory modalities.

It seems likely that each PA will contain many more cells than is suggested by the examples in the figure.

Not shown in the figure are the connections envisaged between neurons within each PA. Each neuron will have connections to its immediate neighbours at least and, perhaps, beyond.

4.1.3   Connections between neural patterns

Each of the neurons within the sensory NA is connected to zero or more C-neurons representing the same low-level feature in zero or more PAs. The figure shows only the connections
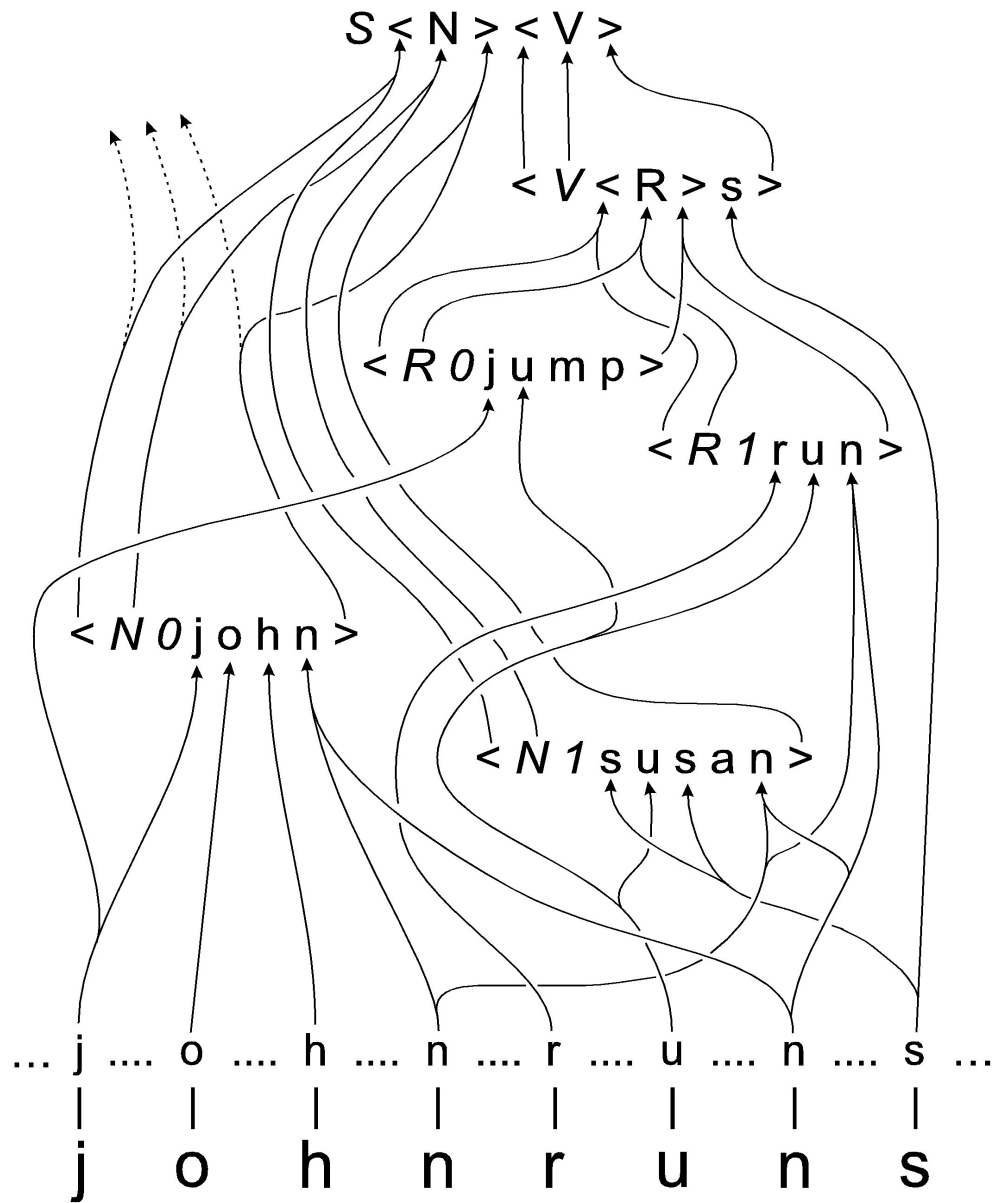
Figure 3: Schematic representation of the proposed neural basis for ICMAUS concepts. The conventions used and other aspects of this figure are explained in the text.

from neurons within the NA that have become active as a result of sensory input. There will be similar connections from each of the many neurons not shown within the sensory NA.

Each of the PAs are connected in a similar way to other PAs. Each of the ID-neurons within each PA is connected to zero or more C-neurons representing the same symbol in zero or more PAs. It is possible for an ID-neuron within one PA to be connected to a C-neuron within the same PA but this 'recursive' kind of connection is not shown in the figure.

The dotted arrows in the figure are included as an indication that any one PA ('< N 0 j o h n >' in this example) may be connected to two or more higher-level PAs. The dotted arrows might, for example, connect with a higher-level PA representing the interrogative form of a sentence.

Each connection comprises a single axon, or a chain of two or more neurons, with branches where necessary. Contrary to what the figure suggests, axons do not join together where they converge on one neuron. That kind of representation is a shorthand for two or more synapses onto the body or dendrites of the given neuron.

No attempt has been made to show any kind of internal NA in the figure. It is envisaged that structures of that kind would have connections to ID-neurons such as '0' in '< N 0 j o h n >' or '1' in '< R 1 r u n >'.

Not shown in the figure are lateral connections between fibres, proposed and discussed in Section 6, below.

5 Representation of knowledge in the brain

The aim in this section and the ones that follow is to flesh out the proposals outlined in the previous section, filling in detail, considering relevant evidence and discussing associated issues.

The central idea—that knowledge is stored in the brain as ICMAUS-style assemblies of neurons—is broadly similar to Hebb's proposals but there are important differences, as we shall see.

The cells in a PA need not be physically contiguous although it seems likely that, in most cases, they will be. Whatever the spatial arrangement of the cells, they should be topologically equivalent to an array of cells in one or two dimensions, or possibly more in some cases.

It is assumed in these proposals that PAs exist largely in the cortex. It seems likely that those concerned with patterns in a particular sensory modality will be in the corresponding area of the sensory cortex. But many others, especially those concerned with a combination of sensory modalities, are likely to be located in the 'association' areas of the cortex.

As in Hebb's proposals, it is envisaged that each cell in a PA is connected to other cells in the PA, at least its immediate neighbours.

As noted above, an ICMAUS symbol may translate into neural terms as a neuron or, if it represents a relatively large chunk of information, it may be realised as a PA. In the latter case, it needs to be connected to the PA representing the pattern in which the given symbol appears. This may be done in exactly the same way as a PA may be linked to another PA, as described below.

5.1 Hierarchical organisation of knowledge

Hebb [1949, p. 130] envisaged an hierarchical organisation for conceptual structures and the possibility of low-level concepts being shared by two or more higher-level concepts (see also the paragraph from Hebb [1958, p. 105] quoted in Section 8.2, below). Hebb's descriptions imply that any one neuron may belong in two or more cell assemblies and this is commonly assumed (see, for example, Huyck [2001, p. 385], Pulvermüller [1999, pp. 257–258] and Sakurai [1998, p. 213]).

This kind of organisation is problematic. If, for example, the cell assembly that recognises the letter 'a' is embedded in the cell assembly for each word that contains that letter, it is hard to see how the cell assembly for 'aardvark' would be organised. It is also hard to see how one could represent the abstract structure of a sentence, for example, where each 'slot' in the pattern may receive a range of *alternative* elements.

These problems may be overcome by the use of *references* between assemblies instead of the literal embedding of one assembly in another.

By contrast with the generally accepted view of cell assemblies, it is envisaged in the ICMAUS-N proposals that any one neuron may belong in one PA and *only* one PA. Hierarchical organisation and the sharing of structures is achieved by connections between ID-neurons and C-neurons that achieve the effect of pointers in software or references in books or articles. There is further discussion in Section 8.2, below.

This kind of reference mechanism means that a given PA can link widely-dispersed areas of the cortex without itself needing to be large (see Section 5.3.2, below).

5.2 New information in brains and nervous systems

As indicated above, sensory information is received by 'neural arrays' such as the visual cortex or areas of the cortex concerned with other sensory modalities. Comparable structures may also exist within the thalamus. It is envisaged that internally-generated patterns that play the role of New information will be stored in some kind of internal NA.

A sensory NA, unlike a PA, must be able to capture any of a wide variety of patterns that arrive at the senses. As a rough generalization, it seems that this is normally achieved by providing a range of different types of receptor cells, each type tuned to some relatively small feature that may be found in sensory data and the whole set repeated across the NA. In the visual cortex, for example, there are cells that respond selectively to short line segments at different angles and the complete range of orientations is repeated within each of a large number of fairly small 'orientation columns' [Barlow, 1982, Hubel, 2000]. It seems likely that a similar organisation may exist in the somatosensory cortex, mirroring the way in which receptors that respond selectively to heat, cold, touch, pressure and so on are repeated across areas of the skin [Nicholls et al., 2001, Chapter 18].

There are many complexities of course (e.g., 'complex' cells in vision that can respond to a line at a given orientation in any of several different positions [Hubel, 2000][6]). But in

---

[6]If complex cells could accommodate large variations of position, this would mean that useful information about the positions of visual entities would be lost at the level complex cells. It has been suggested [Nicholls et al., 2001, p. 422] that the main role of complex cells is to cope with the relatively small changes in position caused by the small saccadic eye movements that are continually made to prevent visual input being suppressed by adaptation in the retina. However, Milner (personal communication) suggests that the receptive fields of many complex cells are significantly bigger than would be needed for that function.

general the overall effect seems to be to translate raw analogue data into a digital language composed with an alphabet of sensory features.

So far, we have assumed that a New pattern arrives all at once like an image projected on to the retina. But it is clear that streams of sensory information are being received in all sensory modalities throughout our lives and that we are very sensitive to patterns in the temporal dimension within those streams of information, especially in hearing. It is assumed here that temporal sequencing in hearing, vision and other modalities is encoded spatially in arrays of cells, probably in the cortex.

5.3 Connections between neurons

As noted above, it is envisaged that, within each PA, neurons will be interconnected, at least to their nearest neighbours. If, as suggested above, the cells within any one PA lie close together, then lateral connections between the cells within any one PA will normally be quite short.

From each neuron with an NA, there may be zero or more connections, each one to a corresponding C-neuron within a PA. And from each ID-neuron within a PA there may be zero or more connections, each one to a corresponding C-neuron within another PA. In this context, the word 'corresponding' means that both neurons represent the same ICMAUS symbol. We may suppose that the connections from any one neuron (within an NA or a PA) are formed from an axon with as many branches as are required.

5.3.1 The volume of connections between neural patterns

There is a possible worry here that the number of connections might get out of hand and that the volume of axons would be too big to fit inside anyone's skull. This might be true if everything was connected to everything else but that is not necessary and seems an unlikely feature of the brain.

Consider, for example, how a neuron within an NA may connect to PAs representing sensory patterns. A neuron representing a short line segment at a particular angle could, conceivably, make a connection to each one of thousands of PAs representing objects or scenes that contain that small feature. It seems more likely, however, that a neuron within an NA would connect only to a relatively small number of PAs each one of which represents some relatively small 'motif'—an angle in some orientation or some other small detail. In a similar way, each of those motifs may connect to a relatively small number of higher-level patterns and so on through several different levels, much like the various levels of description that are familiar in a linguistic context (illustrated in Section 2.4).

By breaking patterns down into parts and sub-parts in this kind of way, with connections chiefly between one level and the next, the number of connections that are required can be dramatically reduced.

As with ICMAUS patterns (or, indeed, the rules of a typical grammar), PAs do not necessarily fall into a strict hierarchy. It would be more accurate to call it a 'loose heterarchy', where any one PA may receive connections from two or more lower-level PAs and may also connect to two or more higher level PAs. Connections need not always be to some strictly-defined 'next level' and, where there are recursive relations, ID-neurons within a given PA

may actually connect to C-neurons of the same PA (examples of this kind of recursion may be found in Wolff [1999a]).

Any untidiness of this kind makes no difference to the basic idea that any one neuron need only be connected to a small subset of all the PAs and thus the volume of interconnections may be kept within reasonable bounds. This economy in interconnections is, indirectly, a reflection of the economy that can be achieved with an ICMAUS style of knowledge representation.

5.3.2 The lengths of connections between neural patterns

How long would the connections between neural patterns have to be? That depends on what kinds of things are being connected.

PAs representing concepts within one domain—some aspect of vision, for example—may all reside within one relatively small area of the sensory cortex and their interconnections need not be very long. However, many of our concepts involve associations across two or more modalities and this seems to demand relatively long interconnections in most cases. Thus, for example, Pulvermüller [1999] has suggested that "most word representations consist of two parts, a perisylvian part related to the word form and a part located mainly outside the perisylvian areas representing semantic word properties." (p. 275).

As another example, consider the concept of 'our house'. This is likely to include the visual appearance of the house (patterns of luminance, colour and motion), sounds, smells and tactile aspects as well as emotional associations. Athough the corresponding PA may be quite compact and located in one small part of the cortex (one of the 'association' areas, perhaps), it must have connections to the areas of the brain concerned with each of the elements, sensory or otherwise, from which the concept is built. Given that each modality is typically processed in a localised area of the cortex and given that these areas are widely dispersed [Nicholls et al., 2001], it seems that relatively long connections will normally be needed.

This has some bearing on learning processes, as we shall see in Section 7.

5.4 Are there enough neurons in the brain?

The idea that the brain stores knowledge in the form of ICMAUS-style cell assemblies raises the question "Are there enough neurons in the brain to support this style of representation"? This is a difficult question to answer because of various uncertainties. Here are some very rough estimates, made with rather conservative assumptions:

- Estimates of the number of neurons in the human brain range from 10^{10} [Sholl, 1956] to 10^{11} [Williams and Herrup, 2001].

- If a neuron in the sensory cortex responds to some small feature of the world—a line at a certain angle or a note of a given pitch—and if there are, at a conservative estimate, about 30 different basic features across all sensory modalities, then each neuron represents $\lceil log_2 30 \rceil = 5$ bits of information. It seems reasonable to assume that other neurons in the brain have a similar capacity. Erring on the side of caution, we shall make the very conservative assumption here that each neuron represents only

1 bit of information. On that basis, the 'raw' storage capacity of the brain is between approximately 10^9 to 10^{10} bytes or between 1000 Mb and 10,000 Mb.

- Bearing in mind that stored knowledge will be heavily compressed using ICMAUS-style mechanisms, the effective storage capacity of the brain will be larger. With an ordinary Lempel-Ziv compression algorithm (e.g., PkZip or WinZip) it is possible to compress English text to one third of its original size (without loss of non-redundant information). This is probably a rather conservative estimate of the amount of compression that may be obtained across diverse kinds of knowledge:

  - LZ algorithms are designed for speed on a conventional serial computer rather than optimal compression. More compression can be achieved when more processing power is available (as in the human brain).
  - Visual information typically contains more redundancy than English text. For example, a good quality compressed JPEG version of a typical photograph is about 5 times smaller than an uncompressed bitmap version of the same image. It is true that JPEG is a lossy compression technique but if the quality is good, the amount of non-redundant information that is lost is relatively small.
  - 'Cinematic' information normally contains very high levels of redundancy because, in each sequence of frames from one cut to the next, each frame is normally very similar to its predecessor. This redundancy allows high levels of compression to be achieved.

  If we adopt the conservative assumption that compression by a factor of 3 may be achieved across all kinds of knowledge, our estimates of the storage capacity of the brain will range from about 3000 Mb up to about 30,000 Mb.

- Is this enough to accommodate what the average person knows? Any estimate here can only be very approximate. Let us assume that the average person knows only a relatively small proportion of what is contained in the Encyclopaedia Britannica. Clearly, each person knows lots of things that are *not* contained in that encyclopaedia—how to walk or ride a bicycle, information about friends, relatives and acquaintances, how the local football team is doing, and so on. If we assume that the 'personal' things that we do know are very roughly equal to the things in the encyclopaedia that we do not know, then the size of the encyclopaedia provides an estimate of the volume of information that the average person knows.

- The Encyclopaedia Britannica can be stored on two CDs in compressed form. Assuming that most of the space is filled, this equates to 1300 Mb of compressed information or approximately 4000 Mb of information in uncompressed form. This estimate of what the average person knows is the same order of magnitude as our range of estimates of what the human brain can store.

Even if the brain stores two or three copies of its compressed knowledge (to guard against the risk of losing it, as discussed in Section 3.3), our estimate of what needs to be stored (12,000 Mb, say) lies within the range of our estimates of what the brain can store. Bearing in mind the *very* conservative assumption that was made about how much information each

neuron represents and other conservative assumptions that have been made, it seems that the proposals in this article about how the brain stores information are not wildly inconsistent with what is required.

5.5 Comparisons

In terms of the 'static' forms of knowledge in the brain (ignoring learning), the ICMAUS-N proposals extend or modify Hebb's original concept in the following ways:

- There is a distinction between NAs that, *inter alia*, serve as temporary buffers for information (from the senses or elsewhere) and PAs that provide for longer-term storage of information.

- Within each PA, it is envisaged that some cells (C-neurons) represent the 'contents' or 'substance' of the pattern while others (ID-neurons) will serve to identify the PA or otherwise define how it relates to other PAs.

- Connections between ID-neurons and C-neurons provide a mechanism by which one PA may be 'referenced' from another PA. This is the key to the expressive power of the ICMAUS framework.

- As noted above, any one neuron may not belong in more than one PA.

- Information is transmitted through the network of PAs by an interplay of excitatory and inhibitory processes as described in Section 6, next.

6 Neural processes for building multiple alignments

Given a structure of PAs, NAs and interconnections as described above, how may the system work to build structures that are comparable with the kinds of multiple alignments created by the SP61 and SP70 computer models? The focus here is mainly on the way in which sensory data may connect with stored patterns, amplifying what was sketched in Section 4, discussing relevant issues and filling in details.

In terms of the schematic representation shown in Figure 3, we are interested in the kind of neural processing that would yield something equivalent to the alignment shown in Figure 4. This is the best alignment found by SP61 with patterns that are comparable with the pattern assemblies shown in Figure 3.

Given a scarcity of direct evidence about what is going on in this area of brain function, it has been assumed that the kinds of principles that have been seen to operate elsewhere in nervous tissue will also apply here. In particular:

1. Contrary to what one might assume, neurons do not normally rest quietly in the absence of stimulation. There is normally a background rate of firing and this may be increased by excitatory input or decreased by inhibitory input [see, for example, Ratliff and Hartline, 1959, Ratliff et al., 1963].

```
0           j o h n              r u n   s   0
            | | | |              | | |   |
1   < N 0 j o h n >              | | |   |   1
    | |             |            | | |   |
2 S < N           > < V          | | |   | > 2
                    | |          | | |   | |
3                   | | < R 1 r u n > | | 3
                    | | | |          | | |
4                   < V < R          > s > 4
```

Figure 4: The best alignment found by SP61 with patterns that are comparable to the pattern assemblies shown Figure 3.

2. Inhibition seems to be a prominent part of the mechanism by which nervous systems remove redundancy from information. For example, lateral inhibition in the retina operates when neighbouring parts of a visual image are the same—which means that there is redundancy in that part of the image. In a similar vein, Barlow [1997] describes a 'law of repulsion': "when two stimuli frequently occur together their representations discourage or inhibit each other so that each is weaker than it is when each stimulus is presented alone." (p. 1144). Since frequent recurrence of two stimuli together is a manifestation of redundancy, this law seems to be in keeping with the idea that a major role of inhibition in neural tissue is to reduce redundancy in information.

3. Lateral inhibition may, in some situations, be seen as a way of forcing a decision amongst competing alternatives so that the 'winner takes all'. In the innervation of the Basilar membrane, where position along the length of the membrane encodes the frequency of sound, lateral inhibition has the effect of strengthening the response where the signal is strongest and suppressing the response in neighbouring regions [von Békésy, 1967].

   Mutual inhibition amongst competitors may also be seen as a means of reducing redundancy because some or all of the information expressed by each element in the competition is also expressed by other elements. Neighbouring regions of the Basilar membrane respond to sounds with similar frequencies—which means that there is information in common amongst those sounds.

4. Where there is a mapping from one part of the nervous system to another (e.g., between the retina and each of the areas of the visual cortex), each point in the origin normally projects to an area in the destination and there is overlap between the projection areas of neighbouring points [Nicholls et al., 2001].

What follows is an outline of two possible ways in which the system may function, in keeping with the principles just described.

6.1 Outline of processing

When sensory input is received, this causes a subset of the neurons within each sensory NA to increase their rate of firing. These signals are transmitted to corresponding C-neurons within a relatively small number of 'low level' PAs like '< N 0 j o h n >' and '< R 1 r u n >' in the schematic representation in Figure 3. At this stage, there seem to be two main possibilities, described in the next two subsections.

6.1.1 'Excitatory' scheme

In the first scheme, which is similar to Hebb's original proposals, signals received by the C-neurons of a PA would have the effect of increasing the rate of firing of those neurons. This excitation would spread via the lateral connections to all the other neurons in the PA, including the ID neurons and any C-neurons that had not received any stimulation on this occasion.

When the rate of firing of ID-neurons is increased as a result of excitatory input to the C-neurons of a PA, these signals will be transmitted to C-neurons in other PAs. As before, the effect will be to increase the rate of firing of those neurons. And, as before, when C-neurons within a PA have become excited, this will have the effect of increasing the rate of firing of ID-neurons within the PA and any C-neurons that did not receive any input.

This process may be repeated so long as there are PAs that can receive signals from other PAs in which the ID-neurons are active.

Although the emphasis here is on the excitation of neurons, inhibition may have a rôle to play in preventing excitation getting out of control [Milner, 1957]. There might, for example, be inhibitory feedback between each C-neuron within a PA and the fibres that bring excitatory signals to that neuron. This would have the effect of dampening down the excitatory signals once they had been received.

This kind of inhibitory feedback may be seen to have the effect of reducing or eliminating redundancy in the temporal dimension corresponding to the persistence of a given sensory input over a period of time.

6.1.2 'Inhibitory' scheme

Notwithstanding some evidence suggesting that the majority of nerve cells in the cortex are excitatory [Pulvermüller, 1999, p. 264] an alternative 'inhibitory' scheme is sketched here that is broadly consistent with the principles described above.

The first of the three principles suggests that, in the absence of stimulation, the neurons in each PA maintain a steady rate of firing.

Since each PA represents a repeating (redundant) pattern in the environment, it is further assumed—in accordance with the second principle—that there is mutual inhibition amongst the neurons of each PA, mediated by the lateral connections between the neurons. There will be a balance between the tendency of the neurons to fire spontaneously and the inhibitory effect that they have on each other.

We may further suppose that, when a neuron within a PA receives stimulation, the effect is inhibitory. Again, this seems to accord with the second principle because the incoming stimulation and the neuron receiving the stimulation represent the same small pattern or symbol, and this repetition of information means redundancy.

27

If the rate of firing of C-neurons within a PA is inhibited, this will have the effect of reducing the inhibitory signals which those neurons have been sending to their neighbours. Hence, the effect of inhibitory input to the C-neurons of a PA will be to *increase* the rate of firing of other neurons in the PA that are not receiving input. The neurons whose rate of firing will increase will be all of the ID-neurons (because they do not receive any input) and any C-neurons that may not have received any input on this occasion. The former will be the focus of attention here. The latter may be seen as inferences that can be derived from the alignment, as described in Section 2.5.4, above.

When the rate of firing of ID-neurons is increased as a result of inhibitory input to the C-neurons of a PA, these signals will be transmitted to C-neurons in other PAs. As before, the effect will be to inhibit the activity of those neurons. And, as before, when C-neurons within a PA have been inhibited, this will have the effect of increasing the rate of firing of ID-neurons within the PA and any C-neurons that did not receive any input.

This process may be repeated so long as there are PAs that can receive signals from other PAs in which the ID-neurons are active.

6.2   Discussion and elaboration

Neither of the schemes just described is sufficient in itself to achieve an effect comparable with the building of multiple alignments in the ICMAUS framework. This subsection considers some other aspects of the proposals and fills in some details that seem to be needed for the system to work as intended.

6.2.1   Inputs to pattern assemblies

In both of the schemes that have been outlined, it is natural to assume that the effect of inputs to each PA would be related to the amount or number of inputs received. If the PA '< N 0 j o h n >' in Figure 3 were to receive inputs from only 'j' and 'h', one would expect the effect to be less than when the PA receives signals from all of 'j', 'o', 'h' and 'n'. Correspondingly, one would expect the increase in firing rate of the ID-neurons to be larger in the second case than the former.

It seems reasonable to assume that the strength of the responses would depend on the proportion of C-neurons in any PA that receive stimulation rather than their absolute numbers.

6.2.2   Divergence

What happens when the signals from a given neuron or combination of neurons in an NA or a PA are transmitted to corresponding C-neurons in two or more other PAs? In the absence of any disambiguating context (see Section 6.2.3, next), all of those PAs should respond, with the strength of the response related to the proportion of C-neurons in each PA that have received an input.

This kind of response would naturally accommodate the kinds of ambiguity mentioned in Sections 2.4.5 and 2.5.2.

6.2.3 Convergence

What happens where two or more sets of fibres converge on one set of C-neurons in one PA? An example can be seen in Figure 3 where fibres from '<, 'N' and '>' in the PA '< N 0 j o h n >' are connected to '< N >' in the PA 'S < N > < V >', and similar connections are made to the same three neurons from the PA '< N 1 s u s a n >'.

In this example, the PAs '< N 0 j o h n >' and '< N 1 s u s a n >' are, in effect, competing for the rôle of noun within the PA 'S < N > < V >' representing the structure of a simple sentence. In the light of the third principle mentioned above, we may suppose that a decision about which of these two PAs should 'win' the competition may be decided by inhibition.

Where fibres representing the same symbol converge on a single neuron, we may suppose that there are lateral connections between them that carry inhibitory signals when the main fibres are excited. This should ensure that the strongest signal gets through and weaker ones are suppressed.

At the highest levels in the hierarchy of PAs, we may suppose that there is a similar competition for the role of 'best overall interpretation' of the sensory input. Where there is ambiguity of the kind mentioned earlier, a disambiguating context may tilt the balance in favour of one interpretation or the other.

Possible evidence in support of the proposed lateral connections between fibres carrying rival signals is the discovery of lateral connections in the visual cortex that connect columns in which the cells respond to edges or lines with similar orientations [see Nicholls et al., 2001, p. 440]. It is not clear whether or not these connections are inhibitory.

Lateral connections are also broadly consistent with the fourth principle described above (projection from neighbouring points to overlapping areas).

6.2.4 Keeping track of order in perception

As the system has been described so far, a stimulus like 'b o y' would produce the same level of response in a PA like '< N 2 y o b >' as it would in a PA like '< N 3 b o y >'. In other words, there is nothing in what has been proposed to keep track of the ordering or arrangement of elements in a pattern.

Our ability to recognise patterns despite some jumbling of their elements (e.g., solving anagrams) is consistent with the scheme that has been described. But the fact that we can see the difference between an anagram and its solution shows that something else is needed. Some kind of mechanism is required to distinguish between two kinds of situation:

1. Fibres leave a PA or one portion of an NA and arrive together, in the same relative positions, at one portion of another PA. Fibres that conform to this rule will be described as *coherent*.

2. Fibres arrive at one portion of a PA from a variety of different sources or, if they come from one source, their relative positions are not preserved. This kind of arrangement may be described as *incoherent*.

When signals arrive at a PA, they should produce a stronger response in the first case than in the second. A possible mechanism to ensure that this happens would be lateral

connections amongst the fibres of a coherent bundle that would have the effect of increasing the rate of firing in the bundle if the majority of them are firing together. When any one fibre in the bundle is firing, then signals will be carried to neighbouring fibres via the lateral connections and these signals should lower the threshold for firing and thus increase the rate of firing in any of the neighbouring fibres that are already firing. These kinds of lateral connections would be broadly consistent with the fourth principle described above.

If signals arrive in a temporal sequence, then other possible mechanisms include 'synfire chains' and temporal 'sequence detectors', as described by Pulvermüller [2002]. However, these would not do for keeping track of order in spatially-distributed patterns, seen in a single glance.

6.2.5   Keeping track of order in production

In speaking or writing or comparable activities not involving language, we need to do things in the right order. The spatial arrangement of neurons in a PA can provide information about that order but supplementary mechanisms would be needed to control the timings of actions and the use of sensory feedback to monitor progress.

6.3   Constancies

A possible objection to the ICMAUS-N proposals is that they are inconsistent with the 'constancy' phenomena in perception. These include:

- *Size constancy.* We can recognise an object despite wide variations in the size of its image on the retina—and we judge its size to be constant despite these variations.

- *Brightness constancy.* We can recognise something despite wide variations in the absolute brightness of the image on our retina (and, likewise, we judge its intrinsic brightness to be constant).

- *Colour constancy.* In recognising the intrinsic colour of an object, we can make allowances for wide variations in the colour of the light which falls on the object and the consequent effect on the colour of the light that leaves the object and enters our eyes.

If the pattern recorded in an NA was merely a copy of sensory input there would indeed be wide variations in the size of the visual pattern projected by a given object and in brightness and colours. The mechanism described in the previous subsection for detecting coherence would not work for patterns whose elements were widely dispersed across the NA. And recognising the elements of a pattern would be more complicated if there were wide variations in brightness or colour.

The suggestion here is that much of the variability of sensory input from a given object has been eliminated at a stage before the information leaves the NA. As was noted in Section 3.4, lateral inhibition in the retina emphasises boundaries between relatively uniform areas. The redundant information within each uniform area is largely eliminated which means that it is, in effect, shrunk to the minimum size needed to record the attributes of that area. Since this minimum will be the same regardless of the size of the original image, the overall effect should be to reduce or eliminate variations in the sizes of images from a given object.

In a similar way, the 'primal sketch' created by lateral inhibition should be largely independent of the absolute brightness of the original image. And adaptation in the early stages of visual processing should to a large extent prevent variations in brightness having an effect on patterns reaching the NA.

Edwin Land's 'retinex' theory suggests that colour constancy is achieved by processing in the retina and in the visual cortex [see Nicholls et al., 2001, pp. 437–439]. This is consistent with the idea that this source of variability has been removed at a stage before sensory input is compared with stored patterns.

7 Neural mechanisms for learning

Like SP61, SP70 tries to encode New information as economically as possible by encoding it in terms of Old information. If a given pattern from New cannot be matched with anything in Old, then the program simply adds the pattern to Old together with system-generated ID symbols. If a given pattern from New can be matched only partially with a pattern in Old, then the program derives new patterns from the partial match and adds them to Old.

From time to time, it is necessary to make a selection from amongst the patterns that have been created, retaining those that are 'good' in terms of MLE principles and discarding the rest. In the SP70 model, no attempt has been made to take account of motivations or emotions but one can envisage a generalization of the model in which these factors have an influence on which patterns are retained and which discarded.

To get the flavour of how SP70 works, consider a simple example. If the current pattern in New is 't h e b o y r u n s' and Old is empty, the program discovers that there is no way to encode the New pattern economically in terms of Old information so it augments the pattern with ID symbols (which converts it into '< %1 3 t h e b o y r u n s >') and adds the augmented pattern to Old. When Old has accumulated a range of patterns like this, it can begin to create multiple alignments. If the current pattern in New is 't h e g i r l r u n s', multiple alignments created by the system will include one like this:

```
0       t h e g i r l r u n s   0
        | | |         | | | |
1 < %1 3 t h e b o y   r u n s > 1
```

From this alignment, the system creates new patterns like this:

```
< %3 8 t h e >
< %4 11 r u n s >
< %5 14 b o y >
< %5 15 g i r l >
< %6 20 < %3 > < %5 > < %4 > >
```

The first four of these patterns represent words while the last one describes the overall structure of two original sentences. Notice how 'b o y' and 'g i r l' have been assigned to the same class, identified by the symbol '%5'.

Notice also how the system picks out entities that we naturally regard as 'discrete'. Similar principles may explain how we learn to see the visual world in terms of discrete objects.

The dynamic programming built into the process of finding multiple alignments means that SP70 can find correlations that bridge arbitrary amounts of intervening structure. It is not necessary to restrict the learning process to second- or third-order correlations (*cf.* Thompson [2000]).

The SP70 model is described much more fully in Wolff [2002b]. Possible ways in which this kind of learning may be done in neural tissue are considered in the next two subsections.

7.1 Creation and storage of patterns

Patterns like those just described would, of course, be realised as PAs. The SP70 style of learning implies that PAs may be created in response to patterns of incoming information that have occurred only once. This contrasts with Hebb's proposal that "Any frequently repeated, particular stimulation will lead to the slow development of a 'cell-assembly'..." [1949, p. xix]. This difference between the two views of learning is discussed briefly in Section 7.4, below.

Another implication of the SP70 style of learning is that there is some kind of mechanism that can add ID-neurons to each newly-created PA. We need also to suppose that the system can create connections between PAs of the kind described earlier.

It seems unlikely that the neurons of each newly-created PA would be grown according to need. Although there is now evidence that new nerve cells can grow in mature brains [Shors and Gould, 2001], it seems likely that the bulk of neurons in the brain are present at birth and that the creation of PAs is largely a process of adapting pre-existing neurons. We may also suppose that the postulated lateral connections between neurons within each PA are also largely present at birth. It is possible that the boundaries of PAs may be established by breaking pre-existing lateral connections. Something like this is suggested by evidence for the progressive weakening of connections between cells that do not normally fire at the same time [Pulvermüller, 1999, pp. 254–255].

In a similar vein, it is not necessary to suppose that connections between PAs are grown from scratch according to need. In many cases, this would be a cumbersome process given that many of the connections between different areas of the cortex are fairly long. It seems much more likely that, as Hebb suggested, nerve fibres between different parts of the cortex are present at birth and that these may be converted into connections between PAs, when required, by the growth of relatively short links.

It may not even be necessary to postulate the growth of short links. It is possible that pre-existing links could be switched on or off somehow according to need. Something like this might enable us to account for the speed with which we can establish memories for names, faces, events and so on (see Section 7.4, below).

7.2 Finding good partial matches between patterns

A key part of the learning process in SP70 is the formation of partial matches between patterns like the one shown above. The kind of neural processing described in Section 6 should be able to achieve this kind of partial matching automatically, without the need for any supplementary mechanisms.

In the 'inhibitory' scheme, lateral inhibition should have the effect of marking the boundaries between parts of a PA that are receiving signals and parts that are not. And this

marking of boundaries would presumably be helpful in the process of creating new PAs, derived from matched and unmatched parts of the given PA.

7.3 Counting

In the SP70 model, each pattern has an associated frequency of occurrence. These frequency values are used periodically to adjust the numbers and sizes of ID-symbols assigned to each pattern, in accordance with the Shannon-Fano-Elias (SFE) coding scheme [see Cover and Thomas, 1991].[7]

It seems possible that information about the frequency of occurrence of each PA may be recorded by the 'strength' of the intra-assembly synapses (as in Hebbian cell assemblies) or, perhaps, by the predisposition of cells to fire in response to incoming stimulation or by the strength of their response when they do fire. It seems possible that this kind of information may lead to adjustments in the numbers of ID-neurons assigned to each PA.

7.4 Speed of learning

In Hebb's theory of learning, cell assemblies are built up gradually by the progressive strengthening of connections between cells that regularly fire at the same time. Although this seems to accord with the slow build up of knowledge throughout childhood and beyond, it does not sit well with the undoubted fact that we can remember things that we have experienced only once—what we had for lunch today, our first day at school, and so on. It is true that unique events are probably encoded in terms of things we already know but we need to be able to store that encoded representation without the need for repetition.

Because the slow growth of cell assemblies does not account for our ability to remember things immediately after a single exposure, Hebb adopted a 'reverberatory' theory for this kind of memory. But, as Milner [1996] points out, it is difficult to understand how this kind of mechanism could explain our ability to assimilate a previously-unseen telephone number. Each digit in the number may be stored in a reverberatory assembly but the way in which we remember the *sequence* of the digits is unexplained.

In the learning scheme outlined above, new PAs can be created in response to a single sensory input. This is consistent with our ability to remember unique events and sequences of events. And, if we suppose that synapses can be switched on or off according to need, we may be able to account for the speed with which immediate memories can be established. But an ability to lay down new PAs relatively quickly does not explain why it takes several years to learn something like a language.

The principles on which SP70 is based suggest why learning a language, and similar kinds of learning, take quite a lot of time. The abstract 'space' of alternative grammars for any natural language is astronomically large and it takes time to search. Although exhaustive search is not feasible, the kinds of heuristic techniques used in the SP70 model—trading accuracy for speed—can bring search times within the bounds of practicality. But finding a tolerably good grammar is still a complex problem and it cannot be solved in an instant.

---

[7]The SFE scheme is similar to the Huffman scheme but has certain advantages in the calculation of probabilities of inferences that may be drawn from alignments.

7.5 Empirical validation

Athough SP70 is not well enough developed to be compared in detail with the way children learn to talk, it is anticipated that, in the sphere of learning, it will behave much like the earlier SNPR model [Wolff, 1988, 1982]. This is because both models are founded on MLE principles and it seems that these principles can account for many features of the way children learn language:

- Unsupervised learning of conjunctive segmental structures (words, phrases etc) from unsegmented linguistic input [see also Wolff, 1980, 1977, 1975].

- Unsupervised learning of disjunctive categories (nouns, verbs etc) from unsegmented linguistic input.

- Unsupervised learning of hierarchical grammars with conjunctive and disjunctive structures at any level.

- Generalization of grammatical rules and unsupervised correction of overgeneralizations.

- The pattern of changes in the rate of acquisition of words as learning proceeds. The slowing of language development in later years.

- The order of acquisition of words and morphemes.

- Brown's (1973) Law of Cumulative Complexity.

- The S-P/episodic-semantic shift.

- The learning of 'correct' forms despite the existence of errors in linguistic data.

- The learning of non-linguistic cognitive structures.

- The word frequency effect.

These things are discussed quite fully in Wolff [1988].

8 Discussion

This section considers a selection of other issues, including some difficulties with Hebb's original proposals [described by Milner, 1996] and how, in the ICMAUS-N proposals, they may be overcome.

8.1 Stimulus equivalence and discrimination

An important motivation for the cell assembly concept was to explain the phenomenon of 'stimulus equivalence' or 'generalization'—our ability to recognise things despite variations in size, shape, position of the image on the retina, and so on. However, Milner writes that "Generalization is certainly an important and puzzling phenomenon, but discrimination is equally important and tends to be ignored in theories of neural representation. Any theory of

engram formation must take into account the relationship between category and instance—the ability we have to distinguish our own hat, house, and dog from hats, houses and dogs in general." (*ibid.*, p. 70).

The ICMAUS-N proposals account for generalization in four main ways:

- As with the original cell assembly concept, a PA can respond provided that a reasonably large subset of its neurons has received inputs. This seems to accommodate our ability to recognise things despite omissions, additions or substitutions in the sensory input relative to the information stored in the PA. The effect should be similar to the dynamic programming built into the SP models.

- In the present scheme, each neuron or feature detector in an NA is connected to each one of the lowest level PAs that contain the corresponding feature. Thus, each of these PAs may respond to appropriate input regardless of the position of the input on the NA.

- The way in which perceptual constancies may be accommodated in the present proposals was considered in Section 6.3.

- Owing to the provision of a 'reference' mechanism in the present proposals (discussed in Section 5.1, above), it is possible to create PAs that represent abstractions from sensory input. An example from Figure 3 is the PA 'S < N > < V >'. This PA represents the abstract structure of a simple sentence and it may be activated by a range of *alternative* inputs: any sentence that conforms to the noun-verb pattern.

The ICMAUS framework provides a neat account of our ability to distinguish specific instances from the categories in which they belong. Given a pattern to be recognised (in New) and appropriate patterns in Old, the SP61 model can build multiple alignments which include several different levels of abstraction, including a level that represents a specific instance (Section 2.5.3).

If the ICMAUS concepts can be realised with neural mechanisms, we can account in neural terms for recognition at multiple levels of abstraction, including the level corresponding to a specific instance.

8.2   Parts, wholes and associations

With respect to Hebbian concepts, Milner raises some other questions: "How do associations between cell assemblies differ from internal associations that are responsible for the assemblies properties? It does not seem likely that both these processes can be the result of similar synaptic changes as is usually assumed. If they were, the interassembly associations would soon become intraassembly loops. A related puzzle is that parts are not submerged in the whole. Doors and windows are integral parts of my concept of a house, but they are also robust, stand-alone concepts." (*ibid.*, p. 71). Later on the same page he writes: "Perhaps the toughest problem of all concerns the fact that we have many associations with almost every engram....The brain must be a veritable rat's nest of tangled associations, yet for the most part we navigate through it with ease."

Hebb [1958, p. 105] provides a possible answer to the way in which parts may be distinguished from wholes: "If two assemblies A and B are repeatedly active at the same

time they will tend to become 'associated,' so that A excites B and vice versa. If they are always active at the same time they will tend to merge in a single systems—that is, form a single assembly—but if they are also active at different times they will remain separate (but associated) systems. (This means that exciting part of A, for example, has a very high probability of exciting all of A, but a definitely lower probability of exciting a separate assembly, B; A may be able to excite B only when some other assembly, C, also facilitates activity in B)."

The reference mechanism in the ICMAUS-N proposals provides a neater solution to this problem and the others raised by Milner:

- Associations between PAs may be encoded by building a new PA containing references to the PAs that are to be associated. These associations are then internal to the new PA and the original PAs retain their identity.

- In the same way, there can be a stand-alone PA for each type of component of a house while the PA for a house comprises a collection of *references* to the components. In this way, the concept of a house does not become muddled with concepts for doors, windows etc.

- It is true that most concepts have many associations but the reference mechanism allows these to be encoded in a way that does not disturb the integrity of each concept. The coherence of a web page is not affected by links to that page from other pages and it makes no difference how many such pointers there may be.

8.3   Direct and distributed neural representations

In the terms described by Gardner-Medwin and Barlow [2001], the ICMAUS-N scheme for representing knowledge may be classified as *direct* because each neuron serves one representation and only one representation. In this respect it contrasts with other *distributed* schemes in which any one cell may be involved in two or more different representations.

Arguments that have been made in favour of distributed representations include:

1. Large numbers of stimulus events may be distinguished with relatively few cells.

2. Provided the representations are 'sparse' (only a small proportion of cells are involved in each recognition event), they can function as content-addressable memories that store and retrieve amounts of information approaching the maximum permitted by their numbers of modifiable elements.

3. Given that counting is needed for learning (see below), distributed representations have the flexibility to represent and count recognition events that have not been foreseen.

Arguments in favour of direct representations include:

1. Starting from the reasonable premise that learning requires an ability to count the frequencies of perceptual entities, Gardner-Medwin and Barlow argue that "compact distributed representations (i.e., ones with little redundancy) enormously reduce the efficiency of counting and must therefore slow reliable learning, but that this is not

the case if they are redundant, having many more cells than are required simply for representation." (*ibid.*, p. 480). This need for redundancy offsets the first of the advantages listed for distributed representations. Gardner-Medwin and Barlow also acknowledge that this kind of redundancy is not required to achieve counting with direct representations.

2. As discussed in Sections 8.1 and 8.2, there are problems associated with distributed representations but these can be overcome in the ICMAUS-N version of direct representation.

3. In the ICMAUS-N scheme, compression mechanisms allow knowledge to be represented in a direct manner with a minimum of redundancy.

4. In the ICMAUS-N scheme, direct representations can function as content-addressable memory.

5. As discussed in Section 7.1, it seems possible that PAs in the ICMAUS-N proposals may be created relatively quickly in response to unforeseen events. And once created, they may count recognition events as described in Section 7.3.

8.4   Compression revisited

With respect to the rôle of IC in neural functioning, Barlow [2001a] writes: "There has been one major change in my viewpoint. Initially I thought that economy was the main benefit to be derived from exploiting redundancy, as it is for AT & T and British Telecom. But ... the physiological and anatomical facts do not fit the idea that the brain uses compressed, economical, representations, and one can see that these would be highly inconvenient for many of the tasks that it performs, such as detecting associations. Therefore I now think the principle is redundancy *exploitation*, rather than *reduction*, since performance can be improved by taking account of sensory redundancy in other ways than by coding the information onto channels of reduced redundancy." (p. 604).

Elsewhere [2001b], Barlow suggests that interest in the statistics of sensory messages "was right in drawing attention to the importance of redundancy in sensory messages because this can often lead to crucially important knowledge of the environment, but it was wrong in emphasising the main technical use for redundancy, which is compressive coding. The idea points to the enormous importance of estimating probabilities for almost everything the brain does, from determining what is redundant to fuelling Baysian calculations of near-optimal courses of action in a complicated world." (p. 242).

With respect to compressive coding (but not the importance of probabilistic reasoning), this revised view is somewhat at odds with what has been proposed in this article (apart from the suggested need for multiple copies of stored information (Section 3.3)). There is insufficient space here to discuss these issues fully but a few comments may be made:

- Barlow [2001a] writes that there is "more than a thousand times as many neurons concerned with vision in the human cortex as there are ganglion cells in the two retinas ... on any plausible assumptions the capacity of the cortical representation is vastly greater than that of the retinal ganglion cells, so redundancy appears to be

increased, not reduced." (p. 605). In the ICMAUS-N proposals, the cortex is the repository of most of our stored knowledge. Even allowing for the compression that can be achieved using reference mechanisms, we need all that space to store even a small fraction of the huge amount of information that impinges on our senses (Section 5.4, above).

- As we saw in Section 2.1.1, counting, in itself, implies IC by the unification of matching patterns.

- As we saw in Section 8.3, the need for efficient counting seems to demand the provision of redundancy in distributed representations. This may be seen as an argument for redundancy in stored information but equally it may be seen as an argument in favour of ICMAUS-style direct representation where no such redundancy is required.

- Contrary to the suggestion that "compressed, economical, representations ... would be highly inconvenient for ... detecting associations", such representations in the ICMAUS-N proposals provide a foundation for the detection of associations and they are indeed an expression of those associations.

- As was argued in Section 3, we would expect organisms that have evolved by natural selection to exploit IC as a means of economising on the transmission or storage of information.

- There is a lot of other evidence for the importance of redundancy reduction in brain function, some of which was briefly reviewed in Section 3.4.

8.5   Motivation, Intention and Attention

The ICMAUS theory and the ICMAUS-N proposals described in this article have little to say about such things as motivation, intention or selective attention. These things are clearly important in perception and learning [Milner, 1999] but, so far, they have remained largely outside the scope of the ICMAUS theory.

Each of the SP models processes New information in sections, one pattern at a time. SP61 is also able to divide a New pattern into 'windows' and process successive windows one at a time. Each of these things may be seen as a very crude version of selective attention. Their main function in the models is to provide a means of reducing the abstract 'space' of alternative solutions that needs to be searched at any one time. This may indeed be one of the functions of selective attention in humans and other animals but more work is clearly needed to understand how concepts of motivation, intention and attention might relate to current or future versions of the ICMAUS theory.

9   Conclusion

The ICMAUS framework provides an account of a range of concepts and phenomena in perception and cognition: the representation and integration of diverse kinds of knowledge, analysis and production of natural language, unsupervised inductive learning, fuzzy pattern recognition, recognition through multiple levels of abstraction, probabilistic 'deduction' and

abduction, chains of reasoning, nonmonotonic reasoning, 'explaining away', solving geometric analogy problems, and others.

This article has attempted to show how the abstract concepts of the ICMAUS scheme may be realised with neural structures and processes. The potential payoff from this exercise is that the substantial explanatory power of the ICMAUS framework may be imported into neuroscience.

What has been proposed may be seen as an extension and development of Hebb's original ideas. By incorporating ideas from the ICMAUS framework, especially the mechanism which allows one pattern to 'refer' to another, many of the weaknesses of the original cell assembly concept may be overcome.

These proposals offer an heuristic framework and a source of hypotheses for further exploration and development.

Acknowledgements

References

J. J. Atick. Could information theory provide an ecological theory of sensory processing? *Network*, 3:213–251, 1992.

F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61: 183–193, 1954.

R. Baddeley, P. Hancock, and P. Földiák, editors. *Information Theory and the Brain.* Cambridge University Press, Cambridge, 2000.

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.

H. B. Barlow. The coding of sensory messages. In W. H. Thorpe and O. L. Zangwill, editors, *Current Problems in Animal Behaviour*, pages 331–360. Cambridge University Press, Cambridge, 1961. Chapter XIII.

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.

H. B. Barlow. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1:371–394, 1972.

H. B. Barlow. David Hubel and Torsten Wiesel: their contribution towards understanding the primary visual cortex. *Trends in Neuroscience*, 5(5):145–152, 1982.

H. B. Barlow. Intelligence, guesswork, language. *Nature*, 304:207–209, 1983.

H. B. Barlow. The knowledge used in vision and where it comes from. *Philosophical Transactions of the Royal Society London B*, 352:1141–1147, 1997.

H. B. Barlow. The exploitation of regularities in the environment by the brain. *Behavioural and Brain Sciences*, 24(4):602–607 and 748–749, 2001a.

H. B. Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–253, 2001b.

N. Chater. Reconciling simplicity and likelihood principles in perceptual organisation. *Psychological Review*, 103(3):566–581, 1996.

N. Chater. The search for simplicity: a fundamental cognitive principle? *Quarterly Journal of Experimental Psychology*, 52 A(2):273–302, 1999.

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

A. R. Gardner-Medwin and H. B. Barlow. The limits of counting accuracy in distributed neural representations. *Neural Computation*, 13:477–504, 2001.

W. R. Garner, editor. *The Processing of Information and Structure*. Lawrence Erlbaum, Hillsdale, NJ, 1974.

M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

D. O. Hebb. *The Organization of Behaviour*. John Wiley & Sons, New York, 1949.

D. O. Hebb. *A Textbook of Psychology*. W. B. Saunders Company, Philadelphia, 1958.

D. H. Hubel. Exploration of the primary visual cortex, 1955–78. In M. S. Gazzaniga, editor, *Cognitive Neuroscience: a Reader*, pages 58–80. Blackwell, Oxford, 2000.

C. R. Huyck. Cell assemblies as an intermediate level model of cognition. In S. Wermter, J. Austin, and D. Willshaw, editors, *Emergent Neural Computational Architectures Based on Neuroscience*, pages 383–397. Springer, Heidelberg, 2001.

M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, New York, 1997.

D. Marr. *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, San Francisco, 1982.

P. M. Milner. The cell assembly: mark ii. *Psychological Review*, 64(4):242–252, 1957.

P. M. Milner. Neural representations: some old problems revisited. *Journal of Cognitive Neuroscience*, 8(1):69–77, 1996.

P. M. Milner. *The Autonomous Brain*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1999.

J. G. Nicholls, A. R. Martin, B. G. Wallace, and P. A. Fuchs. *From Neuron to Brain*. Sinauer Associates, Sunderland, Mass., 2001.

R. C. Oldfield. Memory mechanisms and the theory of schemata. *British Journal of Psychology*, 45:14–23, 1954.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, 1988.

F. Pulvermüller. Words in the brain's language. *Behavioural and Brain Sciences*, 22(2): 253–336, 1999.

F. Pulvermüller. A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Progress in Neurobiology*, 67:85–111, 2002.

F. Ratliff and H. K. Hartline. The response of limulus optic nerve fibres to patterns of illumination on the receptor mosaic. *Journal of General Physiology*, 42:1241–1255, 1959.

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.

J. Rissanen. Modelling by the shortest data description. *Automatica-J, IFAC*, 14:465–471, 1978.

B. Sakitt and H. B. Barlow. A model for the economical encoding of the visual image in cerebral cortex. *Biological Cybernetics*, 43:97–108, 1982.

Y. Sakurai. Cell-assembly coding in several memory processes. *Neurobiology of Learning and Memory*, 70:212–225, 1998.

D. Sankoff and J. B. Kruskall. *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparisons*. Addison-Wesley, Reading, MA, 1983.

D. A. Sholl. *The Organisation of the Cerebral Cortex*. Wiley, New York, 1956.

T. J. Shors and E. Gould. xxx. *Nature*, xx:xx, 2001.

A. W. Snyder and H. B. Barlow. Revealing the artist's touch. *Nature*, 331:117–118, 1988.

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

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

J. A. Storer. *Data Compression: Methods and Theory*. Computer Science Press, Rockville, Maryland, 1988.

M. Thompson. Coding third-order image structure. In Baddeley et al. [2000], pages 62–78.

G. von Békésy. *Sensory Inhibition*. Princeton University Press, Princeton, NJ, 1967.

C. S. Wallace and D. M. Boulton. An information measure for classification. *Computer Journal*, 11(2):185–195, 1968.

S. Watanabe. Pattern recognition as information compression. In S. Watanabe, editor, *Frontiers of Pattern Recognition*. Academic Press, New York, 1972.

R. W. Williams and K. Herrup. The control of neuron number. *The Annual Review of Neuroscience*, 11:423–453, 2001.

J. G. Wolff. An algorithm for the segmentation of an artificial language analogue. *British Journal of Psychology*, 66:79–90, 1975.

J. G. Wolff. The discovery of segments in natural language. *British Journal of Psychology*, 68:97–106, 1977.

J. G. Wolff. Language acquisition and the discovery of phrase structure. *Language & Speech*, 23:255–269, 1980.

J. G. Wolff. Language acquisition, data compression and generalization. *Language & Communication*, 2:57–89, 1982.

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.

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

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

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

J. G. Wolff. Probabilistic reasoning as information compression by multiple alignment, unification and search. Technical report, School of Informatics, University of Wales at Bangor, 1998. Copy: www.cognitionresearch.org.uk/papers/pr/pr.htm.

J. G. Wolff. 'Computing' as information compression by multiple alignment, unification and search. *Journal of Universal Computer Science*, 5(11):777–815, 1999a.

J. G. Wolff. Probabilistic reasoning as information compression by multiple alignment, unification and search: an introduction and overview. *Journal of Universal Computer Science*, 5(7):418–462, 1999b.

J. G. Wolff. Syntax, parsing and production of natural language in a framework of information compression by multiple alignment, unification and search. *Journal of Universal Computer Science*, 6(8):781–829, 2000.

J. G. Wolff. Information compression by multiple alignment, unification and search as a framework for human-like reasoning. *Logic Journal of the IGPL*, 9(1):205–222, 2001a. First published in the *Proceedings of the International Conference on Formal and Applied Practical Reasoning (FAPR 2000)*, September 2000, ISSN 1469–4166.

J. G. Wolff. Information compression by multiple alignment, unification and search as a framework for 'intelligent' computing. In *Proceedings of the AISB'01 Symposium on Nonconscious Intelligence: from Natural to Artificial*, pages 422–428. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), March 2001c. ISBN 1 902956 21 3. The Symposium formed part of the AISB'01 Convention, York.

J. G. Wolff. Information compression by multiple alignment, unification and search as a model of non-conscious intelligence. In *Proceedings of the AISB'01 Symposium on Non-conscious Intelligence: from Natural to Artificial*, pages 72–81. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), March 2001b. ISBN 1 902956 21 3. The Symposium formed part of the AISB'01 Convention, York.

J. G. Wolff. Mathematics and logic as information compression by multiple alignment, unification and search. Technical report, CognitionResearch.org.uk, 2002a. Copy: www.cognitionresearch.org.uk/papers/cml/cml.htm.

J. G. Wolff. Unsupervised learning in a framework of information compression by multiple alignment, unification and search. Technical report, CognitionResearch.org.uk, 2002b.

J. G. Wolff. Information compression by multiple alignment, unification and search as a unifying principle in computing and cognition. *Artificial Intelligence Review*, to appear.

G. K. Zipf. *The Psycho-Biology of Language*. Houghton Mifflin, Boston, 1935.

G. K. Zipf. *Human Behaviour and the Principle of Least Effort*. Hafner, New York, 1949.