The SP Theory of Intelligence as a Foundation for the Development of a General, Human-Level Thinking Machine

J Gerard Wolff   JGW@COGNITIONRESEARCH.ORG
CognitionResearch.org,
Menai Bridge, UK

Editor: Pei Wang

Abstract

This paper summarises how the *SP theory of intelligence* and its realisation in the *SP computer model* simplifies and integrates concepts across artificial intelligence and related areas, and thus provides a promising foundation for the development of a general, human-level thinking machine, in accordance with the main goal of research in artificial general intelligence. The key to this simplification and integration is the powerful concept of *multiple alignment*, borrowed and adapted from bioinformatics. This concept has the potential to be the "double helix" of intelligence, with as much significance for human-level intelligence as has DNA for biological sciences.

Strengths of the SP system include: versatility in the representation of diverse kinds of knowledge; versatility in aspects of intelligence (including: strengths in unsupervised learning; the processing of natural language; pattern recognition at multiple levels of abstraction that is robust in the face of errors in data; several kinds of reasoning (including: one-step 'deductive' reasoning; chains of reasoning; abductive reasoning; reasoning with probabilistic networks and trees; reasoning with 'rules'; nonmonotonic reasoning and reasoning with default values; Bayesian reasoning with "explaining away"; and more); planning; problem solving; and more); seamless integration of diverse kinds of knowledge and diverse aspects of intelligence in any combination; and potential for application in several areas (including: helping to solve nine problems with big data; helping to develop human-level intelligence in autonomous robots; serving as a database with intelligence and with versatility in the representation and integration of several forms of knowledge; serving as a vehicle for medical knowledge and as an aid to medical diagnosis; and several more).

Keywords: artificial general intelligence, information compression, multiple alignment, unsupervised learning, natural language processing, pattern recognition, reasoning

1. Introduction

The *SP theory of intelligence* and its realisation in the *SP computer model* is the product of an extended programme of research,[1] aiming to simplify and integrate observations and concepts in artificial intelligence, mainstream computing, mathematics, and human perception and cognition, with information compression as a unifying theme.

---

[1] From about 1987 to early 2006, and from late 2012 to the present.

In its quest for simplification and integration, the research accords with Occam's Razor, one of the most widely-accepted principles in science, and it accords with the goal of developing a general, human-level thinking machine, a goal pursued by early researchers in AI and, more recently, by researchers in artificial general intelligence (AGI).

The name "SP" derives largely from the fact that information compression—which is central in the workings of the SP system—may be seen as a process of promoting *Simplicity* in a body of information by reducing unnecessary complexity in that information whilst retaining as much as possible of its descriptive *Power*. The name also relates to the importance, in Occam's Razor, of combining simplicity in a system with its descriptive or explanatory power.

The main aim of this paper is to summarise how the SP system simplifies and integrates concepts across artificial intelligence and related areas, and thus provides a promising foundation for the development of a general, human-level thinking machine, in accordance with the main goal of research in artificial general intelligence.

The key discovery in this research has been that a concept of *multiple alignment*, borrowed and adapted from bioinformatics, has great versatility in the representation and processing of diverse kinds of knowledge and great versatility in diverse aspects of intelligence. That two-fold versatility of the multiple alignment framework facilitates the seamless integration of different kinds of knowledge and different aspects of intelligence, an integration that appears to be essential for human-level AI.

I believe it is fair to say that multiple alignment has the potential to be the "double helix" of intelligence, with as much significance for human-level intelligence as has DNA for biological sciences.

The next section describes the SP system in outline, while sections that follow aim to demonstrate how the SP system combines relative simplicity with descriptive or explanatory power.

Since it would not be either feasible or reasonable to repeat what has been published before, this paper is largely a summary of already-published work, with frequent references to relevant sources of information.

2. Outline of the SP System

The SP system is described most fully in Wolff (2006b) and more briefly in Wolff (2013). In the brief summary here, the main features of the SP system are these:

- All kinds of knowledge are represented with arrays of atomic *symbols* in one or two dimensions called *patterns*. At present, the SP computer model works only with one-dimensional patterns but it envisaged that the model will be generalised to work with both 1D and 2D patterns.

- A "symbol" is simply a mark that can be compared with any other symbol to determine whether it is the same or different. Any symbol may have a meaning that derives from its association with other symbols. Otherwise, there is no intrinsic meaning for any symbol, except for some distinctions between different kinds of symbol that serve in the workings of the SP computer model.

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

- Processing of information in the SP system is largely a matter of compressing information by searching for patterns, or parts of patterns, that match each other, with a merging or "unification" patterns, or parts of patterns, that are the same. This process—*information compression via the matching and unification of patterns*—may be referred to as "ICMUP" for short.

- More specifically, processing of information in the SP system is largely a process of compressing information by building and processing the previously-mentioned multiple alignments. Two examples of multiple alignments are shown in Figure 1, below, with some explanation in the text.

- Within that general framework, there are two main functions:

  - *Interpretation of New information via the building of multiple alignments.* Here, the system builds one or more multiple alignments, each of one which provides for the economical encoding one New pattern (sometimes more than one) in terms of one or more Old patterns. Depending on the patterns that are used, this process may be seen to achieve such things as the parsing or understanding or production of natural language, the recognition of patterns, several kinds of reasoning, planning, problem solving, or elements of mathematics.

    Because of the complexity of the search for 'good' multiple alignments amongst the many possible 'bad' ones, it is necessary to use heuristic methods to ensure that the computational complexity of the process stays within reasonable bounds.

  - *Unsupervised learning of grammars.* In the SP system, unsupervised learning incorporates the process for building multiple alignments. In addition: 1) the system derives newly-created Old patterns from good multiple alignments and adds them, at least temporarily, to its store of Old patterns, and 2) it searches through the store of Old patterns to create one or more collections of Old patterns or *grammars* which are relatively effective in the economical encoding of New patterns. As with the building of multiple alignments, the complexity of the process of finding good grammars requires heuristic methods to be applied.

  Although these two functions have here been described separately, they are closely interleaved in practice, in much the same way that human learning is closely interwoven with processes for interpreting what we see, hear, feel, and so on.

- Owing to the close connection between information compression and concepts of prediction and probability (Li and Vitányi, 2014), the SP system is fundamentally probabilistic.

- The SP theory is, at one and the same time, a theory of artificial intelligence, mainstream computing, mathematics, and human learning, perception, and thinking. That said, the main emphasis in work to date has been on aspects of human-level

intelligence and, accordingly, the potential of the SP system as a general, human-level thinking machine.

- An important part of the theory is *SP-neural* (Wolff (2016c), Wolff (2006b, Chapter 11)) which tries to translate abstract concepts like 'pattern' and 'multiple alignment' in to corresponding concepts in terms of neurons and communications between neurons.

Figure 1 (a) shows an example of a multiple alignment that achieves the effect of syntactic analysis or parsing of a sentence, 't w o k i t t e n s p l a y', in terms of grammatical categories including words.

The sentence is shown as a New SP pattern in row 0 and the grammatical categories are shown as Old SP patterns in rows 1 to 8, one pattern per row. This multiple alignment is the best multiple alignment found by the SP computer model when it is supplied with the New pattern and a set of Old patterns representing grammatical categories, including word. Here, 'best' means that the multiple alignment is the one out of many that are produced by the SP computer model that achieves the greatest compression of the New pattern in terms of Old patterns.

A point of interest is that the pattern in row 8 of the multiple alignment marks the sentence as having a plural form ('PL'), with an association between a plural subject in the sentence ('Np') with a plural main verb ('Vp'). This way of marking discontinuous dependencies in syntax has an important role in how the SP system may model the subtle and intricate pattern of dependencies in the syntax of English auxiliar verbs (Wolff (2013, Sections 8.1, 8.2, and 8.3), Wolff (2006b, Section 5.5)).

Figure 1 (b) shows how the SP computer model may produce a plausible parsing despite errors in the New pattern. In this example, there is an error of omission (the absence of 'w' in 't w o'), an error of substitution (the word 'k i t t e n s' contains 'm' instead of 'n'), and an error of addition (the word 'p l a y' has an added 'x').

The robustness of the SP system in the face of errors is not restricted to the processing of natural language. It is a general feature of the SP system that applies in all aspects of the system's workings. The ability of the SP system to tolerate errors in its inputs, within limits, is a by-product of the above-noted probabilistic nature of the system, itself the product of information compression as a central part of how the system works.

The two examples here give only a taste of what can be done with multiple alignment. As noted in the Introduction and outlined below, the multiple alignment concept, as it has been developed in the SP programme of research, is a powerful, versatile vehicle for several different aspects of computing and cognition.

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

(a)

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

(b)
```
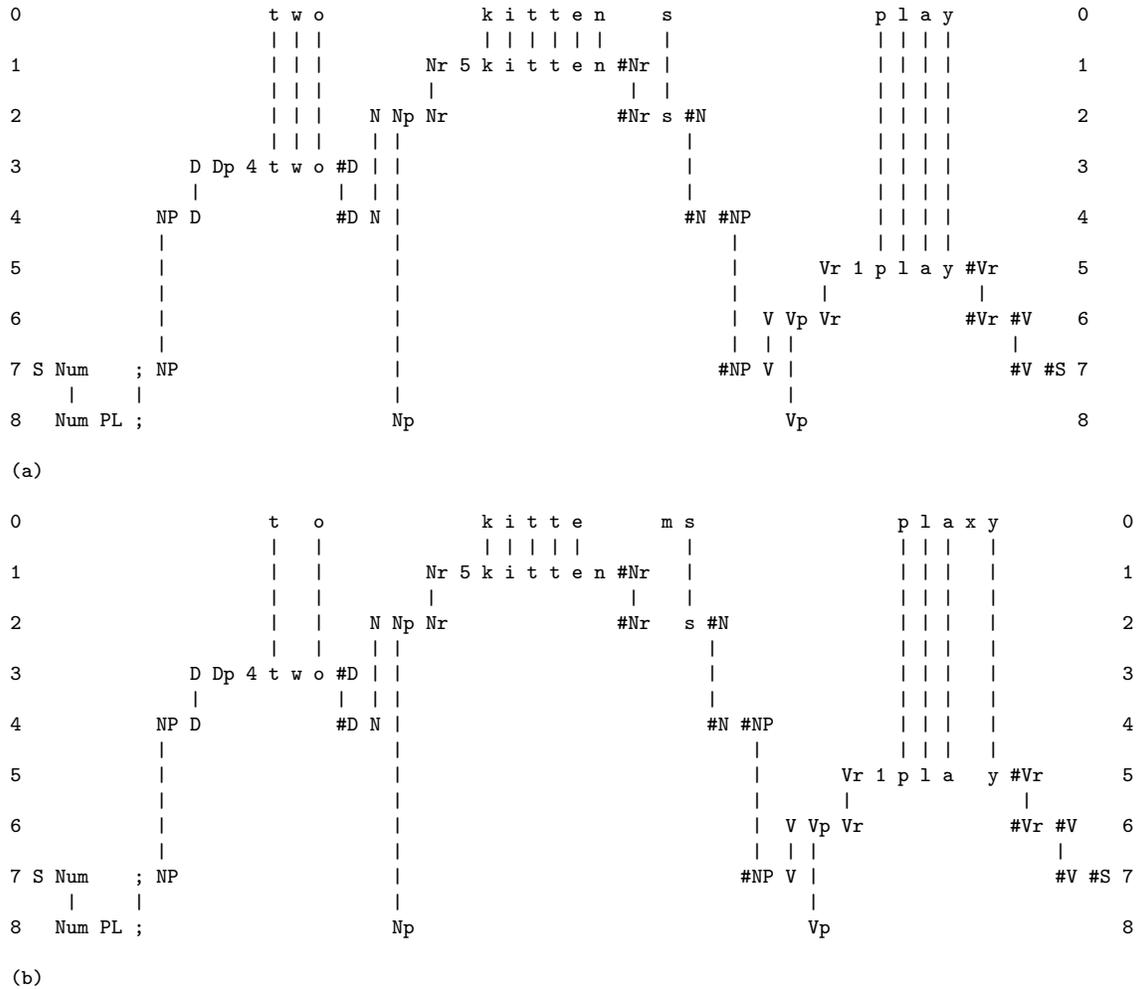
Figure 1: a) The best multiple alignment created by the SP computer model with a store of Old patterns like those in rows 1 to 8 (representing grammatical structures, including words) and a New pattern shown in row 0 (representing a sentence to be parsed). Adapted from Figures 1 and 2 in Wolff (2007), with permission.

3. Evaluation of the SP System in Terms of Simplicity and Power

This section and the ones that follow summarise evidence that, in accordance with Occam's Razor and the goal of developing a general, human-level thinking machine, the SP theory combines simplicity with explanatory or descriptive power.

The essential simplicity of the SP system may be seen from two perspectives:

- The SP system, as it is expressed in the SP computer model, is largely a process of building multiple alignments and creating grammars, with information compression as a touchstone of success. As such, it is remarkably simple, considering the range of things that it can do, outlined in the sections that follow.

- The 'exec' file for the SP computer model takes less than 500 kB of storage space.

The following sections describe aspects of the SP system that may be seen to be part of its explanatory or descriptive power.

At some stage in the future it may be possible to evaluate a scientific theory like the SP theory with mathematically-precise measures of simplicity and power. Since this is not yet feasible, we have to rely on more traditional assessments in terms of 'elegance', 'beauty' and, ultimately, the usefulness of the system in the solution of practical problems.

The strengths and potential of the SP system described in the following sections includes features of the system that have been demonstrated already and others that may reasonably be expected with some further development of the system. Relevant evidence is mainly in Wolff (2006b, Chapters 5 to 10). Other sources of evidence are shown where they apply.

4. Versatility in the Representation of Knowledge

Although SP patterns are not, in themselves, very expressive, they come to life in the building of multiple alignments. Within that framework, they may serve in the representation of several different kinds of knowledge, listed below.

Here are the main kinds of knowledge that, in evidence to date, may be seen to fall within the scope of the SP system: the syntax of natural language; class hierarchies and class heterarchies (the latter term meaning class hierarchies with cross classification); part-whole hierarchies; discrimination networks and trees; rules for reasoning; structures in mathematics; entity-relationship structures (Wolff, 2007, Sections 3 and 4); relational knowledge (*ibid.*, Section 3); patterns in one or two dimensions; images; structures in three dimensions (Wolff, 2014a, Section 6.1 and 6,2); and procedural knowledge (Wolff, 2014d, Section 6.6.1), including parallel procedures (Wolff, 2014b, Sections V-G, V-H, and V-I, and Appendix C).

Since information compression is at the heart of the SP system—in the building of multiple alignments and in unsupervised learning—and since the techniques for the compression of information within the SP system are very general, there is reason to believe that the system may serve in the efficient representation of *any* kind of knowledge, not just the examples that have been listed in this section.

In general, the SP system has potential as a *universal framework for the representation and processing of diverse kinds of knowledge* (UFK). As such, it has potential help solve a significant problem with big data: the great variety of formalisms and formats for knowledge and resulting problems in the analysis of big data and for the discovery of associations and structures in big data (Section 7, Wolff (2014c, Section III)).

5. Versatility in Intelligence

The building of multiple alignment is not merely a means of giving life to SP patterns in the representation of different kinds of knowledge. It gives expression to several different aspects of intelligence, and it lies at the heart of processes within the SP system for unsupervised learning.

In summary, the following aspects of the intelligence fall within the scope of the SP system: unsupervised learning (including the unsupervised learning of segmental structures,

classes of structure, and abstract patterns); the parsing and production of natural language with preliminary evidence that the system supports the understanding of natural language and the production of natural language from meanings; pattern recognition that is robust in the face of errors in data (more below); recognition at multiple levels of abstraction; best-match and semantic forms of information retrieval; several kinds of reasoning (more below); planning; and problem solving (Wolff, 2006b, Chapters 5 to 10).

The versatility of the process for building multiple alignments may be seen most strikingly in how it gives expression to several different kinds of reasoning. These include: one-step 'deductive' reasoning; chains of reasoning; abductive reasoning; reasoning with probabilistic networks and trees; reasoning with 'rules'; nonmonotonic reasoning and reasoning with default values; Bayesian reasoning with "explaining away"; causal reasoning; and reasoning that is not supported by evidence (Wolff, 2006b, Chapter 7). It also supports inheritance of attributes in class-inclusion hierarchies and part-whole hierarchies (Wolff, 2006b, Section 6.4) and there is clear potential for spatial reasoning (Wolff, 2014b, Section IV-F.1), and for what-if reasoning (Wolff, 2014b, Section IV-F.2).

The probabilistic nature of the SP system (Section 2) means that it can yield plausible results in the face of errors of omission, commission, or substitution in any of its New patterns. This can be seen most clearly in such functions as the parsing of natural language (illustrated in Figure 1 (b)) or the recognition of patterns, but it applies to all aspects of its processing.

There is evidence that the SP system may provide a vehicle not only for the representation of mathematical structures (Section 4) but for mathematical reasoning and calculations as well. There is a discussion of those possibilities in Wolff (2006b, Chapter 10) and (Wolff, 2016b, Section 10).

6. Seamless Integration of Diverse Kinds of Knowledge and Diverse Aspects of Intelligence

Multiple alignment, with SP patterns, as a relatively simple vehicle for the representation of diverse kinds of knowledge (Section 4) and for diverse aspects of intelligence (Section 5) brings with it a substantial benefit. It means that there is clear potential for the seamless integration of different kinds of knowledge and different aspects of intelligence, in any combination.

It appears that that kind of integration is essential in any system that aspires to achieve human levels of intelligence. Thus it is a major plank in the claim that the SP system provides a promising foundation for the development of a general, human-level thinking machine (Section 1).

7. Potential Benefits and Applications of the SP system

There are several potential benefits and applications of the SP system, some of them—like the nine potential benefits with big data (below)—were quite unexpected. In evaluating the SP system (Section 3), all the potential benefits and applications are strengths of the system, to be set against its relative simplicity.

Possibilities that have been explored include the following:

- *Several potential benefits and applications have been outlined in Wolff (2014d).* They include:

  - *Simplification of computing systems.* There is clear potential for a substantial simplification of computer systems, including software.
  - *Processing of natural language.* In view of the strengths of the SP system in the representation of the syntax of natural language and non-syntactic or semantic forms of knowledge (Section 4) and in the processing of natural language (Section 5), there is potential for the system to be applied in the understanding and production of natural language and in translating between natural languages;
  - *Bioinformatics.* Since the multiple alignment concept—a key part of the SP system—has been borrowed from bioinformatics, it is likely that the SP system would prove useful in that area of application;
  - *Managing documentation.* There is potential for the system to be applied in managing the diverse kinds of document in a typical software development project, bearing in mind that each document is likely to come in several parts and sub-parts, and that each part may come in several different versions. Likewise for other projects with similar requirements;
  - *Software engineering.* Since the main elements of a conventional computer program may be mapped on to corresponding structures in multiple alignments (Wolff, 2014d, Section 6,7), including the representation of parallel processing (Wolff, 2014b, Sections V-G, V-H, and V-I, and Appendix C), there is potential for the system to be applied in software engineering, both in manual programming of a traditional kind and in automatic or semi-automatic programming via unsupervised learning;
  - *Compression of information.* Since the SP system works entirely via the compression of information, it is likely to prove useful in applications intended for that purpose;
  - *The semantic web.* The strengths of the SP system in the representation of knowledge, in reasoning, in the processing of natural language, in unsupervised learning, and in coping with uncertainty, means that it has potential as an aid to the realisation of the Semantic Web.
  - *Detection of computer viruses.* The strengths of the SP system in pattern recognition (Section 5) suggest that it is likely to be effective in the detection of computer viruses.
  - *Data fusion.* The strengths of the SP system in information compression suggest that it may serve in "data fusion", the merging of streams of data from different sources.

- *Big data.* The SP theory has the potential to help solve nine problems with big data, as described in Wolff (2014c). In summary:

  - *Overcoming the problem of variety in big data.* There is potential to reduce all the many formalisms and formats for data to one *universal framework for the representation and processing of diverse kinds of knowledge* (UFK).

- *Learning and discovery.* The SP system has potential for the unsupervised learning or discovery of 'natural' structures in data (Wolff, 2013, Section 5.2).

- *Interpretation of data.* As indicated in Section 2, the SP system has strengths in the interpretation of data in the widest sense that includes parsing of natural language, pattern recognition and more.

- *Velocity: analysis of streaming data.* The SP system lends itself to an incremental style, assimilating information as it is received, much as people do.

- *Volume: making big data smaller.* Reducing the size of big data via lossless information compression can yield direct benefits in the storage, management, and transmission of data, and indirect benefits in several other areas outlined in this article.

- *Additional economies in the transmission of data.* The SP system in conjunction with model-based coding (Pierce, pp. 139–140) has potential for very substantial economies in the transmission of data, first via the unsupervised learning of a *grammar* for the kind of data to be transmitted, second, via the distribution of a copy of the grammar to all potential senders and receivers of data, and third, via the *encoding* by "Alice" of any one body of data such as a TV programme, the *transmission* of the encoding to "Bob", and his *decoding* of the encoded data using the same grammar ((Wolff, 2017), Wolff (2014c, Section VIII)).

- *Energy, speed, and bulk.* With the SP system, there is potential for big cuts in the use of energy by computers, for greater speed in processing with a given computational resource, and for corresponding cuts in the size and weight of computers.

- *Veracity: managing errors and uncertainties in data.* The SP system can identify possible errors or uncertainties in data, suggest possible corrections or interpolations, and calculate associated probabilities.

- *Visualisation.* Knowledge structures created by the system, and inferential processes in the system, are all transparent and open to inspection. They lend themselves to display with static and moving images.

- *Autonomous robots.* The SP system can help develop human-like versatility and adaptability in autonomous robots (which would be helpful in places like Mars where there is relatively little scope for direct control by people), and the system can help to increase the computational and energy efficiency of robot 'brains' and to reduce their size and weight (which are potential problems since each autonomous robot must carry all its intelligence with it and likewise for its energy supplies) (Wolff, 2014b).

- *Intelligent database.* The SP system may serve as a database system with versatility in the representation and integration of several different forms of knowledge and with versatility in several aspects of intelligence which may work together in any combination (Wolff, 2007).

- *Medical diagnosis.* The SP system may serve as a vehicle for the storage and representation of knowledge about diseases and for the application of that knowledge

in medical diagnosis (Wolff, 2006a). Like other systems of that kind, it would probably be most useful as an assistant to people with medical training rather than as an authority in its own right.

- *Commonsense reasoning.* A recent paper by Ernest Davis and Gary Marcus (2015) argues persuasively that commonsense reasoning—the kind of reasoning we use every day in a wide variety of situations—is surprisingly challenging for artificial intelligence. A paper that I have drafted (Wolff, 2016a) argues, with examples drawn from the Davis and Marcus paper, that the SP system provides a promising foundation for further developments in this area.

8. Other Strengths of the SP System

The SP system has other strengths that don't fall neatly into the areas outlined in the preceding four sections. These other strengths are described briefly here:

- *Building from primitive foundations.* The SP theory excludes ready-made ideas from mathematics, logic, and computer science, like 'equivalence' ('⇔'), 'implies' ('⇒'), 'is-provable' ('⊢'), *variable*', 'equals' ('='), '*if ... then*', and so on. This is because the theory is not only a theory of AI, computing, and mathematics, but also a theory of human learning, perception and cognition (Section 2). Because the primitives in the theory need to be plausible as inborn features of our cognitive machinery, the SP theory has been grounded at a low level that excludes concepts like those mentioned above, which are the products of academic research and which we learn in school or college.

  Of course, the SP theory needs to account for how our minds can accommodate concepts from mathematics, logic, and computer science. How some concepts of those kinds may be modelled in the multiple alignment framework is described in Wolff (2014d, Section 6.6.1), Wolff (2016b, Section 10), and Wolff (2006b, Chapter 10).

- *Only one kind of processing.* In keeping with the remarks just made, all kinds of processing in the SP system is done via the compression of information by the matching and unification of patterns and, more specifically, via the building and processing of multiple alignments. All kinds of inference that have, so far, been demonstrated with the SP system flow from information compression via the matching and unification of patterns without the need to assume that we are born with higher-level capabilities of the kinds that feature in logic, mathematics, and the like.

- *The significance of information compression.* In its underlying philosophy, the SP system is consistent with an accumulation of evidence that information compression is significant in both natural cognition and artificial computing:

  - *Information compression in the workings of brains and nervous systems.* There is a body of research by Attneave (1954), Barlow (2001, 1969), and others, showing the importance of information compression in the workings of brains and nervous systems.

- *Information compression in learning.* There is a body of evidence that information compression is central in the learning of natural language (Wolff, 1988). It appears that the same principles apply to the learning of other kinds of knowledge.

- *Information compression and generalisation.* Information compression provides a neat answer to the problem of generalisation in machine learning without over-generalisation or under-generalisation (Wolff, 2013, Section 5.3). In brief, the product of compressing any given body of information **I** may be seen to comprise a grammar and an encoding of **I** in terms of the grammar. There is quite good evidence that the grammar provides for generalisation beyond **I** without either over-generalisation (underfitting) or under-generaliaation (overfitting).

- *The DONSVIC principle.* There is evidence that, when machine learning is designed to achieve relatively high levels of information compression, the resulting structures take forms that people regard as natural. This principle—*the discovery of natural structures via information compression* (DONSVIC)—provides indirect support for the use of information compression as the foundation for machine learning and other aspects of cognition, as it is in the SP system.

- *Information compression and prediction and probability.* As noted in Section 2, there is an intimate connection between information compression and concepts of prediction and probability. This provides support for information compression as the foundation for computing, cognition, and mathematics, since all those things may be seen to be largely about predicting the future from the past. And it may be seen to be the basis for the above-noted robustness of the SP system in the face of errors, something with clear relevance to the natural selection of living things (next).

- *Inforrmation compression in natural selection.* It is likely that, with any organism that stores and uses information as an aid to living, natural selection would favour compression of that information: by reducing the storage space needed for any given body of information, or increasing the effective capacity of any storage space; by reducing the bandwidth needed to transmit any body of information, or increasing what can be sent with a given bandwidth, or increasing speeds of transmission; and perhaps most importantly by allowing the organism to make predictions and calculate probabilities for events in the future (Wolff, 2016b, Section 4)—and, as mentioned above, to compensate for errors in the system's data, provided they are not too large or numerous.

- *SP-neural.* Some of the abstract concepts in the SP theory—mainly SP patterns and, less certainly, the building of multiple alignments—map fairly straightforwardly into *SP-neural,* a version of the SP theory expressed in terms of neurons, connections between neurons, and communications amongst neurons (Wolff (2016c), (Wolff, 2006b, Chapter 11)). Although SP-neural is still fairly sketchy and needs to be developed and refined as a working computer model, the fact that one can envisage how abstract concepts in the SP theory may be realised in terms of neurons and their interconnections, lends support to the SP theory as a theory of cognition as well as

11

being a theory of AI, mainstream computing, and mathematics. The SP theory would be weaker if it were purely abstract without any possibility of seeing how the concepts might be realised in the brain.

It is not without interest that the concept of a "pattern" in the SP theory is quite similar to Donald Hebb's (1949) concept of a "cell assembly", a concept which itself has evidence in its support and is consistent with a weight of evidence in support of "grandmother" cells and localist models of the brain (Roy, 2013; Gross, 2002).

Although SP-neural is broadly consistent with Hebb's concept of a cell assembly, learning processes in the SP theory overcome a weakness in Hebb's theory of learning, a weakness that Hebb himself recognised. The problem with "Hebbian" learning—the gradual strengthening of connections between neurons that frequently fire at the same time—is that it does not explain the commonplace observation that we can and often do learn from a single exposure or experience. The SP theory overcomes this weakness by providing for learning from a single experience and, at the same time, it explains why it takes time to learn complex knowledge and skills—because of the large abstract space that has to be searched to develop solutions that are 'good' in terms of information compression.

Despite its prominent weakness, Hebbian learning has been enthusiastically adopted as the basis for most neural theories of learning and cognition, including deep learning. This is just one of several problems with deep learning, mentioned under the next bullet point.

- *Distinctive features and advantages of the SP theory.* To highlight what is special about the SP theory, its distinctive features are summarised in the first part of Wolff (2016d). At substantially more length, that paper describes *advantages* of the SP theory compared with a range of AI-related alternatives. In particular, Section V of the paper describes 13 problems with deep learning and how, in the SP system, they may be overcome.

9. Conclusion

In this paper, I have tried to show that the SP system, in accordance with Occam's Razor, combines relative simplicity with a relatively large descriptive and explanatory power: in the representation of several kinds of knowledge; in how it accounts for several aspects of intelligence; in the seamless integration of diverse kinds of knowledge and diverse aspects of intelligence—something that appears to be essential in any model that aspires to human-level intelligence; in several potential benefits and applications of the system; and in other aspects that don't fit into the forgoing categories.

I believe that, in its favourable combination of simplicity and power, the SP system clears the path towards the development of a general, human-level thinking machine or artificial general intelligence. Of course there is more to be done but the SP research appears to provide a promising foundation for the development of AGI. As was mentioned in the Introduction, multiple alignment—which is central in the workings of the SP system—has the potential to be the "double helix" of intelligence, with as much significance for human-level intelligence as has DNA for biological sciences.

9.1 In Some Respects, the Human Mind Is a Kluge

Trying to develop a general, human-level thinking machine is all very well but it seems to be in conflict with quite strong evidence, presented by Gary Marcus (2008), that, in many respects, the human mind is a kluge: "a clumsy or inelegant—yet surprisingly effective— solution to a problem" (*ibid.*, p 2). In view of that evidence, shouldn't the focus of our research be on developing a model of human intelligence, in the spirit of Minsky's "Society of Mind" (Minsky, 1986), that recognises the messiness of human thinking and takes account of the way that evolution always works with what is at hand so that it may not always arrive at a solution that is theoretically ideal?

The suggestion here is that we need to do both things: we need to recognise that, very likely, the human mind is in many respects a kluge, and we need to continue searching for powerful general mechanisms behind the messiness of the human mind. The latter posture is dictated by the central goal of science—to search for simplicity behind the apparent complexity of the world—but it is also dictated by the failure of AI research over many decades to develop any account of human learning, perception and thinking, either as a kluge or as an elegant general mechanism, that does justice to the versatility and adaptability of human intelligence.

9.2 We Need More Than One Approach to the Development of Human-Level AI

As mentioned earlier, I have described in another paper (Wolff, 2016d) distinctive features of the SP theory compared with AI-related alternatives and, in more detail, advantages of the SP theory compared with a selection of those alternatives.

That analysis may suggest a winner-takes-all approach to the management of research. But, to the contrary, there is already an over-strong tendency for one dominant approach to a problem to squeeze out alternative approaches, witness the relatively enormous publicity and resources currently devoted to research in deep learning, despite the 13 shortcomings of deep learning, with solutions in the SP system, described in Wolff (2016d, Section V).

Given the difficulties of achieving human-level AI, we need a multi-pronged strategy. Rather than piling all our eggs into one basket, there should be research on several fronts, including deep learning.

Given the evidence presented in this paper, that multiple alignment may indeed provide the basis for a general, human-level thinking machine, I believe that the SP system is a strong candidate for further research and development.

Acknowledgements

References

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

Barlow, H. B. 1969. Trigger features, adaptation and economy of impulses. In Leibovic, K. N., ed., *Information Processes in the Nervous System*. New York: Springer. 209–230.

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

Davis, E., and Marcus, G. 2015. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM* 58(9):92–103.

Gross, C. G. 2002. Genealogy of the "grandmother cell". *Neuroscientist* 8(5):512–518.

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

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

Marcus, G. 2008. *Kluge: the Hapharzard Construction of the Human Mind.* London: Faber and Faber, paperback edition. ISBN: 978-0-571-23652-7.

Minsky, M., ed. 1986. *The Society of Mind.* New York: Simon & Schuster.

Pierce, J. R. *Symbols, Signals and Noise.* New York: Harper & Brothers. The first edition was published in 1961. The Dover edition, *An Introduction to Information Theory: Symbols, Signals & Noise*, is a second, revised edition, published in 1980. It is available via Kindle.

Roy, A. 2013. An extension of the localist representation theory: grandmother cells are also widely used in the brain. *Frontiers in Psychology* 4:1–3.

Wolff, J. G. 1988. Learning syntax and meanings through optimization and distributional analysis. In Levy, Y.; Schlesinger, I. M.; and Braine, M. D. S., eds., *Categories and Processes in Language Acquisition.* Hillsdale, NJ: Lawrence Erlbaum. 179–215. bit.ly/ZIGjyc.

Wolff, J. G. 2006a. Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, unification and search. *Decision Support Systems* 42:608–625. bit.ly/XE7pRG, arXiv:1409.8053 [cs.AI].

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

Wolff, J. G. 2007. Towards an intelligent database system founded on the SP theory of computing and cognition. *Data & Knowledge Engineering* 60:596–624. bit.ly/Yg2onp, arXiv:cs/0311031 [cs.DB].

Wolff, J. G. 2013. The SP theory of intelligence: an overview. *Information* 4(3):283–341. bit.ly/1hz0lFE.

Wolff, J. G. 2014a. Application of the SP theory of intelligence to the understanding of natural vision and the development of computer vision. *SpringerPlus* 3(1):552–570. bit.ly/1scmpV9.

Wolff, J. G. 2014b. Autonomous robots and the SP theory of intelligence. *IEEE Access* 2:1629–1651. bit.ly/1zrSemu.

Wolff, J. G. 2014c. Big data and the SP theory of intelligence. *IEEE Access* 2:301–315. bit.ly/1jGWXDH. This article, with minor revisions, is reproduced in Fei Hu (Ed.), *Big Data: Storage, Sharing, and Security (3S)*, Taylor & Francis LLC, CRC Press, 2016, pp. 143–170.

Wolff, J. G. 2014d. The SP theory of intelligence: benefits and applications. *Information* 5(1):1–27. bit.ly/1lcquWF.

Wolff, J. G. 2016a. Commonsense reasoning, commonsense knowledge, and the SP theory of intelligence. In preparation. bit.ly/2eBoE9E.

Wolff, J. G. 2016b. Information compression, intelligence, computing, and mathematics. Technical report, CognitionResearch.org. In preparation. bit.ly/1jEoECH, arXiv:1310.8599 [cs.AI].

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

Wolff, J. G. 2016d. The SP theory of intelligence: its distinctive features and advantages. *IEEE Access* 4:216–246. bit.ly/21gv2jT.

Wolff, J. G. 2017. The great potential of model-based coding with the SP theory of intelligence in the transmission of information. Submitted to IEEE Eurocon 2017. bit.ly/2h1h5b9.