

Received November 2, 2015, accepted December 8, 2015, date of publication December 31, 2015, date of current version March 1, 2016.

Digital Object Identifier 10.1109/ACCESS.2015.2513822

The SP Theory of Intelligence: Distinctive Features and Advantages

JAMES GERARD WOLFF, (Member, IEEE)

CognitionResearch.org, Menai Bridge LL59 5LR, U.K.

Corresponding author: J. G. Wolff (jgw@cognitionresearch.org)

ABSTRACT This paper aims to highlight distinctive features of the SP theory of intelligence, realized in the SP computer model, and its apparent advantages compared with some AI-related alternatives. Perhaps most importantly, the theory simplifies and integrates observations and concepts in AI-related areas, and has potential to simplify and integrate of structures and processes in computing systems. Unlike most other AI-related theories, the SP theory is itself a theory of computing, which can be the basis for new architectures for computers. Fundamental in the theory is information compression via the matching and unification of patterns and, more specifically, via a concept of multiple alignment. The theory promotes transparency in the representation and processing of knowledge, and unsupervised learning of natural structures via information compression. It provides an interpretation of aspects of mathematics and an interpretation of phenomena in human perception and cognition. Abstract concepts in the theory may be realized in terms of neurons and their inter-connections (SP-neural). These features and advantages of the SP system are discussed in relation to AI-related alternatives: the concept of minimum length encoding and related concepts, how computational and energy efficiency in computing may be achieved, deep learning in neural networks, unified theories of cognition and related research, universal search, Bayesian networks and some other models for AI, IBM's Watson, solving problems associated with big data and in the development of intelligence in autonomous robots, pattern recognition and vision, the learning and processing of natural language, exact and inexact forms of reasoning, representation and processing of diverse forms of knowledge, and software engineering. In conclusion, the SP system can provide a firm foundation for the long-term development of AI and related areas, and at the same time, it may deliver useful results on relatively short timescales.

INDEX TERMS Artificial intelligence, information compression, multiple alignment, perception, cognition, neural networks, deep learning, unsupervised learning, reasoning, mathematics.

I. INTRODUCTION

The *SP theory of intelligence* is designed to simplify and integrate observations and concepts across artificial intelligence, mainstream computing, mathematics, and human perception and cognition, with information compression via *multiple alignment* as a unifying theme. It is realized in the SP computer model which may be regarded as an early version of the *SP machine*. The theory and some of its potential benefits and applications are described in outline in Appendix I, with pointers to where fuller information may be found.

The name 'SP' stands for *simplicity* and *power*, two ideas which are equivalent to information compression (Appendix I-E) and thus central in the workings of the SP system, and also in Occam's Razor and the evaluation of scientific theories, as described in Appendix II.

The SP theory is one of many theories, systems, or schemes that are vying for the attention of researchers in AI and

related areas. Since it takes a fair amount of work to fully understand any one of these competing systems, researchers are naturally selective about where they will concentrate their efforts. Accordingly, the main aim of this paper is to present a case for the SP theory.

The beginnings of the argument are simply to show potential benefits and applications of the theory. Much of this has been done already in previous publications referenced in Appendix I.

But this paper goes further. Instead of merely describing what the theory is good for, the paper highlights what is *distinctive* about the SP theory compared with alternatives and, more importantly, it argues, immodestly, that in terms of the long-term development of AI-related research, the SP theory has *advantages* compared with alternatives, and, at the same time, it has potential in some areas on relatively short timescales. Some of the pros and cons of long- and

short-term perspectives in research are discussed briefly in Appendix III.

In comparing the SP system with others, this paper aims to avoid unnecessary repetition of material that has already been published elsewhere. Trying to include, within this paper, a comprehensive description of the SP system and its applications would not be feasible or reasonable. But to aid understanding and readability, some repetition is unavoidable. Many aspects of the SP system are described in outline or summarised. The paper includes several figures, several of them new, showing output from the SP computer model, with explanations in the text.

The next section provides a broad-brush view of distinctive features of the SP theory, and its strengths. Sections that follow aim to highlight apparent advantages of the SP concepts compared with AI-related alternatives, acknowledging intellectual debts of the SP system, and its shortcomings.

II. OVERVIEW OF DISTINCTIVE FEATURES AND STRENGTHS OF THE SP THEORY

This section is an expanded and revised version of [74, Sec. II-G], summarising distinctive features and strengths of the SP system.

A. SIMPLIFICATION AND INTEGRATION OF OBSERVATIONS AND CONCEPTS

As noted in the Introduction and Appendix I, the SP theory aims to simplify and integrate observations and concepts across a broad canvass. Although the theory is not complete (Appendix I-J), there is now much evidence that the attempt is proving successful—that the SP theory, in accordance with Occam's Razor (Appendix II-B), combines relative simplicity with descriptive and explanatory power across a wide range of observations and concepts (Appendix I-H) and across a wide range of potential benefits and applications (Appendix I-I).

Combining relative simplicity with descriptive and explanatory power in AI-related areas is perhaps the most distinctive feature of the SP theory, and its major strength.

B. SIMPLIFICATION AND INTEGRATION IN COMPUTING SYSTEMS

The provision of one simple format for knowledge (Appendix I-D), and one framework for the processing of knowledge (Appendix I-E), promotes an overall simplification of computing systems, including both hardware and software [77, Sec. 5].

Those two things also promote seamless integration of diverse kinds of knowledge and diverse aspects of intelligence [77, Sec. 7], an integration that appears to be necessary if we are to achieve human-like versatility and adaptability in AI [74, Sec. IV-A].

The SP computer model is not yet a rival to some systems that are dedicated to specific functions. But its long-term potential is likely to be greater than dedicated systems because of its strong theoretical foundations, and because it

is designed for the simplification and integration of diverse forms of knowledge, diverse kinds of processing, and diverse aspects of intelligence.

C. THE SP THEORY IS ITSELF A THEORY OF COMPUTING

Most other AI-related systems are founded on the concept of 'computing' as the workings of the universal Turing machine [61] or equivalent models such as lambda calculus [2] or Post's canonical system [47].¹

By contrast, *the SP theory is itself theory of computing* which appears to be Turing-equivalent as argued in [70, Chapter 4]. The gist of the argument is that the SP machine can do much the same as a Post canonical system, a type of computing system which is recognised as being equivalent to a universal Turing machine.

Another reason for believing that the SP system is likely to have the generality of a universal Turing machine is that the main elements of computer programming can be modelled by the multiple alignment concept, which is itself central in the workings of the SP system (Section XI-F).

What is distinctive about the SP theory as a theory of computing is that it provides much of the human-like intelligence that is missing from earlier models (Appendices I-H and I-I. See also Appendix II-B).

D. NEW ARCHITECTURES FOR COMPUTERS

The SP theory may provide the foundations for two new architectures for computers:

- The *multiple alignment* framework (Section II-F and Appendix I-E2) may itself provide the basis for a new kind of computer.
- *SP-neural*—the realisation of the multiple alignment framework in terms of neurons and their interconnections (Section II-L and Appendix I-G)—may provide the basis for a different but related new architecture for computers.

The potential significance of these architectures in terms of the computational and energy efficiency of computers is outlined in Section IV.

E. INFORMATION COMPRESSION VIA THE MATCHING AND UNIFICATION OF PATTERNS

In trying to cut through some of the mathematical complexities associated with information compression, the SP research programme focuses on a simple, 'primitive' idea: that information compression may be understood as a search for patterns that match each other, with the merging or 'unification' of patterns that are the same. This will be referred to as 'ICMUP' meaning "information compression via the matching and unification of patterns". The potential advantage of this approach is that it can help us avoid old trammings, and open doors to new ways of thinking ([76, Sec. 2], Appendix I-E1).

¹ An apparent exception is the concept of a "neural Turing machine" [14].

F. INFORMATION COMPRESSION VIA MULTIPLE ALIGNMENT

More specifically, ICMUP provides the basis for a concept of *multiple alignment*, borrowed and adapted from that concept in bioinformatics (Appendix I-E2). Developing this idea as a framework for the simplification and integration of concepts across a broad canvass has been a major undertaking. Some of the versatility of this concept may be seen in multiple alignments shown in this paper (Figures 1, 2, 3, 4, 6, 8, 10, and 11) and in other publications about the SP system.

Multiple alignment is a distinctive and powerful idea in the SP programme of research. This concept, as it has been developed in the SP programme of research, has the potential to be as significant in computing and cognition as the double helix has proved to be in biological sciences.

Because of the significance of ICMUP and multiple alignment in the SP theory, no attempt has been made in this paper to review or consider the vast literature on information compression via other techniques.

As a general rule, applications in which information compression has a role to play have been developed for specific purposes and do not aspire to simplify and integrate concepts across a broad canvass as the SP system is intended to do.

Section III discusses how the SP system relates to minimum length encoding and related topics, and Section V-I makes some remarks about information compression in deep learning with artificial neural networks.

G. THE SP SYSTEM IS INTRINSICALLY PROBABILISTIC

For reasons outlined in Appendix I-F, the SP system is probabilistic at its deepest levels, but the all-or-nothing nature of conventional systems may be imitated if required.

H. TRANSPARENCY IN THE REPRESENTATION AND PROCESSING OF KNOWLEDGE

By contrast with sub-symbolic approaches to artificial intelligence, and notwithstanding objections to symbolic AI,² knowledge in the SP system is transparent and open to inspection, and likewise for the processing of knowledge.

I. UNSUPERVISED LEARNING AND THE DONSVIC PRINCIPLE

A related point is that unsupervised learning in the current version of the SP computer model conforms to the ‘DONSVIC’ principle—*The Discovery of Natural Structures Via Information Compression* [72, Sec. 5.2]. Evidence to date suggests that, by contrast with sub-symbolic approaches to artificial intelligence, structures created via unsupervised learning in the SP system will normally be structures that people regard as natural and comprehensible.

J. MATHEMATICS

By contrast with other approaches to artificial intelligence, mainstream computing, or human perception and cognition,

²See, for example, “Hubert Dreyfus’s views on artificial intelligence”, *Wikipedia*, <http://bit.ly/1hGHV8m>, retrieved 2014-08-19.

the SP theory has quite a lot to say about the nature of mathematics. In brief, it appears that several aspects of mathematics may be understood in terms of ICMUP and, potentially, in terms of multiple alignment ([70, Ch. 10], [76]). Although logic has received less attention in the SP programme of research, it seems likely that similar principles will apply there [70, Ch. 10].

K. HUMAN PERCEPTION AND COGNITION

The SP theory draws extensively on research on human perception and cognition. In particular, it is founded, in part, on research developing computer models of the learning of natural language, underpinned by empirical evidence. This research is summarised in [68] and in the web page “Language learning” in <http://www.cognitionresearch.org>, with download links to papers.

L. SP-NEURAL

The SP theory includes proposals—*SP-neural*—for how abstract concepts in the theory may be realized in terms of neurons and neural processes. The SP-neural proposals (Appendix I-G) are significantly different from artificial neural networks as commonly conceived in computer science, and arguably more plausible in terms of neuroscience.

III. MINIMUM LENGTH ENCODING, ALGORITHMIC INFORMATION THEORY, AND KOLMOGOROV COMPLEXITY

This section and the ones that follow consider how the SP theory relates to a selection of AI-related concepts, emphasising distinctive features of the theory and its apparent advantages compared with alternatives, but also acknowledging shortcomings in the SP system as it is now, and where it has drawn inspiration from earlier work.

As mentioned in Appendix I-E, information compression in the SP theory may be seen as an example of the principle of *minimum length encoding* (MLE) [56, 65, 48]. Also, information compression and MLE are closely related to *algorithmic information theory* (AIT), and *Kolmogorov complexity* (KC) [35].

Amongst these inter-related areas of study, distinctive features of the SP theory are:

- Most research on information compression, MLE, AIT, and KC, is founded on the assumption that ‘computing’ is defined by the universal Turing machine. By contrast, the SP theory is itself a theory of computing (Section II-C).
- By contrast with most research in these areas, there is a central role in the SP theory for ICMUP (Section II-E) and, more specifically, the concept of multiple alignment (Section II-F).

IV. COMPUTATIONAL AND ENERGY EFFICIENCY OF COMPUTERS

The book *Smart Machines* from IBM [26] argues persuasively that much of the value of big data will be lost because of

shortcomings in today's computers,³ and that, while some gains in efficiency may be achieved with optimisations in the design of present-day computers, radically new, brain-like kinds of computer will be needed that are "many orders of magnitude more energy efficient" than existing computers:

"The human brain is a marvel. A mere 20 watts of energy are required to power the 22 billion neurons in a brain that's roughly the size of a grapefruit.⁴ To field a conventional computer with comparable cognitive capacity would require gigawatts of electricity and a machine the size of a football field. So clearly something has to change fundamentally in computing for sensing machines to help us make use of the millions of hours of video, billions of photographs, and countless sensory signals that surround us. . . . Unless we can make computers many orders of magnitude more energy efficient, we're not going to be able to use them extensively as our intelligent assistants." [26, pp. 75, 76, and 88].

The two architectures outlined in Section II-D are, together, a fundamentally new approach to the design of computers that draw extensively on research in human cognition and neuroscience, with clear potential to make computers "many orders of magnitude more energy efficient" and very much smaller, as described in [74, Sec. III] (see also [75, Sec. IX] and Sec. X-A). It appears that there are very few alternatives with that kind of potential.

In brief, the arguments are as follows. Since computation in the SP framework is largely a matter of searching for patterns that match each other, gains in efficiency may be achieved by increasing the efficiency of search. This can be done in four main ways that are not mutually exclusive:

- Compression of information, which is at the heart of the SP system, reduces the size of the data to be searched.
- Potentially very large gains in efficiency may be achieved by concentrating searches in areas where success is most likely, exploiting the statistical information that the SP system gathers as a by-product of how it works (Appendix I-F).
- With SP-neural (Appendix I-G), there is potential to cut out a lot of searching altogether by making direct connections to structures that are already known (somewhat like making URL links to known web pages in a list of bookmarks or favourites), instead of using search for everything, whether or not it is already known (rather like using an internet search engine to find web pages

on all occasions, without taking advantage of shortcuts in a list of bookmarks or favourites).

- There is also potential to integrate processing and data in the manner of 'datacentric' computing [26, Ch. 5].

The potential of the SP system to make deep cuts in the energy consumption of computers contrasts sharply with the large computational resources used by artificial neural networks with deep learning (Section V-E).

V. DEEP LEARNING IN NEURAL NETWORKS

This section, about deep learning (DL) in artificial neural networks (ANNs), draws extensively on a review by Schmidhuber [52], who has achievements and long experience in the field.

Without in any way wishing to diminish the undoubted successes of DL in ANNs (both of which, together, will be referred to as NNs for short), the aim here is to highlight potential advantages of the SP system. This may seem unduly presumptuous since the SP system, unlike some NNs, has not won any competitions and has not been adopted or promoted by any company or incorporated in any products. But for reasons given in the subsections that follow, it appears that the SP system is built on firmer foundations than the current generation of NNs, and its long-term prospects are better.

The great variety of NNs makes it difficult to say things that are true of all of them. For that reason, the subsections that follow attempt to say things that are at least true of the majority.

A. SCOPE FOR ADAPTATION

There is a superficial resemblance between NNs and multiple alignments (especially if the latter are realized as SP-neural, as outlined in Appendix I-G) because they both have layers or levels and they both have connections between the levels. But NNs are not multiple alignments and provide much less scope for adaptation:

- Standardly, the number of layers of an NN and the size of each layer are pre-defined, whereas the number of columns or rows in a multiple alignment, and their sizes, depend entirely on the incoming and stored information from which it is built.⁵
- Normally, there is just one set of layers in an NN with a structure that is fixed, although its behaviour may be changed via the creation of new links within the structure, or changes in the strengths of links. By contrast, the SP system works by building what is normally a great diversity multiple alignments, each one created by drawing patterns from what is normally quite a small pool of New patterns and what may be a very large pool of Old patterns.

³In this connection, it has been reported ("Big data: 20 mind-boggling facts everyone must read", *Forbes*, 2015-09-30, <http://onforb.es/1YOdT2K>) that only about 0.5% of all data is ever analysed.

⁴As noted in [74], the brain's power consumption is probably about 12.6 watts ("Does thinking really hard burn more calories?", *Scientific American*, 2012-07-18, bit.ly/1qJmCBG), and it may contain as many as 86 billion neurons [18].

⁵But in the 'Group Method of Data Handling' (GMDH), the number of layers and the number of neurons in each layer depend on the problem being solved [52, Sec. 5.3]. However, it is evident from a review of research in this area [23] that NNs of this type are quite different from multiple alignments in the SP system, and they appear to be much less versatile and adaptable.

- Standardly, any given layer in an NN connects only with the layer immediately above (if any) and immediately below (if any).⁶ By contrast, any given column or row in a multiple alignment may have connections with any other column or row, depending on what the multiple alignment represents.
- Perhaps most importantly, learning in the SP system (Appendix I-E3) is quite different from gradualist styles of learning in an NN (Section V-D). Instead of varying the strengths of links between neurons in a pre-defined structure, the SP system learns by creating Old patterns, which may be derived directly from New patterns or, more commonly, from multiple alignments containing New and Old patterns ([70, Ch. 9 and Sec. 3.9.2], [72, Sec. 5]). There is potential for the creation of large numbers of different Old patterns, with a corresponding potential for the learning of diverse kinds of knowledge and skills.

B. BIOLOGICAL VALIDITY

It is generally recognised that NNs are only vaguely related to biological systems. For example: “‘Neural networks are actually very loosely inspired by the brain,’ says Oren Etzioni, CEO of the Allen Institute for Artificial Intelligence in Seattle. ‘They are distributed computing elements, but they’re very simple as compared with neurons; the connections are extremely simple as compared with a synapse.’”⁷

Although there are still big gaps in our knowledge about neural structures in the brain, and their functions, it appears that the organisation and workings of the SP system is better supported by available evidence:

- *Models of language learning.* As noted in Section II-K, the SP programme of research derives largely from earlier research developing computer models of language learning in children, drawing extensively on relevant empirical evidence.
- *Avoidance of over- or under-fitting.* A related point is that the solution to the problem of over-generalisation and under-generalisation that has been developed in research on language learning is likely to provide a better solution to the problems of over-fitting and under-fitting in NNs than other proposals in that area (Section V-H).
- *Cell assemblies.* In SP-neural (Appendix I-G, [70, Ch. 11]), abstract concepts in the SP theory map neatly into structures that are quite similar to Hebb’s [17] concept of a “cell assembly”, itself derived from neurophysiological evidence.
- *One-trial learning.* By contrast with gradualist styles of learning in NNs, the way in which unsupervised learning is done in the SP system provides an explanation for the phenomenon of one-trial learning and, at the same time,

explains why it takes time to learn complex knowledge or skills (Section V-D).

- *Grandmother cells.* In SP-neural, a concept such as one’s grandmother would be represented by a ‘pattern assembly’—the neural equivalent of an SP pattern. By contrast, the organisation and workings of NNs suggest that the neural representation of a concept such as one’s grandmother would normally be distributed across many widely-dispersed neurons and many connections amongst neurons (see also Section V-J1). On balance, available evidence suggests that grandmother cells, or something like them, do exist in mammalian brains: 1) it is sometimes suggested that the concept of a grandmother cell or cells is implausible because death of the cell or cells would mean that one could no longer recognise one’s grandmother. But that is exactly the kind of thing that can happen with people who have suffered a stroke or are suffering from dementia; and 2) there is relatively direct neurophysiological evidence for the existence of grandmother cells in the brain [16].

Although the ‘simple’ and ‘complex’ types of neuron discovered by Hubel [20] appear to have provided some inspiration for DL concepts [52, Sec. 5.2], they may also be seen to provide empirical support for hierarchical structures in the SP system, especially SP-neural (Appendix I-G).

C. LEARNING PARADIGMS

Supervised learning, unsupervised learning, and reinforcement learning—three forms of learning with NNs—are, in that connection, normally treated as alternatives with equal status.⁸

By contrast, in the SP perspective, unsupervised learning is seen as a foundation for all other forms of learning, including such things as learning by being told and learning by imitation [74, Secs. V-A.1, V-A.2, and V-J], and the learning of minor and major skills [74, Secs. V-G to V-I]. The main reasons are, in brief, that:

- As a matter of ordinary observation, much learning occurs without the benefit of labelled examples, help from a teacher, or carrots and sticks⁹: “Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object.” [30, p. 442].
- Extraction of redundancy from data, which is central in the SP theory, may be seen to operate not only in unsupervised learning but also in supervised learning—where there is redundancy in the associations between labels and corresponding examples—and in reinforcement learning—where there is redundancy in

⁸Although Schmidhuber acknowledges that unsupervised learning may facilitate supervised learning and reinforcement learning [52, Sec. 4.2].

⁹With regard to the last point, it is clear that motivations have an influence on learning—we tend to learn things best if they interest us and if we give them our attention. But, contrary to the central dogma of Skinnerian learning theory, it is unlikely that motivations are fundamental in learning.

⁶An exception would be a fully-recurrent neural network.

⁷“A robot finds its way using artificial ‘GPS’ brain cells”, *MIT Technology Review*, 2015-10-19, <http://bit.ly/1RVwsxc>.

the associations between actions and corresponding rewards or punishments.

Overall, the SP system promises to provide a unifying framework for learning of all kinds, potentially more satisfactory than when different forms of learning are treated separately.

D. LEARNING FROM A SINGLE OCCURRENCE OR EXPERIENCE

Most NNs incorporate some variant of the idea, proposed by Hebb [17] and known as ‘Hebbian’ learning, that neurons that repeatedly fire at about the same time will tend to become connected, or for existing connections between them to be gradually strengthened.

This kind of ‘gradualist’ mechanism for learning leads to slow changes in the behaviour of NNs, in keeping with the observation that it normally takes time to learn things like how to talk, or how to play the piano. That correspondence between the workings of NNs and a familiar feature of how we learn may strengthen the belief that NNs are psychologically valid.

But this feature of NNs conflicts with the undoubted fact that we can and often do learn things from a single occurrence or experience. Getting burned once will teach us to be careful with fire. We may retain memories for many years of significant events in our lives that occurred only once. And we may recognise a face that we have seen only briefly, we may recognise music that we have heard only once before, and likewise for films. It is true that we may rehearse things mentally but often it seems that, with little or no rehearsal, we remember things that have been seen or heard only once.

Because the slow strengthening of links between neurons does not account for our ability to remember things after a single exposure, Hebb adopted a ‘reverberatory’ theory for this kind of memory [17, p. 62]. But, as Milner has pointed out [39], it is difficult to understand how this kind of mechanism could explain our ability to assimilate a previously-unseen telephone number: for each digit in the number, its pre-established cell assembly may reverberate, but this does not explain memory for the *sequence* of digits in the number. We may add that it is unclear how the proposed mechanism would encode a phone number in which one or more of the digits is repeated.

The SP theory provides an explanation, both for learning from a single experience, and for the fact that some kinds of learning are slow:

- *Learning from a single experience.* Learning in the SP system (sketched in Appendices I-A and I-E3) starts by assimilating New information directly, followed by a possible encoding of the information in terms of any existing Old patterns, and the creation of newly-minted Old patterns via information compression [70, Secs. 3.9.2 and 9.2.2], [72, Sec. 5.1]. The taking in of New information, with or without its encoding in terms of existing Old patterns, means that the system can

learn from a single exposure to a pattern or event, much like an electronic recording system. This is essentially what is sometimes called ‘episodic memory’.

- *Slow learning of complex knowledge or skills.* Learning something like a natural language is much more complicated than remembering one’s first day at school or when one had a ride on a camel. With the learning of complex knowledge or skills, the main challenge is heuristic search through the vast abstract space of possible knowledge structures to find one or two that are reasonably good (Appendix I-E4). The learning of this kind of knowledge—sometimes called ‘semantic memory’—is necessarily a gradual process.

E. COMPUTATIONAL RESOURCES, SPEED OF LEARNING, AND VOLUMES OF DATA

In addition to apparent problems with learning from a single experience, there seem to be related issues with NNs concerning the computational resources and volumes of data they require for learning, and their speed of learning. For example:

- One news report¹⁰ describes how an NN with “16,000 computer processors” and “one billion connections” was exposed to “10 million randomly selected YouTube video thumbnails”, “over the course of three days”. Then, “after being presented with a list of 20,000 different items”, it began to recognize pictures of cats.
- Another report [9] refers to “... billions or even hundreds of billions of connections that have to be processed for every image” and “Training such a large network requires quadrillions of floating point operations ...”.
- And “... the new millennium brought a DL breakthrough in [the] form of cheap, multiprocessor graphics cards or GPUs. ... GPUs excel at the fast matrix and vector multiplications required ... for NN training, where they can speed up learning by a factor of 50 and more.” [52, Sec. 4.5].

Although it is true that it takes time for a person to learn his or her native language or languages (Section V-D) and the human brain contains billions of neurons, the current generation of NNs appears to overlook what can be achieved with ICMUP (Section II-E and Appendix I-E1), with small amounts of data, and quite modest computational resources:

- In accordance with the theory developed by Marr and Poggio [37] in which ICMUP may be seen to play a central role, a computer program developed by Grimson could discover the hidden image in a random-dot stereogram ([25], [73, Sec. 5.1]) with performance on a late-1970s computer that “coincides well with that of human subjects” [15, Sec. 5]. Although Grimson does not give run times, it looks as if his program finds the hidden image in a random-dot stereogram about as fast as people—normally in less than a minute.

¹⁰“Google’s artificial brain learns to find cat videos”, *Wired*, 2012-06-26, <http://wrd.cm/18YaV5I>.

- With run times of only a few minutes on a PC, the SP computer model, founded on ICMUP, demonstrates unsupervised learning of plausible generative grammars for the syntax of English-like artificial languages (Section XI-C). Similar results have been obtained with earlier models of language learning [68], also founded on ICMUP.
- In a similar vein, Joshua Tenenbaum of MIT has been quoted as saying “With all the progress in machine learning, it’s amazing what you can do with lots of data and faster computers, . . . But when you look at children, it’s amazing what they can learn from very little data. . . .”¹¹

It is true that what is being attempted with many NNs is relatively ambitious, but we should not forget that biological neurons are very much slower than electronic components. On balance, there appear to be problems with NNs relating to the computational efficiency of their learning and the volumes of data they require to obtain useful results.

Although researchers with NNs do not normally record the amount of energy required to run them, it is clear that, with the large computational resources and large volumes of data described in the quotes above, the current generation of NNs consume very much more energy than the *ca.* 13 watts of the human brain, and the trend appears to be upwards. NNs, as currently conceived, do nothing to solve the problems of energy consumption in computing outlined in Section IV. By contrast, there is potential, via information compression and the exploitation of statistical information which the SP system gathers as a by-product of how it works, for very substantial reductions in the amount of energy consumed in AI and, more generally, in computing (Section IV, [74, Sec. III]).

F. RECOGNITION OF IMAGES AND SPEECH

With some qualification (Sections V-E, V-G, and V-H), NNs do well in tasks such as the recognition of images (eg, [79]) or speech (eg, [4]). A particular strength of NNs in this connection is that they may be applied to ‘raw’ digitised data without the need for pre-processing to discover such features as edges or angles [30, p. 436]. But the SP system has potential for the discovery of such features, as discussed in [73, Sec. 3].

Notwithstanding the strengths of NNs just mentioned, it appears that the SP system provides a firmer foundation for the development of human-like capabilities in computer vision, and perhaps also in the processing of speech. As outlined in Section XI-A, the SP system has relative strengths and potential in several different aspects of pattern recognition and vision.

G. DEEP NEURAL NETWORKS ARE EASILY FOOLED

A recent report describes how “We can cause [a deep neural network] to misclassify an image by applying a certain hardly

perceptible perturbation.” [60, Abstract]. For example, the NN may correctly recognise a picture of a car but may fail to recognise another slightly different picture of a car which, to a person, looks almost identical (*ibid.*, Figure 6).

Another report [46] describes how one kind of deep neural network can be fooled quite easily into assigning an image with near certainty to a recognisable class of objects such as ‘guitar’ or ‘penguin’, when people judge the given image to be something like white noise on a TV screen or an abstract pattern containing nothing that resembles a guitar or a penguin or any other object.

Of course, these kinds of failures are a potential problem in any kind of application where recognition needs to be reliable. And without a good theory for how NNs work (Section V-M), they may be difficult to weed out.

With regard to the first kind of error—failing to recognise something that is almost identical to what has been recognised—there is already evidence that the SP computer model would not make that kind of mistake. It can recognise words containing errors of omission, commission and substitution (Section XI-A, [70, Sec. 6.2.1]), and likewise for diseases in medical diagnosis viewed as pattern recognition [69, Sec. 3.6] and in the parsing of natural language [72, Sec. 4.2.2].

No attempt has been made to test experimentally whether or not the SP computer model is prone to the second kind of error—recognising abstract patterns as ordinary objects—but a knowledge of how it works suggests that it would not be.

H. UNDER-GENERALISATION AND OVER-GENERALISATION

An issue with any learning system is its ability to generalise from the data (**I**) that is the basis of its learning, without under-generalisation (‘overfitting’) or over-generalisation (‘underfitting’). If, for example, the system has learned the concept ‘horse’, it should, in its later recognition of horses, not be too closely constrained to recognise only horses that are identical to or very similar to those in **I** (under-generalisation) and, at the same time, it should not make such mistakes as assigning cows, sheep or dogs to the category ‘horse’ (over-generalisation).

1) UNDER-GENERALISATION IN NNS

It is widely recognised that NNs may suffer from overfitting, and various solutions have been proposed. For example, Srivastava and colleagues [58] suggest that some neurons in an NN, together with their connections, may be randomly dropped from the NN during training, to prevent them co-adapting too much; while Zeng and colleagues [81] suggest that, in a multi-stage classifier, unsupervised pre-training and specially-designed stage-wise supervised training can help to avoid overfitting; and Wiesler and colleagues [66] say that they have found that a “factorized structure” can be effective against overfitting.

¹¹From “A learning advance in artificial intelligence rivals human abilities”, *New York Times*, 2015-12-10.

2) OVER-GENERALISATION IN NNS

The problem of underfitting in NNs has also drawn attention. For example, Dauphin and Bengio [5] show how underfitting may arise from the failure of some big neural networks to take full advantage of their computational capacity, and they make suggestions for overcoming the problem; and Ganin and Lempitsky [12] describe how a “two-stage architecture” can help overcome problems of underfitting.

3) ALMOST CERTAINLY, INFORMATION COMPRESSION SOLVES BOTH PROBLEMS

Without attempting a detailed comparison with alternative accounts of overfitting and underfitting in NNs, the suggestion here is that, in the SP theory, information compression provides a simple, elegant solution for both problems.

In connection with how children learn their first language or languages, information compression can explain how they learn to generalise correctly beyond the language that they have heard—not too little and not too much ([70, Sec. 9.5.3], [72, Sec. 5.3]). This includes correction of the over-generalisations (such as “hitted” or “gooses”) that are prominent in the early stages of a child’s learning of language. There is evidence that these things can be achieved without the need for correction by a ‘teacher’ or anything equivalent.

In brief, the argument depends on the idea that, when a body of information, **I**, is compressed, the result comprises a *grammar*, which we may refer to as **G**, and an *encoding* of **I** in terms of **G**, which we may refer to as **E** (Appendix I-E5). The two things together represent a lossless compression of **I**, without any generalisation. But **G** by itself represents the recurrent or redundant features of **I**, without the non-redundant parts of **I**. As such, it appears to provide for generalisation beyond **I**, without either over-generalisation or under-generalisation.

It seems likely that the same principles would apply to the learning of grammars for the recognition of such things as images or speech, thus solving two problems—the problems of overfitting and underfitting—with one over-arching principle.

The SP system may also help to solve the problem of overfitting in the way that it can recognise patterns via multiple alignment in the face of errors of omission, commission, and substitution (Section XI-A, [72, Sec. 4.2.2]).

I. INFORMATION COMPRESSION

Schmidhuber’s review [52] contains a short section (4.4) about “Occam’s Razor: compression and minimum description length (MDL)”, and it mentions information compression in some other sections. Although he suggests (in Section 5.10) that “much of machine learning is essentially about compression”, the overall thrust of the review is that information compression is merely one of several “recurring themes” in deep learning, without any great significance.

By contrast, information compression is fundamental in the SP theory, running through it like *Blackpool* in a stick of rock, in its foundations ([76], Appendix I-B), in the matching and unification of patterns (Appendix I-E1), in the building of multiple alignments (Appendix I-E2), and in unsupervised learning (Appendix I-E3).

In view of evidence for the importance of information compression in intelligence, computing, and mathematics (Appendix I-B), it appears that the peripheral status of information compression in the design and operation of NNs weakens them conceptually in comparison with the SP system.

J. TRANSPARENCY IN THE REPRESENTATION AND PROCESSING OF KNOWLEDGE

A problem with NNs is that there is considerably uncertainty about how they represent knowledge and how they process it:

“... we actually understand surprisingly little of why certain models work and others don’t. ... One of the challenges of neural networks is understanding what exactly goes on at each layer.” [41].

“... no one knows how neural networks come up with their answers. ... A programmer need adjust only the number of nodes and layers to optimise how it captures relevant features in the data. However, since it’s impossible to tell exactly how a neural network does what it does, this tweaking is a matter of trial and error.”¹²

With regard to the first quote, it is true that, as described in the blog, NNs can be made to reveal some of their knowledge. But, while many of the resulting images have artistic appeal, they are not transparent representations of knowledge (see also Section V-J2), it’s not clear how they are learned or how they function in such tasks as recognition, and they certainly do not conform to the DONSVIC principle (Section II-I).

By contrast with these uncertainties:

- In the SP system, all kinds of knowledge, including those detailed in Section XI-E, are represented transparently as SP patterns.
- In the SP computer model, an audit trail can be provided for all processing, including the building of multiple alignments and the creation of grammars.
- In the SP system, it is anticipated that unsupervised learning will conform to the DONSVIC principle (Section II-I, [72, Sec. 5.2]), and this is confirmed by evidence from the SP computer model. To the extent that this remains true in future versions of the model, structures created via unsupervised learning in the SP model are likely to be transparent and comprehensible by people.

¹²“The rapid rise of neural networks and why they’ll rule our world”, *New Scientist*, 2015-07-08, <http://bit.ly/1IkbbuC>.

1) DISTRIBUTED OR LOCALIST ENCODINGS

There is further uncertainty about whether knowledge in an NN is represented and processed in a ‘distributed’ or ‘localist’ scheme. The dominant view is that, in neural networks, knowledge of a concept such as one’s grandmother is encoded in neurons that are widely-distributed, with links between them (Section V-B). In this ‘sub-symbolic’ view, it would not be possible to identify any single neuron or local cluster of neurons that represent any given concept: “there is no distinction between individual high level units and random linear combinations of high level units, according to various methods of unit analysis”, suggesting that “it is the entire space of activations, rather than the individual units, that contains the bulk of the semantic information” [60, Abstract].

But some researchers suggest that “it is possible to train [artificial] neurons to be selective for high-level concepts ... In our experiments, we obtained neurons that function as detectors for faces, human bodies, and cat faces by training on random frames of YouTube videos.” [29, Conclusion]. This research appears to support the ‘localist’ or ‘symbolic’ view that a concept such as one’s ‘grandmother’ may be represented by a single neuron or, perhaps, a small cluster of neurons (*cf.* Section V-B).

Although the SP theory may turn out to be wrong empirically, it is at least clear on this issue. As noted in Section V-B, it is very much in the localist camp. In SP-neural, it is envisaged that knowledge of concepts in the form of SP patterns is recorded on the cortex very much like writing on a sheet of paper (Appendix I-G). As with writing on a sheet of paper, any one section or paragraph, in one location, may be seen to describe a concept, although it may contain pointers to supporting information elsewhere.

2) CLASS-INCLUSION HIERARCHIES AND PART-WHOLE HIERARCHIES

With regard to hierarchical structures, there is further uncertainty about whether NNs discover:

- Class-inclusion hierarchies:

“Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level.” [30, p. 436].

“Hidden layers: these learn more abstract representations as you head up.”¹³

- Or part-whole hierarchies:

“An image ... comes in the form of an array of pixel values, and the learned features in the first layer of representation typically represent the presence or absence of edges at particular

orientations and locations in the image. The second layer typically detects motifs by spotting particular arrangements of edges, regardless of small variations in the edge positions. The third layer may assemble motifs into larger combinations that correspond to parts of familiar objects, and subsequent layers would detect objects as combinations of these parts.” [30, p. 436].

“... the first layer maybe looks for edges or corners. Intermediate layers interpret the basic features to look for overall shapes or components, like a door or a leaf. The final few layers assemble those into complete interpretations ... such as entire buildings or trees.” [41].

and, either way, the representation is obscure.

While it is clear that a human face is part of a human body (Section V-J1), it seems reasonable to assume that, in either a distributed or localist scheme (Section V-J1), a concept such as ‘horse’, ‘cow’, or ‘sheep’, would be most fully encoded in the highest layer of an NN. If that is accepted, the question arises how the NN would encode knowledge of something like an agricultural exhibition which is likely to contain representations of all three of the concepts mentioned, and with multiple instances of each one of them. Likewise, we may ask how such concepts may be encoded as examples of abstractions such as ‘mammal’ or ‘vertebrate’. It is not clear how concepts that are most fully encoded in the top layer of an NN could ever be part of something that is larger or more abstract.

By contrast with these uncertainties with NNs about the representation of class-inclusion hierarchies or part-whole hierarchies, the SP system can represent both types of hierarchy, with a clear distinction between the two. This can be seen in the class hierarchies in Figure 1 and in [70, Fig. 6.7], and in the part-whole hierarchies in Figure 2 and in [70, Fig. 6.8]). At the same time it provides for seamless integration of both kinds of hierarchy, as illustrated in [72, Fig. 16]. Any concept, of any size or level of abstraction, may always be embedded in something larger or more abstract.

The SP system also provides the means of representing class-inclusion heterarchies (cross-classification) and part-whole heterarchies. It appears that this is well beyond anything that can be done with NNs.

Figure 1 shows how the SP computer model may recognise a car at multiple levels of abstraction. The figure shows the best multiple alignment created by the SP computer model—the one yielding most compression—with a New pattern (column 0) and a set of Old patterns representing categories related to cars (columns 1 to 5). Column 5 shows that the unknown entity is “my car”, and other columns show that it is a sports car (column 4), a car (column 1), and a vehicle (column 2). The order of the categories across the columns has no special significance—it depends purely on how the multiple alignment was built up.

Figure 2 shows how, in a similar way, the SP system may, via a multiple alignment, model the recognition of an

¹³Slide 3 in “Deep learning for NLP (without magic)”, slide show by R. Socher and C. Manning, dated 2013, <http://stanford.io/1bmBsKK>, retrieved 2015-07-25.

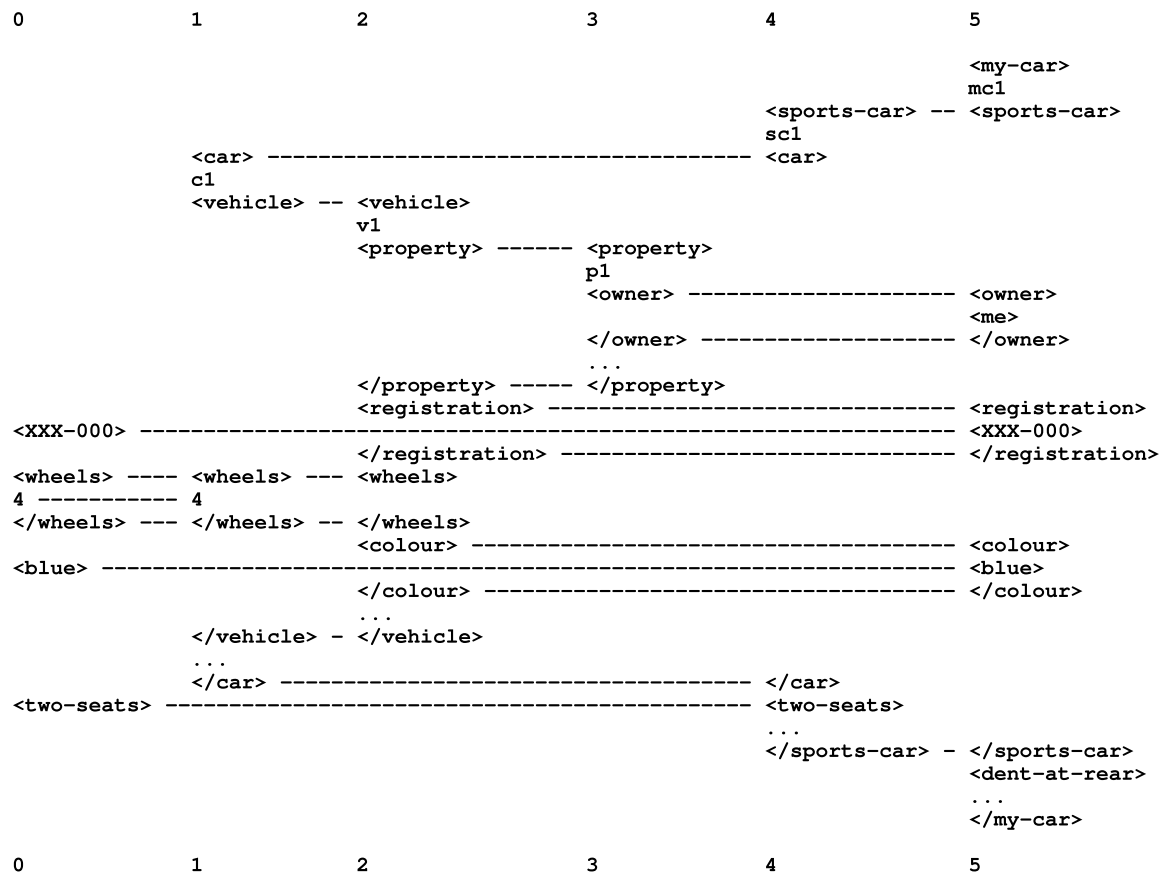


FIGURE 1. A multiple alignment illustrating recognition at multiple levels of abstraction, as described in the text.

unknown entity in terms of its parts and sub-parts. In this example, the features of the unknown entity are shown in column 0. That the entity is a car is shown in column 2, with a match for one of its features—a gearbox. Column 1 shows one part of a car—the engine—with a match for one of its parts—the crankshaft. And column 3 represents the body of the car with a match for one of its parts—the steering wheel.

3) ITERATION AND RECURSION

Related to issues just discussed are questions about how NNs may encode repeated instances of this or that category (such as the many instances of horses, cows, and sheep that one would expect to see at an agricultural exhibition (Section V-J2)), and how NNs may encode the kinds of recursive structures that are prominent in natural language, such as *This is the horse and the hound and the horn, That belonged to the farmer sowing his corn, That kept the rooster that crowed in the morn, . . .*

At first sight, one might think that ‘recurrent neural networks’ would be an appropriate means of modelling recursion in natural language. But, as noted in Section V-K1, these kinds of NNs do not, apparently, respect the word, phrase, clause and sentence structures in natural language that have been prominent in generative and computational linguistics for many years.

By contrast with uncertainties about how NNs may handle repeated structures, the SP theory provides crisp answers:

- Any given SP pattern may appear one, two, or more times within any one multiple alignment. One example is how the pattern representing a noun phrase (‘NP’) appears twice in the multiple alignment shown in Figure 10: once in column 2 and again in column 11. Another example is shown in Figure 3, below, where the pattern ‘ri’ appears in columns 5, 7, and 9.
- In the interpretation of any given body of data, there should be no difficulty arising from the occurrence of multiple instances of any given category. There is more detail in [70, Sec. 3.4.6] about how a given pattern may appear more than once in any multiple alignment.
- Multiple alignments can accommodate recursive structures, as illustrated in Figures 3 and 8.
- Recursion may also be accommodated in SP-neural, as illustrated in [70, Fig. 11.10].

With regard to Figure 3, native speakers of English know that, in a phrase like *the very . . . expensive jewellery*, the word *very* may, in principle, be repeated any number of times—although it would be considered poor style to use it more than 4 or 5 times. In the figure, this kind of recursion may be seen in repeated appearances of the self-referential pattern ‘ri ril i #i ri #ri #ri’.

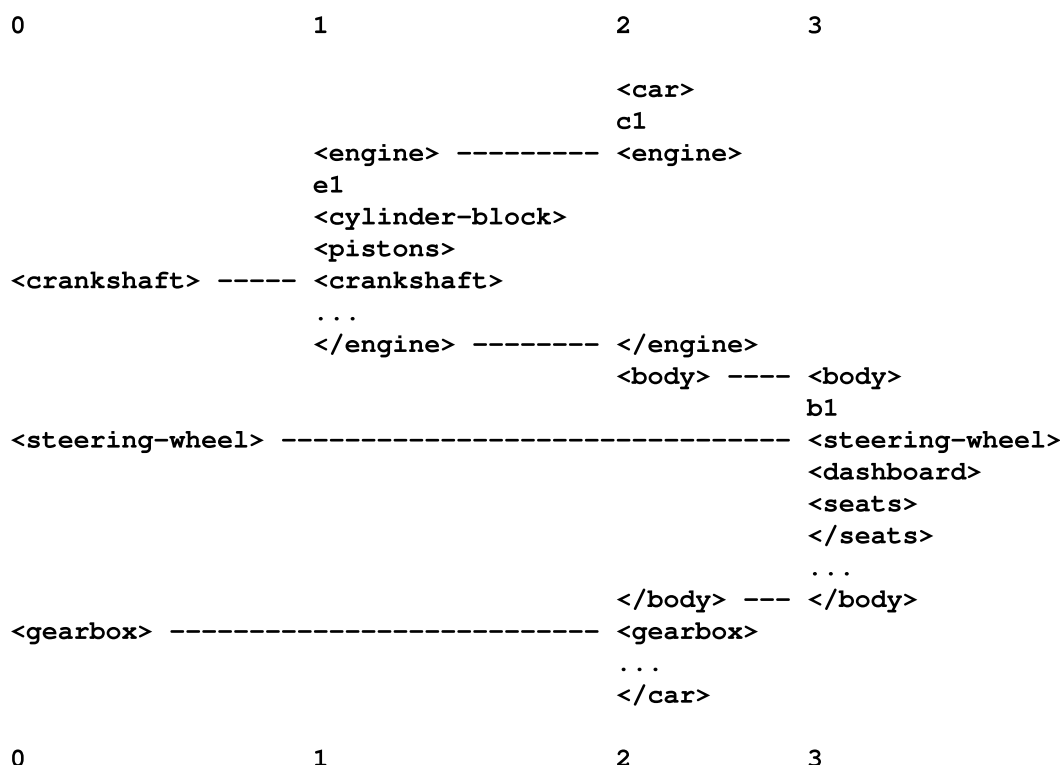


FIGURE 2. A multiple alignment illustrating recognition of a car via its parts and sub-parts, as described in the text.

It is self-referential because the symbols ‘ $r_i \dots \#r_i$ ’ at the beginning and end of the pattern can be aligned with the symbols ‘ $r_i \#r_i$ ’ in the body of another appearance of the pattern.

K. THE PROCESSING OF NATURAL LANGUAGE

A recent review of research in deep learning [30] suggests that “deep learning has produced extremely promising results for various tasks in natural language understanding, particularly topic classification, sentiment analysis, question answering and language translation.” (p. 436). This section first examines some research in the application of NNs to natural language, and then, in Section V-K5, assesses their prospects in this area in comparison with the strengths and potential of the SP system.

To anticipate a little, it seems that NNs may indeed produce some useful results in the processing of natural language but, taking a longer view, the SP system promises more precision and higher standards of performance. As one commentator has remarked: “Deep learning’s usefulness for speech recognition and image detection is beyond doubt, but it’s still just a guess that it will master language and transform our lives more radically.” [54, p. 73]. It seems unlikely that NNs will, for example, ever be able to represent and process the subtle structure of inter-locking dependencies in English auxiliary verbs, as the SP system can do ([70, Sec. 5.5], [72, Secs. 8.2 and 8.3]).

1) PARSING AND PRODUCTION OF NATURAL LANGUAGE

In addressing issues in the parsing of natural language by NNs, Socher and colleagues [55] employ the concept of ‘Compositional Vector Grammar’, but their results are not as clear or straightforward as one might wish, something which two of them (Socher and Manning) acknowledge in a slide show:

“*Concern: problem with model interpretability.* No discrete categories or words, everything is a continuous vector. We’d like have symbolic features like NP, VP, etc. and see why their combination makes sense.”¹⁴

Notwithstanding some qualifying remarks, their concern remains valid.

By contrast, the SP system has several strengths in both the analysis and production of natural language (Section XI-B). In particular, the system respects and fully represents the kinds of hierarchical structure that have long been recognised in generative linguistics, which are quite different from the kinds of cyclic representations of language sequences by recurrent neural networks described in, for example, [59, 64].

¹⁴Slide 201 in “Deep learning for NLP (without magic)”, slide show by R. Socher and C. Manning, dated 2013, <http://stanford.io/1bmBsKK>, retrieved 2015-07-27.

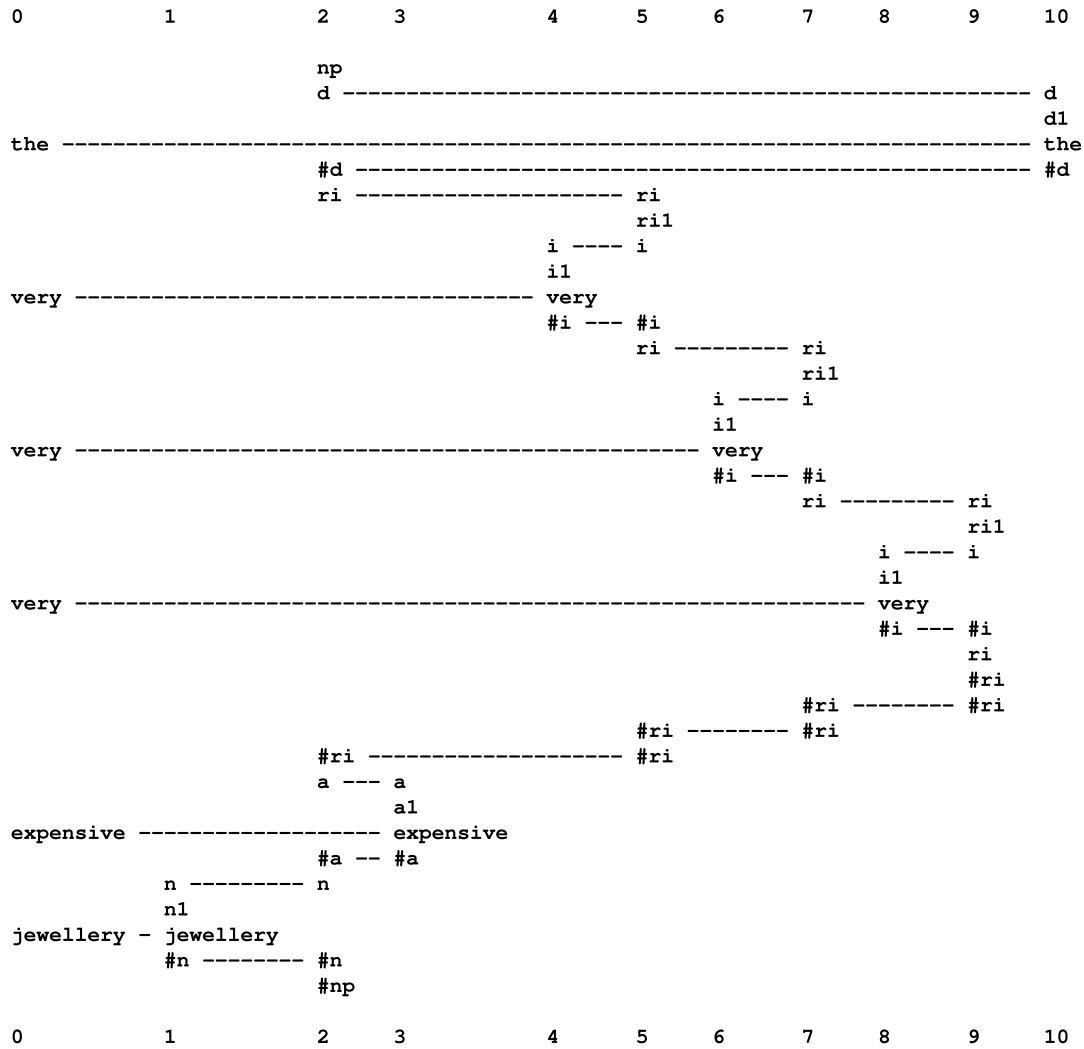


FIGURE 3. A multiple alignment from the SP computer model showing recursion, as described in the text.

2) TRANSLATION BETWEEN LANGUAGES

In some recent work (eg, [24, 53, 59]) NNs have been applied to the already-established technique of translating between languages by mapping phrases in the source language to corresponding phrases in the target language.

The main problem here is that, while phrase-to-phrase or syntax-to-syntax systems can produce useful translations, and may generalise usefully from the training data [53], it is unlikely that they will ever attain the standard of a good human translator. To reach that ‘Holy Grail’, it appears to be essential to translate the source language into a representation of its meaning, and then translate this ‘interlingua’ into the target language. This appears to be what people do, it is in any case needed for high-quality ‘understanding’ of natural language (Sections V-K3 and XI-B), and it can reduce substantially the number of mappings that are required.

In the quest for that higher standard of translation, it appears that the SP system has more potential than NNs: it has clear potential for the learning of the syntax of natural

language, for the learning of semantic structures, and for learning syntactic-semantic structures (Section XI-C).

3) UNDERSTANDING OF NATURAL LANGUAGE

Zhang and LeCun [82, Sec. 1] write: “In this article we show that text understanding can be handled by a deep learning system without artificially embedding knowledge about words, phrases, sentences or any other syntactic or semantic structures associated with a language.” But this claim needs to be qualified:

- In the examples that are given, what is learned is essentially associations between short pieces of text such as questions and answers in the *Yahoo! Answers* website. This notion of ‘understanding’ falls a long way short of, for example, the understanding that a motor engineer would have for an expression like “gear box”, or that a musician would have for an expression like “Handel’s Messiah”. For the great majority of ‘content’ words or phrases in any language, the full meaning is a rich

structure of knowledge, and, normally, most of it is not about words or combination of words.

- With the very simplistic concept of ‘understanding’ that is employed in [82], it may indeed be possible to find associations between words and ‘meanings’ without worrying about the syntactic and semantic structure of language. But for a thorough understanding of, for example, a legal, philosophical or scientific argument, it seems unlikely that we can by-pass the need for a good knowledge of such structures.

What is perhaps a better example is described in [64], where NN techniques for syntax-to-syntax translation are borrowed and converted to an image-to-syntax scheme, where “image” means the output of an NN that recognises images and “syntax” means a natural language caption for each image. In this case, it is perhaps more reasonable to describe each image as the ‘meaning’ of the corresponding caption. But we would still be a long way from the richly-structured forms of knowledge which provide most of the meanings for natural language.

In summary, it seems that, compared with NNs, the SP system provides a much sounder basis for learning those rich, complex structures of knowledge which are the meanings of natural language and which need to be in place if we are to attain anything like human-level ‘understanding’ of natural language. It seems unlikely that NNs will ever be able to meet the “The Winograd Schema Challenge” [31].¹⁵

4) UNSUPERVISED LEARNING OF NATURAL LANGUAGE

Zhang and LeCun [82, Sec. 1] also write: “... we hypothesize that when trained from raw characters, temporal ConvNet is able to learn the hierarchical representations of words, phrases and sentences in order to understand text.” (Section 1). In view of the problems of transparency with NNs (Section V-J), uncertainties about distributed or localist encodings (Section V-J1), uncertainties about class-inclusion hierarchies and part-whole hierarchies (Section V-J2), and the failure of NNs to respect the DONSVIC principle (Section V-J, introduction), this hypothesis appears to be over-optimistic.

By contrast, the SP system has already demonstrated unsupervised learning of segmental structures, classes of structure, and abstract patterns, and it has good prospects for further development (Section XI-C).

5) DISCUSSION

There is little doubt that, in natural language processing, NNs will yield some useful new applications on short to

¹⁵Winograd’s original example [67, p. 33] demonstrates the subtlety of natural language with two sentences: *The city councilmen refused the demonstrators a permit because they feared violence* and *The city councilmen refused the demonstrators a permit because they advocated revolution*. In the first case, we understand that the word *they* refers to the city councilmen, while in the second case it seems more reasonable to assume that *they* is a reference to the demonstrators. Our understandings in the two cases depend on extensive knowledge of how things happen in the world and the psychology of people, including city councilmen and demonstrators.

medium timescales. But, because of the several problems with NNs, described elsewhere in Section V, it is unlikely that they can match human capabilities with natural language. In particular, they are unlikely to make much impact on the difficult problems that need to be solved to reach that standard: learning syntactic structures, learning the kinds of knowledge structures that provide the meanings for language, and learning syntactic-semantic connections.

It appears that, taking a long view, the SP system provides a much sounder basis for the understanding, production, translation, and learning of natural language, with firmer theoretical foundations.

L. REASONING, PROGRAMMING, AND OTHER KINDS OF ‘SYMBOLIC’ PROCESSING

“As regards knowledge representation in the brain, one of the key challenges is to understand how neural activations, which are widely distributed and sub-symbolic, give rise to behavior that is symbolic, such as language and logical reasoning.” [6, Sec. 2].

“[Deep learning techniques] have no obvious ways of performing logical inferences . . .”¹⁶

This section briefly reviews aspects of AI and other areas of computing which seem to pose problems for NNs, and where the SP system is relatively strong. In broad terms, these seem to be areas where the ‘symbolic’ tradition has proved to be relatively successful.

Some years ago, there was quite a lively interest in how NNs might do reasoning (eg, [34]) but that interest seems to have subsided, probably because NNs really are not well suited to much more than the relatively simple kinds of inference that, in pattern recognition, correct errors of omission, commission, or substitution.

By contrast, the SP system demonstrates several kinds of reasoning, with clear potential for further development (Section XI-D).

Similar things may be said about areas of AI which appear to be problematic for NNs and where symbolic AI and the SP system are relatively strong, such as planning (see Figure 4, below, and [70, Ch. 8]), problem solving ([70, Ch. 8]), and unsupervised learning of language (Section XI-C, [70, Ch. 9], [72, Sec. 5]).

Again, it is difficult to envisage how NNs might be used for the kind of ‘programming’ which is the mainstay of software engineering as it is practiced now. By contrast, the SP system has clear potential for application in that area (Section XI-F).

There seem to be three main reasons for the relative strength of the SP system in the areas mentioned:

- Unlike NNs, there is transparency in both the representation and processing of knowledge (Sections II-H and V-J).

¹⁶“Is ‘deep learning’ a revolution in artificial intelligence?”, *The New Yorker*, Gary Marcus, 2012-11-25, <http://nry.kr/1Be7S22>.

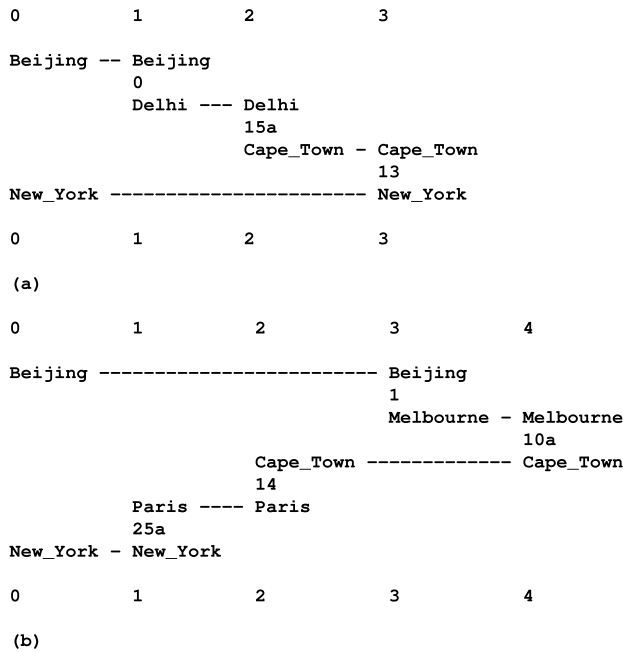


FIGURE 4. Two multiple alignments produced by the SP computer model showing how, with appropriate data, the system may plot alternative routes between two cities, as described in the text. This figure has been adapted, with permission, from [70, Figs. 8.1 and 8.5].

- Within the multiple alignment framework, it is possible to model concepts from mainstream computing and the symbolic tradition of AI such as *variable*, *value*, and *type* (Section XI-D).
- The multiple alignment framework is much more adaptable than the deep learning framework (Section V-A).

As an example of the kind of thing that the SP system can do, Figure 4 shows how multiple alignments produced by the SP computer model may serve to plot alternative routes between two cities. In each of the two multiple alignments, the two cities, Beijing and New York in this example, are shown as a New pattern in column 0. Each of the remaining columns in each multiple alignment shows an Old pattern, drawn from a pool of such patterns, showing a one-stop flying route between another pair of cities.

M. THEORETICAL FOUNDATIONS

Several things point to deep problems with NNs—theoretical foundations that are weak or absent. These include: doubts about their biological validity (Section V-B), failure to account for learning from a single experience (Section V-D), concerns about the computational resources required by NNs (Section V-E), failures in recognition (Section V-G), weaknesses in under-generalisation and over-generalisation (Section V-H), problems of transparency in the representation and processing of knowledge (Section V-J), problems with the processing of natural language (Section V-K), and problems in modelling reasoning and other ‘symbolic’ kinds of processing (Section V-L).

That there are deficiencies in theory for NNs is also suggested by their many different versions: ‘time-delay’ neural networks, ‘gradient-based deep learners with alternating convolutional and down-sampling layers’, ‘weight-sharing feed-forward’, ‘nonlinear auto-regressive with exogenous inputs recurrent’, ‘max-pooling convolutional’, ‘multi-column GPU max-pooling convolutional’, ‘bi-directional long short-term memory recurrent’, and more [52]. Although these many versions indicate high levels of interest in NNs, they are also an indication that researchers are still not happy with the NN framework.

VI. UNIFIED THEORIES OF COGNITION AND RELATED RESEARCH

As mentioned in Appendix II, ‘unified theories of cognition’ and ‘artificial general intelligence’ are honourable exceptions to the rather marked tendency for research in AI and in human perception and cognition to have become fragmented into many subfields, without much communication between them. This section will not attempt to review all the projects in these and related areas. It will say something briefly about two of them: Soar and CogPrime.

The first of these, Soar [28], was originally developed as a response to Allen Newell’s call for ‘unified theories of cognition’, in his book of the same name [44]. Up to and including Soar 8 (*ca.* 2004), “... there has been a single framework for all tasks and subtasks (problem spaces), a single representation of permanent knowledge (productions), a single representation of temporary knowledge (objects with attributes and values), a single mechanism for generating goals (automatic subgoal), and a single learning mechanism (chunking)”, in accordance with the principle that “... the number of distinct architectural mechanisms should be minimized.”¹⁷

But in later versions of Soar, the stress on “unified” seems to have diminished: “We have revisited [the assumption that architectural mechanisms should be minimised] as we attempt to ensure that all available knowledge can be captured at runtime without disrupting task performance. This is leading to multiple learning mechanisms (chunking, reinforcement learning, episodic learning, and semantic learning), and multiple representations of long-term knowledge (productions for procedural knowledge, semantic memory, and episodic memory)” (*ibid.*).

Although Newell uses the expression “cognitive architecture” [44, p. xii and elsewhere], an increased willingness now for Soar to encompass multiple structures and multiple processes appears to be the reason for an increased emphasis on the concept of a cognitive architecture rather than that of a unified theory of cognition.

The variety of structures and mechanisms in the early Soar, and their expansion in the later Soar, suggests that, despite the commendable early quest for a unified model of cognition,

¹⁷From the home page of the official Soar website at <http://bit.ly/1NXnGP9>, retrieved 2015-09-02.

the system has been and is a kluge (Appendix II-D). It appears that the SP system scores better in terms of simplicity and power (Appendix II-B), doing at least as much but with significantly less structure or mechanism.

Similar things may be said about CogPrime [13]. To be fair, it does not aspire to be a unified theory of cognition or intelligence, since “We have not discovered any one algorithm or approach capable of yielding the emergence of [intelligence].” (*ibid.*, p. 2). Rather, it is an “architecture” for artificial general intelligence containing a variety of structures and mechanisms, with “*cognitive synergy*: the fitting together of different intelligent components into an appropriate cognitive architecture, in such a way that the components richly and dynamically support and assist each other ...” (p. 2, emphasis in the original), based on the hypothesis that such synergy “can yield robust intelligence at the human level and ultimately beyond.” (p. 3).

In effect, the argument is that, although CogPrime is a kluge, “cognitive synergy”, also described as “secret sauce” (p. 2), will somehow overcome that handicap. Although [13] contains outline examples of how cognitive synergy may be applied, it is difficult to see how, via such applications, CogPrime’s relatively complex collection of structures and mechanisms might yield “robust intelligence”, or achieve the overall simplicity that one would wish to see in any good theory.

In [13], there is a welcome recognition of the importance of Occam’s Razor, which is described as “an overall design principle that underlies nearly every part of the [CogPrime] system.” (p. 36). But, from an SP perspective, this falls far short of what is needed: it does not recognise the importance of ICMUP (Section II-E) or multiple alignment (Section II-F), and it does not recognise how, via those two principles, one may achieve an overall simplification in theory (Section II-A), and an overall simplification in computing systems (Section II-B), cutting out much of the complexity that arises when a collection of diverse concepts and mechanisms, from different areas, are bolted together, with or without “secret sauce”.

It seems that researchers in AI have largely given up on the quest for a truly unified theory of cognition. The nearest thing seems to be a relatively large number of ‘cognitive architectures’ (many of which, up to 2010, are reviewed in [50]). It appears that the SP system is unique in its attempt to simplify and integrate concepts across a broad canvass. As such, it appears to be unusually strong in combining conceptual simplicity with descriptive or explanatory power (Section II-A).

VII. UNIVERSAL SEARCH

Some ideas, that may be grouped together loosely under the heading ‘universal search’, seem, at first sight, to offer comprehensive solutions to problems in AI and beyond.

Solomonoff [57] has argued that the great majority of problems in science and mathematics may be seen as either ‘machine inversion’ problems or ‘time limited optimization’

problems, and that both kinds of problem can be solved by inductive inference using the principle of minimum length encoding.

In ‘Levin search’ [32], [33], which aims to solve inversion problems, the system is designed to solve any given problem by interleaving all possible programs on a universal Turing machine, sharing computation time equally among them, until one of the executed programs manages to solve the target problem.

Ideas of this kind have been developed by Hutter (eg, [21]), Schmidhuber (eg, [51]), and others.¹⁸ From the perspective of the SP research programme, the main sources of concern are:

- The apparent difficulty of translating the abstract concepts of universal search into working models that exhibit aspects of intelligence or are potentially useful.
- The apparent difficulty of squeezing the subtlety and complexity of human intelligence into the procrustean bed of ‘well defined problems’ [21]—something that appears to be a prerequisite for universal search.
- With problems in AI, it is rarely possible to guarantee solutions that are theoretically ideal—a focus of interest in research on universal search. Normally, via heuristic search, we should aim for solutions that are ‘reasonably good’ and not necessarily perfect (Appendix I-E4).
- In terms of the trade-off between simplicity and power (Appendix II), it appears that theories in the area of universal search are running the risk of being too simple and over-general, and correspondingly weak in terms of descriptive or explanatory power.

By contrast, the SP system provides mechanisms for finding good full and partial matches between patterns (Appendix I-E1), for building multiple alignments (Appendix I-E2), and for creating grammars (Appendix I-E3), and it has plenty to say about a range of observations and concepts in AI and beyond.

VIII. BAYESIAN NETWORKS AND SOME OTHER MODELS FOR AI

This section considers briefly some other systems or techniques that have been proposed as models for AI, or aspects of AI: Bayesian networks, support vector machines, hidden Markov models, Kalman filters, self-organising maps, Petri nets, cellular automata, pushdown automata, dimensionality reduction, and finite-state automata. It is likely that there are other systems to which the remarks in this section apply.

The suggestion here, which is admittedly a rather sweeping generalisation, is that, while these models are admirably simple, and while they may have useful applications, they lack the descriptive and explanatory range of the SP system. In terms of the quest for a favourable combination of simplicity and power (Appendix II and Section II-A), their descriptive and explanatory range is too limited.

With appropriate data, the SP system provides an alternative to Bayesian networks in modelling such phenomena as

¹⁸For a useful overview, see “Universal search” by Matteo Gagliolo [11].

‘explaining away’ ([72, Sec. 10.2], [70, Sec. 7.8]). In addition to its relative strength in terms of simplicity and power, the SP system, compared with a Bayesian network, has three apparent advantages:

- *Simplicity in the representation of statistical knowledge.* Each node in a Bayesian network contains a table of conditional probabilities for possible combinations of inputs, and these tables can be quite large. By contrast, the SP framework only requires a single measure of frequency for each pattern. Absolute or conditional probabilities can be derived from that frequency measure, as required.
- *Objectivity.* Arguably, the SP system has an advantage compared with Bayesian networks because it derives ‘objective’ probabilities in a straightforward way from frequencies of occurrence, and it eliminates the subjectivity in Bayes’ theorem in its concept of probability as a degree of belief.
- *Creation of ontologies.* Bayes’ theorem assumes that the categories that are to be related to each other via conditional probabilities are already established. By contrast, the SP system provides an account of how a knowledge of categories and entities may be developed via unsupervised learning (Appendix I-E3).

IX. IBM’S WATSON

As is now well known, a team of researchers at IBM developed a computing system, called Watson, that, in 2011, beat the best human players at the TV quiz game *Jeopardy!*¹⁹

Of course winning at *Jeopardy!* is a major achievement, with potential benefits in terms of ideas and, perhaps, applications. But there are concerns about aspects of Watson and its development, described in the following subsections.

A. SIGNIFICANCE OF WATSON FOR AI

Doubt has been expressed about the significance of Watson for AI:

“... systems that seem to have mastered complex language tasks, such as IBM’s *Jeopardy!* winner Watson, do it by being super-specialized to a particular format. ‘It’s cute as a demonstration, but not work that would really translate to any other situation,’ [says Yann LeCun].” [54, p. 73].

B. IGNORING RESEARCH ON HUMAN PERCEPTION AND COGNITION

There is concern about how, in the development of Watson, research on human perception and cognition was ignored. Dave Ferrucci, leader of the team that developed Watson, has been quoted as saying:

“Did we sit down when we built Watson and try to model human cognition? Absolutely not. We

just tried to create a machine that could win at *Jeopardy!*”²⁰

An implication of this remark is that, by ignoring the large body of research into human cognition and the many insights that have been gained, the developers of Watson have probably reduced the chances of Watson providing a meaningful route to ‘cognitive computing’ as described in [26].

C. WATSON AS A KLUGE

The original Watson was created by combining natural language understanding with statistical analysis of very large amounts of text. IBM has now added capabilities in translation, speech-to-text, and text-to-speech, and they plan to add capabilities in deep learning,²¹ with large numbers of medical images as data for learning.²²

Further, Watson was created as a “holistic combination of many diverse algorithmic techniques” [10, p. 2] or “hundreds of different cooperating algorithms” [10, pp. 2 and 3] which are, as Ferrucci remarks, reminiscent of Minsky’s *Society of Mind* [40].

In short, Watson was originally developed, and is continuing to be developed, as a kluge: “a clumsy or inelegant—yet surprisingly effective—solution to a problem.” [36, p. 2].

Does this matter? There are several possible answers to this question. On the plus side (the kluge is a good thing):

- It is probably possible to create a system that does useful things and earns money.
- Combining technologies may help to overcome weaknesses in individual technologies and it may help to overcome fragmentation in AI. (“If deep learning can be combined with other AI techniques effectively, that could produce more rounded, useful systems.”²³)
- Since the human mind is a kluge [36], it should not matter if AI systems are the same (but see Appendix II-D).

But on the minus side:

- Creating a kluge may yield short-term gains but is unlikely to be satisfactory on longer timescales [77, Secs. 2, 6, and 7].
- Creating a kluge may be a distraction from the long-term goal of developing ‘cognitive computing’:

“The creation of this new era of [cognitive] computing is a monumental endeavor ... no company can take on this challenge alone. So we look to our clients, university researchers, government policy makers, industry partners, and entrepreneurs—indeed the entire tech industry—to take this journey with us.” [26, Preface].

²⁰“The man who would teach machines to think”, *The Atlantic*, November 2013, <http://theatlantic.com/1fi9AFv>.

²¹“IBM pushes deep learning with a Watson upgrade”, *MIT Technology Review*, 2015-07-09, <http://bit.ly/1Nq0bMg>.

²²“Why IBM just bought billions of medical images for Watson to look at”, *MIT Technology Review*, 2015-08-11, <http://bit.ly/1P6lcVQ>.

²³“IBM pushes deep learning with a Watson upgrade”, *MIT Technology Review*, 2015-07-09, <http://bit.ly/1Nq0bMg>.

¹⁹See, for example, “Watson (computer)”, *Wikipedia*, <http://bit.ly/1DwVKiC>, retrieved 2015-08-12.

On balance, a two-pronged strategy is probably best: take advantage of short-term gains that may accrue from developing applications such as Watson as kluges and, at the same time, develop cleaner and more elegant solutions to problems in AI, drawing on insights gained from kluges (Appendix III). The SP system is a good candidate for inclusion in the second strand of research.

D. WATSON AND ENERGY CONSUMPTION

Apart from the concerns outlined above, a major objection is that Watson does nothing to solve the problem of energy consumption in the processing of big data, outlined in Section IV. Indeed, the projected addition of deep learning to Watson, mentioned in Section IX-C, is likely to make things worse owing to inefficiencies with deep learning and artificial neural networks, outlined in Section V-E.

E. A POSSIBLE CONTRIBUTION FROM THE SP SYSTEM

In keeping with the two-pronged strategy mention in Section IX-C, there is probably a case for recasting Watson in the SP framework. There is potential to create a system that combines the strengths of Watson with the strengths of the SP system: in aspects of AI, in its origins in research on human perception and cognition, in its potential to simplify and integrate diverse kinds of knowledge and diverse kinds of processing, and in its potential for very substantial cuts in energy consumption.

X. BIG DATA AND AUTONOMOUS ROBOTS

Potential benefits and applications for the SP theory are summarised in Appendix I-I. This section gives a bit more detail about two areas of potential application: big data and autonomous robots.

A. BIG DATA

The paper “Big data and the SP theory of intelligence” [75] describes how the SP theory may help to solve nine problems with big data:

- *Helping to overcome the problem of variety in big data.* The SP system may serve as a universal framework for the representation and processing of knowledge (U FK), helping to tame the great variety of formalisms and formats for data, each with its own mode of processing (Section XI-E).
- *Learning and discovery.* In accordance with the DONSVIC principle (Section II-I, [72, Sec. 5.2]), the system has strengths in the unsupervised learning or discovery of ‘natural’ structures in data, with potential for further development.
- *Interpretation of data.* The SP system has strengths in areas such as pattern recognition, information retrieval, parsing and production of natural language, translation from one representation to another, several kinds of reasoning, planning and problem solving.

- *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 compression can yield direct benefits in the storage, management, and transmission of data, and indirect benefits in several of the other areas outlined in this subsection and discussed in [75].
- *Supercharging the transmission of data.* In addition to economies in the transmission of data via simple reductions in volume, there is potential for additional and very substantial economies in the transmission of data by judicious separation of ‘encoding’ and ‘grammar’.
- *Computational and energy efficiency.* There is potential for large gains in the computational efficiency of computers, with corresponding savings in the use of energy in computing, and for reductions in the size and weight of computers (Section IV).
- *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.

Considering these proposed solutions collectively, and in several cases individually, it appears that there are no alternatives that can rival the potential of what is described in [75].

B. AUTONOMOUS ROBOTS

The paper “Autonomous robots and the SP theory of intelligence” [74] describes how the SP theory may help in the design of the information-processing ‘brains’ of autonomous robots:

- *Computational and energy efficiency.* This is a revised version of the discussion in [75, Sec. IX].
- *Towards human-like versatility in intelligence.* The strengths of the SP system in diverse areas, summarised in Appendix I-H, can help in the development of human-like versatility in autonomous robots.
- *Towards human-like adaptability in intelligence.* It appears that unsupervised learning in the SP framework has potential as a key to human-like adaptability in intelligence, both directly and as a basis for other kinds of learning.

This approach to the development of intelligence in autonomous robots is quite different from others, and arguably more promising.

XI. DISTINCTIVE FEATURES AND ADVANTAGES OF THE SP SYSTEM IN COMPARISON WITH SOME SYMBOLIC ALTERNATIVES

This section summarises distinctive features and advantages of the SP system in symbolic kinds of computing. *A key advantage of the SP system in these areas (and others*

discussed elsewhere in this paper) is that it provides for the seamless integration of each function with other aspects of intelligence (Section II-B). This applies in each of the subsections below but will not be constantly repeated.

A. PATTERN RECOGNITION AND VISION

The main strengths and potential of the SP system in pattern recognition are described in [70, Ch. 6] and [72, Sec. 9]. In brief, these are: recognition at multiple levels of abstraction (Figure 1) and via hierarchies of parts and sub-parts (Figure 2); recognition of ‘family resemblance’ or polythetic categories; recognition that is robust in the face of errors of omission, commission or substitution (Section XI-A); for each act of recognition, calculation of associated probabilities; modelling the way in which context may influence recognition.

In addition to the strengths just mentioned, there is, in computer vision [73], potential in the SP system in: scene analysis via the parsing of an image; the learning of visual entities and classes of entity; the creation of 3D models of objects and of their surroundings; explaining how we can see things that are not objectively present (Section XI-A2); providing insights into the phenomena of size constancy, lightness constancy, and colour constancy; and finding alternative interpretations of images and scenes.

The two subsections that follow amplify some of the points above.

1) RECOGNITION IN THE FACE OF ERRORS OF OMISSION, COMMISSION, AND SUBSTITUTION

Figure 5 shows how, via the creation of multiple alignments, the SP system may recognise words despite errors of omission ((a) in the figure), commission (b), and substitution (c).²⁴

These examples are very simple but this capability of the SP system applies in much more complex examples of pattern recognition and also in other areas such as the processing of natural language (Section XI-B).

2) AMODAL PERCEPTION AND COMPLETION

“A remarkable property of human perception is the ease with which our visual system interpolates information not directly visible in an image. A particularly prominent example of this ... is *amodal perception*: the phenomenon of perceiving the whole of a physical structure when only a portion of it is visible” ([83, pp. 1 and 2], emphasis in the original).

A familiar example of amodal perception is how we can recognise something like a vase with flowers even when it is partly hidden or ‘occluded’ by something like a mug. Here, we have no difficulty with ‘amodal completion’: imagining the parts of the vase and flowers that we would see if the mug was removed.

²⁴By contrast with most multiple alignments shown in this paper, the ones in this figure and also in Figures 6 and 11 have been rotated by 90°, replacing columns with rows. The choice between these two representations, which are equivalent, depends on what fits best on the page.

0		t	r	a	s	f	o	r	m	a		o	n		0						
1	W	w	1		t	r	a	n	s	f	o	r	m	a	t	i	o	n	#	W	1

(a)

0		i	n	t	e	r	n	a	t	i	p	q	o	n	a	l		0				
1	W	w	2		i	n	t		e	r	n	a	t	i		o	n	a	l	#	W	1

(b)

0		f			x	y		u	r	i	s	h		z	n	g		0			
1	W	w	3		f	l	o			u	r	i	s	h	i		n	g	#	W	1

(c)

FIGURE 5. Three multiple alignments illustrating recognition with errors of omission (a), commission (b), and substitution (c), as described in the text.

0				v	a	s	e	-	w	i	t			m	u	g			l	o	w	e	r	s		0
1	OBJ	obj1		v	a	s	e	-	w	i	t	h	-	f					l	o	w	e	r	s	#OBJ	1

(a)

0				v	a	s	e	-	w	i	t	m	u	g					l	o	w	e	r	s	0
1	OBJ	obj2												m	u	g	#OBJ								1

(b)

FIGURE 6. Two multiple alignments modelling perception and occlusion, as described in the text. (a) This is a one-dimensional analogue of how a vase containing flowers may be recognised despite its partial occlusion by a mug. (b) This shows how the previously-mentioned mug may be recognised in the multiple alignment framework.

Figure 6 shows how the SP computer model may solve a one-dimensional analogue of this perceptual problem. Row 0 in each of the two multiple alignments contains the pattern ‘v a s e - w i t m u g l o w e r s’, which is an analogue of a vase with flowers which is partly obscured by a mug. The first multiple alignment (a) shows how the vase with flowers—represented by the pattern ‘OBJ obj1 v a s e - w

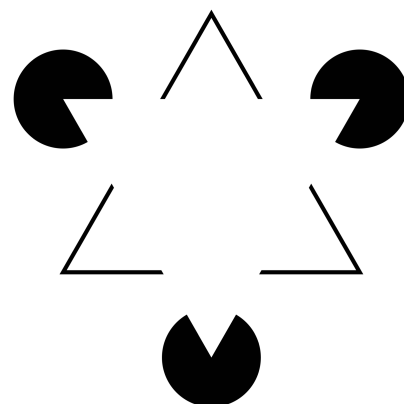


FIGURE 7. Kanizsa's triangle.

i t h - f l o w e r s #OBJ' in row 1—may be recognised despite its partial occlusion by 'm u g', and how the missing characters ('h - f') may be interpolated. The second multiple alignment (b) shows how the mug (represented by the pattern 'OBJ obj2 m u g #OBJ' in row 1) may be recognised as a separate entity.

This example does not model the way we perceive the mug to lie in front of the vase with flowers. But, as described in [73, Secs. 6.1 and 6.2], the SP system has potential for the building of 3D models. If that potential can be realized, the

capability may potentially be married with amodal perception of one object that is partly obscured by another.

A popular example of amodal completion is Kanizsa's triangle, shown in Figure 7. Here, we 'see' a white equilateral triangle in the middle of the figure with its apex at the bottom, although it is marked only with a minimum of features: three corners, each in a black disk, and breaks in the sides of another equilateral triangle with its apex at the top.

Figure 8 shows how the SP computer model may recognise a one-dimensional analogue of such a figure via the building

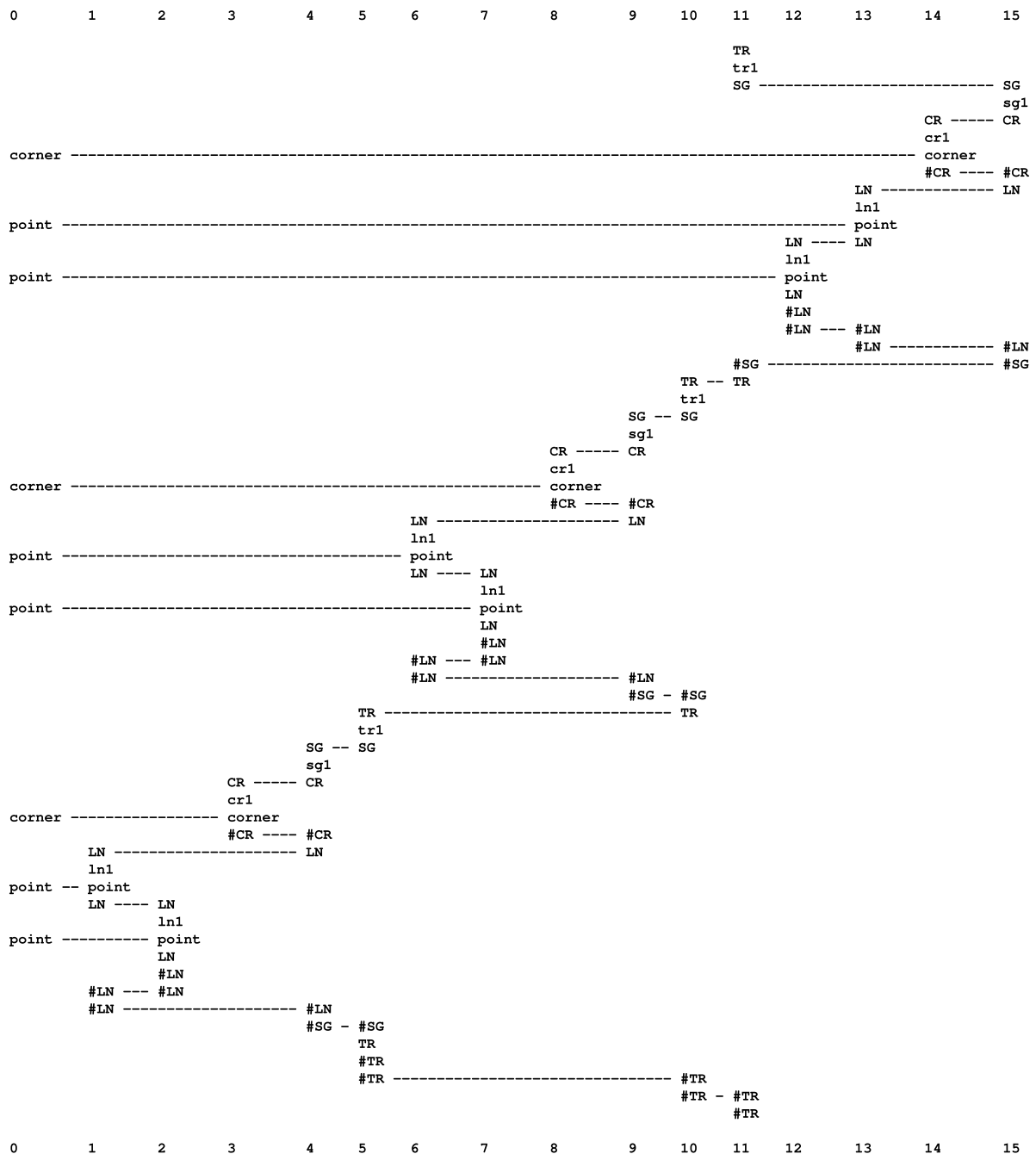


FIGURE 8. A multiple alignment modelling a one-dimensional analogue of how Kanizsa's triangle may be perceived, as described in the text.

of a multiple alignment, using the features that are available. Column 0 shows a New pattern representing the minimal features: each ‘corner’ in the multiple multiple alignment represents a corner in a black disk in Figure 7, and each ‘point’ in the multiple alignment is a point on a side of the target triangle where there is break in the side of the other triangle.

In this example, recognition is achieved via the parsing of the New pattern in terms of Old patterns representing elements of the triangle: the pattern ‘TR’ in column 11 represents the whole triangle, ‘SG’ in columns 4, 9, and 15 represents a segment of the triangle comprising a corner (‘CR’) and a line (‘LN’) which is itself a sequence of points.

In this example, there is recursion (like that shown in Figure 3) at two levels: the triangle is a recursive grouping of segments (so that, strictly speaking, it represents a polygon, not a triangle), and the line is built from a recursive sequence of points.

The example illustrates amodal completion because a mere two points on each side of the Kanizsa triangle are cues for what is potentially an infinite sequence of points between one corner and another.

B. THE PROCESSING OF NATURAL LANGUAGE

With the processing of natural language, the main strengths and potential of the SP system are described in [70, Ch. 5] and [72, Sec. 8]. In brief: it can model the parsing of natural language directly and transparently, as illustrated in Figure 10; it provides a simple, direct means of encoding discontinuous dependencies in syntax; in parsing, it can accommodate syntactic ambiguities, and the resolution of ambiguities via the provision of appropriate context; parsing may be achieved via recursive structures in syntax (Figure 3); parsing is robust against errors of omission, commission, or substitution (much as in pattern recognition—Section XI-A); and one mechanism may achieve both the parsing and production of natural language.

A key advantage of the SP system compared with most other systems for the processing of natural language is that, because all kinds of knowledge—both syntactic and semantic—are represented in one simple format (SP patterns) and because all kinds of knowledge are processed in the same way (via the formation and processing of multiple alignments), the system has clear potential to facilitate the seamless integration of syntax and semantics. In a similar vein, the processing of natural language may be integrated seamlessly with other aspects of intelligence.

C. UNSUPERVISED LEARNING OF NATURAL LANGUAGE

Although grammatical inference has been the subject of research for many years, automatic learning of the syntax of a natural language remains a major challenge. Even more difficult is unsupervised learning of the rich structure of non-linguistic knowledge which provides the ‘meanings’ for language, and unsupervised learning of the kinds of syntactic-semantic structures that are needed for such things as interpreting the meaning of natural language, the production of

speech or writing from meanings, and, when it is done at a high standard, translation from one language to another (Section V-K2).

As mentioned in Appendix I, the SP programme of research grew out of earlier research developing computer models of language learning, but it has required a radical reorganisation of earlier models to meet the new goals of the programme. Now, the SP computer model demonstrates unsupervised learning of plausible generative grammars for the syntax of English-like artificial languages, including the learning of segmental structures, classes of structure, and abstract patterns [70, Ch. 9], in accordance with the DONSVIC principle (Section II-I).

A key insight from this research, discussed briefly in Section V-H3, is that information compression appears to solve the problems of both over- and under-generalisation in the learning of language ([70, Sec. 9.5.3], [72, Sec. 5.3]). This appears to be a significant advantage compared with, for example, deep learning in neural networks (Section V-H3).

It appears that, with some further work, the potential is considerable: in the learning of syntactic structures, in the learning of semantic structures, and in the learning of syntactic-semantic structures. The potential of the SP system in these areas appears to be a major advantage of the system compared with alternatives.

D. EXACT AND INEXACT FORMS OF REASONING

From the General Problem Solver [45], through Prolog [3], to such systems as Description Logics [19], concepts derived from classical logic have been prominent in AI and related fields such as the semantic web.

Although the all-or-nothing certainties of classical logic can be useful, it has been recognised for some time that much of human thinking and reasoning revolves around judgements that may have varying levels of certainty. This has led to several proposals for systems that, in one way or another, combine exact or ‘logical’ reasoning with inexact, ‘fuzzy’ or ‘probabilistic’ kinds of reasoning (eg, [27], [80]). Some recent developments are described in [49].

As described in [70, Ch. 7], a distinctive feature and apparent advantage of the SP system is that it is fundamentally probabilistic (Appendix I-F) but levels of confidence may be increased by gathering more evidence, or by concentrating on probabilities that are close to 0 or 1, or both those things. And the SP system can model several of the concepts that are familiar in logic and mathematics (such as *variable*, *value*, and *type*), as described in [70, Ch. 10] and [77, Sec. 6.6]. There is clear potential for the system to model both probabilistic and exact forms of reasoning and to switch between them according to need.

The SP system also demonstrates several more specific kinds of reasoning within one unified framework ([70, Ch. 7 and 10, Secs. 6.4], [72, Sec. 10]): one-step ‘deductive’ reasoning; abductive reasoning; reasoning with probabilistic decision networks and decision trees; reasoning with ‘rules’; nonmonotonic reasoning and reasoning with

default values; reasoning in Bayesian networks, including ‘explaining away’ (Section VIII); causal diagnosis; reasoning which is not supported by evidence; and inheritance of attributes in an object-oriented class hierarchy or heterarchy. There is also potential for spatial reasoning [74, Sec. IV-F.1] and what-if reasoning [74, Sec. IV-F.2].

Because these several kinds of reasoning all flow from one unified framework, they may be used together in any combination according to need. In the same vein, these several forms of reasoning may integrate seamlessly with other aspects of intelligence: pattern recognition, natural language processing, unsupervised learning, and so on. This kind of flexibility is a distinctive feature and apparent advantage of the SP system compared with alternatives (Section II-B).

E. REPRESENTATION AND PROCESSING OF DIVERSE FORMS OF KNOWLEDGE

A problem with AI and other areas of computing as they have developed to date is that knowledge may be represented with a large number of different formalisms, and often, for each one, there is a large number of different formats. This complexity is compounded by the fact that, normally, each formalism and each format has its own mode of processing. Until recently, this complexity has been easy to ignore. But with the advent of big data, it has become a major problem, a problem that the SP system may help to solve (Section X-A).

As described in outline in [75, Sec. III-B], and in more detail in [70] and [72], the SP system promotes the seamless integration of a wide variety of kinds of knowledge, with seamless integration of their processing. These include: the syntax of natural language; class hierarchies, part-whole hierarchies, and their integration; trees and networks, including Bayesian networks; entity-relationship structures; relational knowledge (tuples); if-then rules, associations, and other knowledge in support of reasoning; patterns and images; structures in three dimensions; and sequential and parallel procedures.

Since information compression via ICMUP and multiple alignment is at the heart of how knowledge is represented in the SP system, since information compression can in principle be an efficient means of representing any kind of knowledge, and since the multiple alignment framework appears to be a very general means of compressing information, there is reason to believe that *any* kind of knowledge—perhaps including the kind of ‘commonsense’ knowledge that is described and discussed in [7]—may be represented effectively in the SP system. Similar things may be said, *mutatis mutandis*, about reasoning and other kinds of processing of knowledge in the SP system.

F. SOFTWARE ENGINEERING

Although machine learning is beginning to make an impact, controlling what computers do is still done largely in the traditional manner, via the running of programs and their creation by people. Although machine learning may eventually become dominant, it is likely that, with all kinds of computing system, there will be a continuing need for the foreseeable

future, for each system to be controllable by human-created software.

At first sight, the SP system fails this test. SP patterns don’t look like a conventional program and the building of multiple alignments does not look much like the running of a conventional program. But, as noted in Section XI-D, the multiple alignment framework can model such concepts as *variable*, *value*, and *type*. More generally, the SP system can model the kinds of concepts used in software engineering including *procedure*, *function* with *parameters*, conditional statements, repetition of procedures, the integration of programs and data, and elements of object-oriented design, including *class hierarchies* and *inheritance* [77, Sec. 6.6]. The SP system also has potential for the processing of parallel streams of information [74, Secs. V-G, V-H, and V-I, and Appendix III].

Because of these features of the SP system, and because of its strengths in unsupervised learning ([70, Ch. 9], [72, Sec. 5]), it has potential for ‘automatic programming’—the integration of programming with unsupervised learning—as outlined in [77, Sec. 6.6.4]. This is a distinctive feature of the system and an apparent advantage compared with systems that are specialised only for learning.

XII. CONCLUSION

Preceding sections of this paper have aimed to highlight distinctive features of the *SP theory of intelligence* and its apparent advantages compared with some AI-related alternatives.

Section II summarises distinctive features and strengths of the SP system:

- Simplification and integration of observations and concepts;
- Simplification and integration of structures and processes in computing systems;
- The SP theory is itself a theory of computing;
- The theory provides the basis for new architectures for computers;
- Information compression via the matching and unification of patterns is central in the theory;
- More specifically, all processing is done via the building of *multiple alignments*, a concept borrowed and adapted from bioinformatics;
- Transparency in the representation and processing of knowledge;
- The unsupervised learning of ‘natural’ structures via information compression (DONSVIC);
- Interpretation of aspects of mathematics in terms of the SP theory;
- Interpretation of phenomena in human perception and cognition;
- Realisation of abstract concepts in terms of neurons and their inter-connections (*SP-neural*).

In several sections, distinctive features and advantages of the SP system have been highlighted in comparison with AI-related alternatives:

- The concept of minimum length encoding and related concepts;

- How computational and energy efficiency in computing may be achieved;
- Deep learning in neural networks;
- Unified theories of cognition and related research;
- Concepts of universal search;
- Bayesian networks and some other models for AI;
- IBM's Watson;
- Solving problems associated with big data;
- Solving problems in the development of intelligence in autonomous robots;
- The processing of natural language;
- Unsupervised learning of natural language;
- Exact and inexact forms of reasoning;
- Representation and processing of diverse forms of knowledge;
- Software engineering.

The main conclusion of the paper is that, while some alternatives to the SP system may deliver applications fairly quickly, a strength of the SP system is that it can provide a firm foundation for the long-term development of AI, with many potential benefits and applications, and, at the same time, it may deliver useful results on relatively short timescales.

It is envisaged that a high-parallel, open-source version of the SP machine will be created, hosted on an existing high-performance computer, and derived from the existing SP computer model (Appendix I-J). This would be a means for researchers everywhere to explore what can be done with the system, and to create new versions of it.

APPENDIX I OUTLINE OF THE SP THEORY OF INTELLIGENCE AND SP MACHINE

As noted in the Introduction, the *SP theory of intelligence* is designed to simplify and integrate observations and concepts across artificial intelligence, mainstream computing, mathematics, and human perception and cognition, with information compression via *multiple alignment* as a unifying theme. As outlined in Appendix I-C, the SP theory is realized in the SP computer model which may be regarded as a version of the *SP machine*.

The SP theory originates in part from an earlier programme of research on grammatical inference and the unsupervised learning of natural language, with information compression at centre stage [68]. However, meeting the goals of the SP research programme has meant a radical reorganisation of the system, with the development of a concept of *multiple alignment* (Appendix I-E2) as a framework for the simplification and integration of diverse structures and functions [74, Sec. V-A.4].

This Appendix is intended to provide readers with sufficient information about the SP system to make the rest of the paper intelligible. The outline starts with a short informal account of the SP system, followed by subsections that

describe the main elements of the system in a little more detail.

A. AN INFORMAL ACCOUNT OF HOW THE SP SYSTEM WORKS

The SP theory is conceived as an abstract brain-like system that, in an 'input' perspective, may receive *New* 'patterns' via its senses, and compress some or all of it to create *Old* 'patterns', as illustrated schematically in Figure 9.

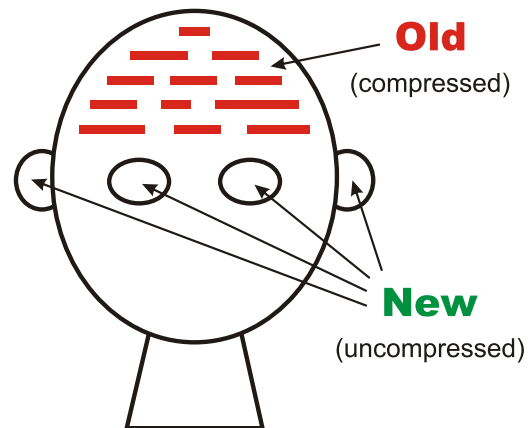


FIGURE 9. Schematic representation of the SP system from an 'input' perspective.

In the early stages, when there is little or no Old information in store, the system simply stores New patterns directly as Old patterns, except that 'identification' ('ID') symbols are added to each pattern for use later.

After a while, New patterns are received that are fully or partially the same as Old patterns. Then, the system builds multiple alignments as outlined in Appendix I-E2. From the multiple alignments:

- When a New pattern is exactly the same as a stored Old pattern, the system records the occurrence of that pattern in terms of its ID-symbols, not the pattern itself. Since the ID-symbols are normally relatively short compared with their associated pattern, the effect is to store the New pattern in a compressed form.
- When a New pattern is received that is partially the same as one or more Old patterns, the system constructs patterns from the parts of the New and Old patterns that match each other and from the parts that do not match each other. Each newly-created pattern is assigned its own ID-symbols and is stored as an Old pattern. The system also constructs 'abstract' patterns that, using relevant ID-symbols, record the sequential relationship of the newly-constructed patterns.
- Similar principles apply when the system builds multiple alignments containing previously-constructed abstract patterns. And it is envisaged that similar principles will apply when the SP computer model has been developed for the representation and processing of 2D patterns.

Many of the Old patterns that are created in this way may be seen as ‘good’, but, normally, the system also creates many that people are likely to regard as ‘bad’. To get rid of the bad patterns, SP system applies a process for evaluating sets of Old patterns (called *grammars*) in terms of their ability to compress incoming New information. The one or two grammars that yield high levels of compression are retained by the system while residual patterns—those not found in the successful grammars—are discarded.

It is interesting to see that, very often, patterns that are ‘good’ in terms of information compression are also ‘good’ in terms of human judgements, and *vice versa*. This correspondence between what people regard as ‘natural’ and what yields high levels of compression is termed the ‘DONSVIC’ principle (Section II-I).

The building of multiple alignments provides the means by which New information may be encoded economically in terms of Old information. It also provides much of the versatility of the system in such functions as unsupervised learning, the parsing and production of natural language, pattern recognition, computer vision, information retrieval, several kinds of reasoning, planning, problem solving, and information compression.

B. FOUNDATIONS AND SCOPE OF THE SP THEORY

The SP theory is founded on a range of observations suggesting the fundamental importance of information compression in natural and artificial intelligence, in computing, in mathematics, and in neuroscience ([70, Chapter 2], [76]).

Like most theories, the SP theory is narrower in its scope than one might wish. It is certainly not a comprehensive theory of human psychology. For example, it has little to say about emotions and motivations and their impact, in people, on such things as perception, learning, and reasoning (but see [74, Sec. V-A.2]). At some stage, there is likely to be a case for examining whether or how those kinds of things may be accommodated in the theory.

C. THE SP COMPUTER MODEL AND THE SP MACHINE

The SP theory is realized in the form of a computer model, SP71, which may be regarded as a version of the *SP machine*.

An outline of the organisation and workings of the SP computer model may be found in [70, Sec. 3.9], with more detail, including pseudocode, in [70, Secs. 3.10 and 9.2].²⁵ Fully commented source code for the SP71 computer model may be downloaded via a link near the bottom of <http://www.cognitionresearch.org/sp.htm>, and via ‘Ancillary files’ under <http://arxiv.org/abs/1306.3888>.

All the multiple alignments shown in this paper are output from the SP computer model.

²⁵These sources describe SP70, a slightly earlier version of the model than SP71 but quite similar to it. The description of SP70 includes a description, in [70, Secs. 3.9.1 and 3.10], of a subset of the SP70 model called SP61.

D. PATTERNS AND SYMBOLS

In the SP system, knowledge is represented with arrays of atomic symbols in one or two dimensions called *patterns*. The SP71 model works with one-dimensional patterns but it is envisaged that the system will be generalised to work with patterns in two dimensions [72, Sec. 3.3].

Each SP pattern has an associated frequency of occurrence that has a role in the calculation of probabilities, as outlined in Appendix I-F.

An ‘atomic symbol’ in the SP system is simply a mark that can be matched with any other symbol to determine whether it is the same or different—no other result is permitted.

Patterns in two dimensions are likely to have a role in the processing of images ([70, Ch. 13], [73]) and also in the processing of sensory or motor streams of information that occur in parallel [74, Secs. IV-A.4, IV-H, V-G to V-I, and Appendix III].

In themselves, SP patterns are not particularly expressive. But within the multiple alignment framework (Appendix I-E2), they support the representation and processing of a wide variety of kinds of knowledge (Section XI-E, Appendix I-H). It appears that the system has potential as a *universal framework for the representation and processing of knowledge* (UFK) [75, Sec. III].

E. INFORMATION COMPRESSION

In the SP system, all kinds of processing is done by compression of information. This is essentially the principle of *minimum length encoding* (MLE) [48, 56, 65]²⁶ but with qualifications described in Section III.

The default assumption in the SP theory is that compression of information is always lossless, meaning that all non-redundant information is retained. In particular applications, there may be a case for discarding non-redundant information (see, for example, [75, Sec. X-B]) but any such discard is reversible.

The name ‘SP’ is short for *Simplicity* and *Power*, because compression of any given body of information, **I**, may be seen as a process of reducing ‘redundancy’ of information in **I** and thus increasing its ‘simplicity’, whilst retaining as much as possible of its non-redundant descriptive and explanatory ‘power’. As noted in Appendix II, it is no accident that the same two concepts are prominent in Occam’s Razor as a touchstone of success for scientific theories.

In the SP system, information compression is achieved via the matching and unification of patterns, or parts thereof (see ‘ICMUP’ in Section II-E and Appendix I-E1). More specifically, it is achieved via the building of multiple alignments and via the unsupervised learning of grammars. These three things are described briefly in the following three subsections.

²⁶MLE is an umbrella term for ‘minimum message length’ encoding (MML), ‘minimum description length’ encoding (MDL), and similar concepts.

1) INFORMATION COMPRESSION VIA THE MATCHING AND UNIFICATION OF PATTERNS

The basis for information compression in the SP system is a process of searching for patterns that match each other with a process of merging or ‘unifying’ patterns that are the same: ‘information compression via the matching and unification of patterns’ or ‘ICMUP’ [76].

At the heart of the SP computer model is a method for finding good full and partial matches between sequences, with advantages compared with classical methods [70, Appendix A].²⁷

The emphasis on ICMUP is motivated partly by evidence of the importance of such processes in human perception and cognition, and partly by its potential to cut through much complexity and to achieve a new perspective on AI, mainstream computing, and mathematics [76]. Because a goal of the SP theory is to develop a new perspective on those four areas, without theoretical ‘baggage’, the theory minimises the use of concepts from those disciplines, including mathematics [76, Sec. 2.1].²⁸

This is not intended to be in any way disrespectful of mathematics as a discipline or the monumental achievements of mathematicians. It is merely an observation that, if one is trying to achieve a new perspective on any discipline, it is probably best to avoid using too many concepts from that discipline. With regard to mathematics, it would be surprising if it had not been shaped, at least in part, by its exceptionally long history, and the fact that, for most of its existence, the discipline has depended largely on human brainpower, with nothing but the simplest kinds of artificial aid.

2) INFORMATION COMPRESSION VIA THE BUILDING OF MULTIPLE ALIGNMENTS

The process for finding good full and partial matches between patterns is the foundation for processes that build *multiple alignments* like the one shown in Figure 10. This concept is similar to multiple alignment in bioinformatics but with important differences [70, Sec. 3.4]. It is a powerful and distinctive feature of the SP system.

In Figure 10, the SP pattern in column 0 is a New pattern representing a sentence to be parsed, while each of columns 1 to 11 contains an Old SP pattern representing a grammatical form (where ‘grammatical form’ includes words). This example is the best multiple alignment created by the SP computer model with the New pattern as shown

²⁷The main advantages are [70, Sec. 3.10.3.1]: 1) That it can match arbitrarily long sequences without excessive demands on memory; 2) For any two sequences, it can find a set of alternative matches (each with a measure of how good it is) instead of a single ‘best’ match; 3) The ‘depth’ or thoroughness of the searching, which has the effect of controlling the amount of backtracking, can be controlled by parameters.

²⁸And, bearing in mind that the SP theory should be consistent with the biological origins of human intelligence, an attempt has been made to ensure that the frequency information that is stored with each SP pattern, and the probability calculations that are performed by the SP computer model, are, potentially, the kinds of things that could be modelled, at least approximately, via analogue processes in biological systems.

in column 0 and a set of pre-existing Old patterns, including those shown in columns 1 to 11 in the figure.

Here, the ‘best’ multiple alignment is the one in which the New pattern may be encoded most economically in terms of the Old patterns—and this means a multiple alignment in which there is a relatively large number of symbols that match each other from column to column, aligned in rows.

The way in which an encoding is derived from a multiple alignment is explained in [70, Sec. 3.5] and [72, Sec. 4.1]. Like all other kinds of knowledge in the SP system, encodings derived from multiple alignments are recorded using SP patterns (Appendix I-D).

The overall effect of this multiple alignment is to analyse the sentence into its grammatical parts and sub-parts, an analysis that is, in its essentials, the same as a conventional parsing.

A point of interest is that the pattern for the whole sentence in column 6 marks the grammatical dependency between the singular subject of the sentence (‘a stitch’)—marked with ‘Ns’—and the singular verb (‘saves’)—marked with ‘Vs’. Notice how the dependency neatly bridges the subordinate phrase (‘in time’). This method of encoding discontinuous dependencies in syntax contributes to the compression that is achieved by this multiple alignment and is, arguably, simpler than existing techniques for encoding such dependencies.

3) INFORMATION COMPRESSION VIA UNSUPERVISED LEARNING

From a set of New patterns, the SP system may, without assistance from a ‘teacher’ or anything equivalent, derive one or more plausible context-sensitive grammars, including segmental structures, classes of structure, and abstract patterns. The learning process is outlined in [70, Sec. 3.9.2] and [72, Sec. 5.1], and described more fully in [70, Ch. 9]. In that process, multiple alignment has a central role as a source of SP patterns for possible inclusion in any grammar ([70, Sec. 9.2.5], [72, Sec. 5.1.1]). Although the current model has some shortcomings (Appendix I-J, [72, Sec. 3.3]), it appears that these may be overcome.

A key part of the learning process is the formation of multiple alignments in which there are mismatches between patterns like those shown in Figure 11.

From the matched and unmatched parts of a multiple alignment like this, the system derives such patterns as ‘B 1 t h a t #B’, ‘C 1 r u n s #C’, ‘D 1 g i r l #D’, and ‘D 2 b o y #D’, each one with system-generated ID-symbols at the beginning and end. The system also creates a pattern like this: ‘E 1 B #B D #D C #C #E’, that records the sequence of structures in terms of the ID-symbols. These patterns are the beginnings of a simple grammar.

In practice, many of the multiple alignments, and many of the derived segments, are much less tidy than this example may suggest. But, via heuristic search (Appendix I-E4) through the space of alternative grammars,

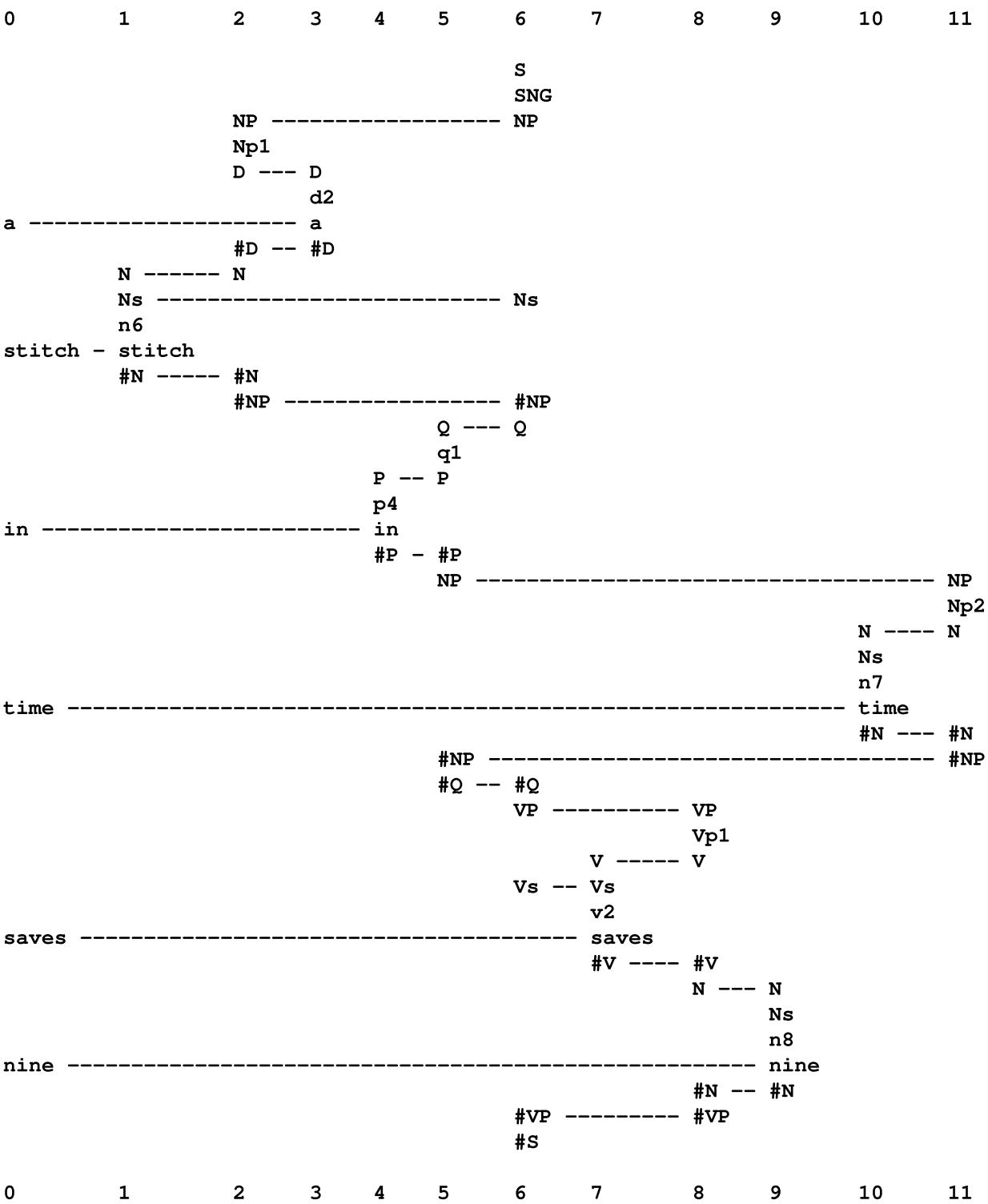


FIGURE 10. A multiple alignment illustrating parsing of a sentence, as described in the text.

the system is able to derive one or two grammars that score relatively well in terms of information compression. Normally, these are also grammars that appear to be most natural, in accordance with the DONSVIC principle (Section II-I).

4) HEURISTIC SEARCH

Like most problems in artificial intelligence, each of the afore-mentioned problems—finding good full and partial matches between patterns, finding or constructing good multiple alignments, and inferring one or more good grammars

```

0      t h a t g i r l r u n s      0
      | | | |      | | | |
1 A 1 t h a t b o y   r u n s #A 1

```

FIGURE 11. A simple multiple alignment from which other patterns may be derived.

from a body of data—is normally too complex to be solved by exhaustive search.

With intractable problems like these, it is often assumed that the goal is to find theoretically ideal solutions. But with these and most other AI problems, “The best is the enemy of the good”. By scaling back one’s ambitions and searching for ‘reasonably good’ solutions, it is often possible to find solutions that are useful, and without undue computational demands.

As with other AI applications, and as with the building of multiple alignments in bioinformatics, the SP71 model uses heuristic techniques—‘hill climbing’ or ‘descent’—in all three cases mentioned above [70, Appendix A; Secs. 3.9 and 3.10; Ch. 9]. This means searching for solutions in stages, with a pruning of the search tree at every stage, guided by measures of compression, and with backtracking where appropriate to increase the chances of success. With these kinds of techniques, acceptably good approximate solutions can normally be found without excessive computational demands and with ‘big O’ values that are within acceptable limits. Possible alternatives to the heuristic techniques used in SP71 include ‘simulated annealing’ and ‘genetic programming’.

5) GRAMMARS AND ENCODINGS, SIMPLICITY AND POWER

In unsupervised learning in the SP system, compression of a body of information, **I**, produces two distinct results: a *grammar* and an *encoding* of **I** in terms of the grammar, both of them expressed as SP patterns. The two together represent a lossless compression of **I**.

The term ‘grammar’ has been adopted because the SP programme of research derives largely from earlier research on models of language learning and grammatical inference [68] but, because of the versatility of SP patterns in the multiple alignment framework (Appendix I-D), the term is applied, in this research, to any kind of knowledge, not just natural language.

Often but not invariably, there is a trade-off between the size of the grammar and the size of the encoding: as a general rule, small grammars yield large encodings and large grammars yield small encodings. Normally, the greatest overall compression of **I** is obtained with grammars that are not at the extremes of size (small or big), and likewise for encodings. It appears that this means learning that avoids both under-generalisation and over-generalisation (Sections V-H and XI-C).

From the trade-off we can see that there is a direct relationship between the concepts of ‘grammar’ and ‘encoding’ on the one hand, and the concepts of ‘simplicity’ and ‘power’ on the other: for a given **I**, there is simplicity in any grammar

when the grammar is small, and the grammar has power when the encoding is small. Any reasonably thorough compression of **I** is likely to yield a good balance between the two.²⁹

F. INFORMATION COMPRESSION, PREDICTION, AND PROBABILITIES

Owing to the close connection between information compression and concepts of prediction and probability [35], the SP system is fundamentally probabilistic. As noted in Appendix I-D, each SP pattern has an associated frequency of occurrence. Using that frequency information, probabilities may be calculated for each multiple alignment and for any inference that may be drawn from any given multiple alignment [70, Sec. 3.7].

Although the SP system is fundamentally probabilistic: it can be constrained to answer only those kinds of questions where probabilities are close to 0 or 1; and, via the use of error-reducing redundancy, it can deliver decisions with high levels of confidence. Contrary to what one may suppose, there is no conflict between the use of error-reducing redundancy and the notion that ‘computing’ may be understood as information compression. The two things are independent, as described in [70, Sec. 2.3.7].

G. SP-NEURAL

Part of the SP theory is the idea, described most fully in [70, Ch. 11], that the abstract concepts of *symbol* and *pattern* in the SP theory may be realized more concretely in the brain with collections of neurons in the cerebral cortex.³⁰

The neural equivalent of an SP pattern is called a *pattern assembly*. The word ‘assembly’ has been adopted in this term because the concept is quite similar to Hebb’s [17] concept of a *cell assembly*. The main difference is that the concept of pattern assembly is unambiguously explicit in proposing that the sharing of structure between two or more pattern assemblies is achieved by means of ‘references’ from one structure to another, as described and discussed in [70, Sec. 11.4.1]). Also, learning in the SP system is quite different from the gradualist ‘Hebbian’ kinds of learning that are popular in artificial neural networks (see Sections V-A and V-D).

Figure 12 shows schematically how pattern assemblies may be represented and inter-connected with neurons. Here, each pattern assembly, such as ‘< NP < D > < N > >’, is represented by the sequence of atomic symbols of the corresponding SP pattern. Each atomic symbol, such as ‘<’ or ‘NP’, would be represented in the pattern assembly by one neuron or a small group of inter-connected neurons. Apart from the inter-connections amongst pattern assemblies, the cortex in SP-neural is somewhat like a sheet of paper on which knowledge may be written in the form of pattern assemblies and their inter-connections.

²⁹Here, the qualification, “reasonably thorough” is quite important. Compression algorithms like the popular LZ algorithms are ‘quick and dirty’—they are designed for speed on low-powered computers. They are not very thorough and will normally miss quite large amounts of redundancy.

³⁰See also [70, Sec. 2.3.1].

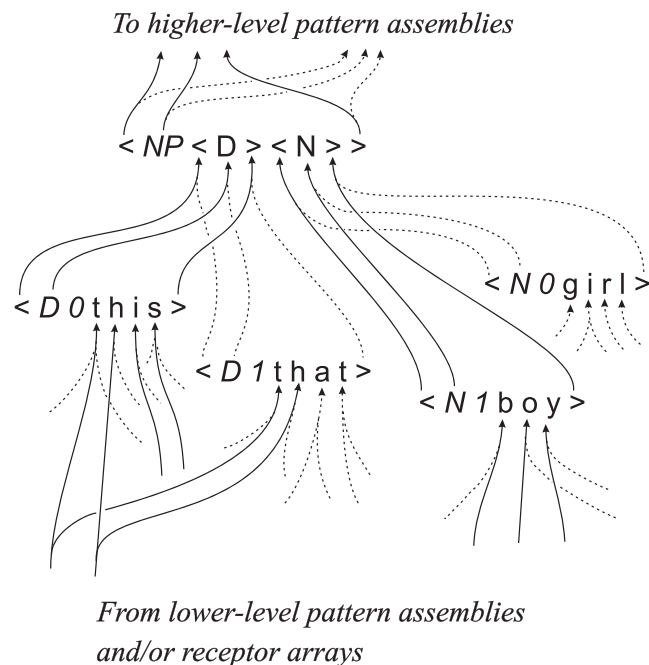


FIGURE 12. Schematic representation of inter-connections amongst pattern assemblies as described in the text. Not shown in the figure are lateral connections within each pattern assembly, and inhibitory connections elsewhere, as outlined in [70, Secs. 11.3.3 and 11.3.4]. Reproduced, with permission, from [70, Fig. 11.2].

As noted in Section V-B, the hierarchical relations that can be seen in Figure 12 may be seen to be broadly in keeping with the hierarchical relations between ‘simple’ and ‘complex’ cells discovered by Hubel [20].

It is envisaged that any pattern assembly may be ‘recognised’ if it receives more excitatory inputs than rival pattern assemblies, perhaps via a winner-takes-all mechanism in which inhibitory processes have a role to play [70, Sec. 11.3.4]. And, once recognised, any pattern assembly may itself be a source of excitatory signals leading to the recognition of higher-level pattern assemblies.

H. EMPIRICAL AND CONCEPTUAL SUPPORT FOR THE SP THEORY

As noted in Appendix II-B, the SP theory has non-trivial things to say about a wide range of observations and concepts in artificial intelligence, mainstream computing, mathematics, and human perception and cognition. These things are described most fully in [70], more briefly in [72], and in extended summaries in [74, Secs. IV and V]. In a bare-bones summary, the main strengths of the SP system are in:

- Natural language processing ([70, Ch. 5], [72, Sec. 8]).
- Pattern recognition and vision ([70, Ch. 6], [72, Sec. 9], [73]).
- Information storage and retrieval ([70, Ch. 6], [72, Sec. 11], [71]).
- The representation and processing of diverse kinds of knowledge ([72, Sec. 7], [75, Sec. III-B] and, more generally, [70, Ch. 5 to 10]).
- Benefits accruing from the seamless integration of

diverse kinds of knowledge and diverse aspects of intelligence ([77, Secs. 2, 5, and 7]).

- Several kinds of reasoning ([70, Ch. 7], [72, Sec. 10]).
- Planning and problem solving ([70, Ch. 8], [72, Sec. 12]).
- Unsupervised learning ([70, Ch. 9], [72, Sec. 5], [74, Sec. V]).
- Implications for our understanding of human perception and cognition, including neural processing ([70, Ch. 11 and 12], [73]).
- Implications for our understanding of the nature of mathematics ([70, Ch. 10], [76]).

There is more detail about some of these capabilities in the body of the paper.

I. POTENTIAL BENEFITS AND APPLICATIONS

In summary, potential benefits and applications of the SP system include:

- Helping to solve nine problems associated with big data [75] (see also Section X).
- The development of versatility and adaptability in autonomous robots, with potential for gains in computational efficiency [74] (see also Section X).
- The development of computer vision and pattern recognition, and the interpretation of aspects of natural vision ([72, Sec. 9], [73]).
- The system may be developed as a versatile database management system, with intelligence [71].
- The system may serve as a repository for medical knowledge and as an aid for medical diagnosis [69].
- There are several other potential benefits and applications described in [77]: simplification of computing systems, including software; unsupervised learning; the processing of natural language; software engineering; information compression; the semantic web; bioinformatics; the detection of computer viruses; data fusion; new kinds of computer; the development of scientific theories; and the seamless integration of diverse kinds of knowledge and processing.

As describe in Appendix I-J, next, some potential applications may be developed on relatively short timescales.

J. DEVELOPMENT OF THE SP SYSTEM

Like most scientific theories, the SP system is not complete [72, Sec. 3.3]. As it is now, the main shortcomings in the SP computer model are:

- The process for finding good full and partial matches between one-dimensional patterns needs to be generalised to patterns in two dimensions;
- A better understanding is needed of how the system may be applied to the discovery and recognition of low-level features in speech and images;
- In unsupervised learning, the model does not learn intermediate levels of abstraction or discontinuous dependencies in data;

- And a better understanding is needed of how the system may be applied in the representation and processing of numbers.

It appears that none of these problems are fatal—that all of them are soluble.

Apart from this work, developing the core model, there is much to be done in exploring how the system may be applied in areas such as pattern recognition, natural language processing, and so on (Appendix I-H), and in applications such as medical diagnosis, crime prevention and detection, and more (Appendix I-I).

Since there are many more avenues to be explored than could be tackled by any one research group, it is envisaged that the SP computer model will be the basis for the creation of a high-parallel, open-source version of the SP machine, hosted on an existing high-performance computer [78]. This will be a means for researchers everywhere to explore what can be done with the system, and to create new versions of it. If any reader is interested in the possibility of using such a facility or believes the facility may be useful, it would be helpful if he or she were to let the author know (via jgw@cognitionresearch.org).

Some potential applications of the SP system may be developed on relatively short timescales using existing high-performance computers or even ordinary computers. These include the SP system as an intelligent database [71], and applications in such areas as medical diagnosis [69], pattern recognition ([70, Ch. 6], [72, Sec. 9]), information compression [77, Sec. 6.7], highly-economical transmission of information [75, Sec. VIII], bioinformatics [77, Sec. 6.10.2], and natural language processing [77, Sec. 6.2].

APPENDIX II

OCCAM'S RAZOR: SIMPLICITY AND POWER

One of the most widely accepted principles in science—Occam's Razor—is that a good theory should combine conceptual *simplicity* with explanatory or descriptive *power*. Albert Einstein expressed it thus: "A theory is more impressive the greater the simplicity of its premises, the more different things it relates, and the more expanded its area of application."³¹

In these terms, a theory can be weak because it is too complex—like the Ptolemaic earth-centred system of epicycles as a theory of the movements of the planets and the sun, or merely redescribing the data that it is meant to explain. Or a theory can be weak because it is too simple and too general, explaining everything and nothing—like the over-enthusiastic use of the concept of 'instinct' as an explanation of animal behaviour. A good theory—like the Copernicus/Kepler heliocentric theory of the solar system—strikes a balance between the two. This relates to the issue of under-generalisation and over-generalisation discussed in Sections V-H and XI-C.

³¹Quoted in [22, p. 512].

Alan Turing's concept of a 'universal Turing machine', and equivalent models such as Post's [47] 'canonical system', are good models of 'computing' in the widest sense. But notwithstanding Turing's vision that computers might become intelligent [62], the concept of a universal Turing machine, does not tell us how!³² Something with more specific capabilities is needed for AI.

Since the beginnings of electronic computers, several decades of research in AI have yielded some useful insights and some impressive applications but I believe it is fair to say that AI has been and is suffering from an excess of narrow subfields and, with some honourable exceptions, insufficient attention to the need to simplify and integrate observations and concepts across different areas.³³ Hence the SP programme of research (Appendix II-A).

It is no accident that 'simplicity' and 'power'—key ideas in evaluating scientific theories—are also prominent in the SP theory:

- The two terms, together, are equivalent to 'information compression', which is central in the SP theory (Appendix I-E);
- Cosmologist John Barrow has written that "Science is, at root, just the search for compression in the world" [1, p. 247].

A. MOTIVATION

Part of the motivation for developing the SP theory has been to try to overcome the problems of narrow focus and over-specialisation, mentioned above and identified by other authors:

- Neisser [42] writes of the need to avoid 'microtheories' in psychology.
- In a similar vein, Newell, in his famous essay "You can't play 20 questions with nature and win" [43], urges researchers to develop theories with wide scope (pp. 284–289) dealing with "a genuine slab of human behaviour" (p. 303). This thinking led on to Newell's *Unified Theories of Cognition* [44] and related work, discussed in Section VI.
- "Today, as scientists labor to create machine technologies to augment our senses, there's a strong tendency to view each sensory field in isolation as specialists focus only on a single sensory capability. Experts in each sense don't read journals devoted to the others senses, and they don't attend one another's conferences. Even within IBM, our specialists in different sensing technologies don't interact much." [26, location 1004].
- And McCorduck writes "The goals once articulated with debonair intellectual verve by AI pioneers appeared unreachable . . . Subfields broke off—vision, robotics, natural language processing, machine learning, decision theory—to pursue singular goals in solitary splen-

³²But Turing began to address that problem in a report about "unorganised machines" [63] and also in [62].

³³Honourable exceptions include research aiming to develop 'unified theories of cognition' and 'artificial general intelligence'.

dor, without reference to other kinds of intelligent behaviour.” [38, p. 417]. Later, she writes of “the rough shattering of AI into subfields . . . and these with their own sub-subfields—that would hardly have anything to say to each other for years to come.” (*ibid.*, p. 424). She adds: “Worse, for a variety of reasons, not all of them scientific, each subfield soon began settling for smaller, more modest, and measurable advances, while the grand vision held by AI’s founding fathers, a general machine intelligence, seemed to contract into a negligible, probably impossible dream.”

B. EVALUATION OF THE SP THEORY IN TERMS OF SIMPLICITY AND POWER

Although, in comparing one theory with another, we must rely on relatively informal assessments of simplicity and power, the SP theory, in those terms, appears to do well [77, Sec. 4]:

- The key concept of multiple alignment, with associated processes, are, in the SP computer model, expressed in an ‘exec’ file that requires less than 500 KB of storage space.
- The theory has non-trivial things to say about a wide range of observations and concepts in artificial intelligence, mainstream computing, mathematics, and human perception and cognition (Appendix I-H) and it has many potential benefits and applications (Appendix I-I).

It appears that the SP theory avoids what are perhaps the two most common pitfalls in the development of scientific theories, noted above: it is not over-specific and it is not over-general. There appears to be a good balance between simplicity and power.

C. IF IT WORKS, WHO CARES?

Some people may argue that these concerns are misplaced—that researchers should concentrate on creating things that work and not worry about the development of good theory. It is true that a suck-it-and-see approach can produce useful results and may indeed be helpful in the development of theory. But no theory or bad theory is almost always a handicap: imagine the difficulties of space travel using Ptolemy’s epicycles as a guide, or the impoverishment of biology without an understanding of DNA.

D. SINCE THE HUMAN MIND IS A KLUGE, WHY WORRY ABOUT GOOD THEORY FOR AI?

Marcus has argued persuasively [36] that, as a result of the way biological evolution builds on what comes to hand, the human mind is, in many respects, a kluge, without the coherence or elegance of a well-designed piece of engineering. In keeping with that idea, Minsky [40] writes: “What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.” (p. 308).

Since human perception and cognition is a source of inspiration and a touchstone of success for AI, some people may conclude from arguments like those just mentioned that we need not worry about good theory for AI: just bundle together a collection of applications from different areas of AI. But this is really a counsel of despair which should not distract us:

- Although it is clear that the human mind has many shortcomings, it also has extraordinary versatility and adaptability. This is still a major challenge for AI, a challenge which appears to demand the development of good theory.
- A kluge of different applications is unlikely to provide the smooth inter-working of different kinds of knowledge and different aspects of intelligence that appears to be essential if, in AI, we are to achieve human-like versatility and adaptability in intelligence (Section II-B).
- Much progress in science has depended on a willingness to look for simplicity within the apparent complexity of the world, witness Newton’s laws of motion. We should not give up on the quest for good theory merely because it is difficult.

Of course, it is known that different areas of the brain are specialised for different functions. But there is no contradiction between that observation and the likelihood that human intelligence depends on powerful general principles. The SP concepts are promising candidates for those roles.

APPENDIX III LONG-TERM AND SHORT-TERM PERSPECTIVES

In research and development in ICT, it is often possible to achieve useful results on relatively short timescales. For example, Google Translate delivers translations between natural languages that are often useful even though they normally fall short of what can be achieved by a good human translator.

To achieve higher standards and crack harder problems, an over-enthusiastic focus on short-term results can be more costly than a more strategic approach. Any one idea that shows a little promise may turn out to be a blind alley (in the same way that climbing a tree, while it takes us closer to the moon, is not a means of reaching the moon [8]). Big bets on ideas like that can mean big losses. It is probably better to maintain a broad view unless or until there is sufficient evidence to justify a narrower focus.

The SP system, itself the product of a 20-year programme of research, draws on earlier research developing computer models of language learning. There is now much evidence in support of the framework but, as noted in Appendix I-J, there are still many avenues to be explored. Hence the proposal to create a new research facility to enable researchers anywhere to explore what can be done with the SP machine and to create new versions of it. Such a development would be a relatively inexpensive way of taking things forward on a

broad front, very much in keeping with the quote from [26] in Section IX-C: “The creation of this new era of [cognitive] computing is a monumental endeavor . . .”.

REFERENCES

- [1] J. D. Barrow, *PI in the Sky: Counting, Thinking, and Being*. Harmondsworth, U.K.: Penguin, 1992.
- [2] A. Church, *The Calculi of Lambda Conversion* (Annals of Mathematics Studies), vol. 6. Princeton, NJ, USA: Princeton Univ. Press, 1941.
- [3] W. F. Clocksin and C. S. Mellish, *Programming in Prolog: Using the ISO Standard*. Berlin, Germany: Springer-Verlag, 2003.
- [4] G. E. Dahl, D. Yu, L. Deng, and A. Acero, “Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, pp. 30–42, Jan. 2012.
- [5] Y. N. Dauphin and Y. Bengio, “Big neural networks waste capacity,” Dept. d’informatique Recherche Opérationnelle, Univ. Montréal, Montréal, QC, Canada, Tech. Rep., Mar. 2013. [Online]. Available: <http://arxiv.org/pdf/1301.3583.pdf>
- [6] A. d’Avila Garcez et al., “Neural-symbolic learning and reasoning: Contributions and challenges,” in *Proc. AAAI Spring Symp. Knowl. Represent. Reason., Integr. Symbolic Neural Approaches*, Mar. 2015, pp. 1–4.
- [7] E. Davis and G. Marcus, “Commonsense reasoning and commonsense knowledge in artificial intelligence,” *Commun. ACM*, vol. 58, no. 9, pp. 92–103, 2015.
- [8] H. L. Dreyfus, *What Computers Still Can’t Do—A Critique of Artificial Reason*. New York, NY, USA: MIT Press, 1992.
- [9] C. Edwards, “Growing pains for deep learning,” *Commun. ACM*, vol. 58, no. 7, pp. 14–16, 2015.
- [10] D. A. Ferrucci, “Introduction to ‘this is Watson,’” *IBM J. Res. Develop.*, vol. 56, nos. 3–4, pp. 1–15, 2012.
- [11] M. Gagliolo, “Universal search,” *Scholarpedia*, vol. 2, no. 11, p. 2575, 2007.
- [12] Y. Ganin and V. Lempitsky, “ N^4 -fields: Neural network nearest neighbor fields for image transforms,” presented at the 12th Asian Conf. Comput. Vis., Singapore, Nov. 2014.
- [13] B. Goertzel, “CogPrime: An integrative architecture for embodied artificial general intelligence,” Open Cognit. Project, Tech. Rep. 2012-10-02, 2012. [Online]. Available: <http://bit.ly/1EyOOlb>
- [14] A. Graves, G. Wayne, and I. Danihelka, “Neural turing machines,” Google DeepMind, London, U.K., Tech. Rep., Dec. 2014. [Online]. Available: <http://arxiv.org/pdf/1410.5401.pdf>
- [15] W. E. L. Grimson, “A computer implementation of a theory of human stereo vision,” *Philos. Trans. Roy. Soc. London B, Biol. Sci.*, vol. 292, no. 1058, pp. 217–253, 1981. [Online]. Available: <http://bit.ly/1Jq8zYk>
- [16] C. G. Gross, “Genealogy of the ‘grandmother cell,’” *Neuroscientist*, vol. 8, no. 5, pp. 512–518, 2002.
- [17] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. New York, NY, USA: Wiley, 1949.
- [18] S. Herculano-Houzel, “The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost,” *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 1, pp. 10661–10668, 2012.
- [19] I. Horrocks, “Ontologies and the semantic Web,” *Commun. ACM*, vol. 51, no. 12, pp. 58–67, 2008.
- [20] D. H. Hubel, “Exploration of the primary visual cortex, 1955–78,” in *Cognitive Neuroscience: A Reader*, M. S. Gazzaniga, Ed. Oxford, U.K.: Blackwell, 2000, pp. 58–80.
- [21] M. Hutter, “The fastest and shortest algorithm for all well-defined problems,” *Int. J. Found. Comput. Sci.*, vol. 13, no. 3, pp. 431–443, 2002.
- [22] W. Isaacson, *Einstein: His Life and Universe*. London, U.K.: Pocket Books, 2007.
- [23] A. G. Ivakhnenko and G. A. Ivakhnenko, “The review of problems solvable by algorithms of the group method of data handling (GMDH),” *Pattern Recognit. Image Anal.*, vol. 5, no. 4, pp. 527–535, 1995.
- [24] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, “On using very large target vocabulary for neural machine translation,” in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics, 7th Int. Joint Conf. Asian Fed. Natural Lang. Process. (ACL-IJCNLP)*, 2015, pp. 1–10.
- [25] B. Julesz, *Foundations of Cyclopean Perception*. Chicago, IL, USA: Chicago Univ. Press, 1971.
- [26] J. E. Kelly and S. Hamm, *Smart Machines: IBM’s Watson and the Era of Cognitive Computing*. New York, NY, USA: Columbia Univ. Press, 2013.
- [27] P. Klinov and B. Parsia, “Pronto: A practical probabilistic description logic reasoner,” in *Uncertainty Reasoning for the Semantic Web II* (Lecture Notes in Computer Science), vol. 7123. Heidelberg, Germany: Springer, 2013, pp. 59–79.
- [28] J. E. Laird, *The Soar Cognitive Architecture*. Cambridge, MA, USA: MIT Press, 2012.
- [29] Q. V. Le et al., “Building high-level features using large scale unsupervised learning,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 8595–8598.
- [30] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, May 2015.
- [31] H. J. Levesque, E. Davis, and L. Morgenstern, “The Winograd schema challenge,” in *Proc. 13th Int. Conf. Principles Knowl. Represent. Reason.*, 2012, pp. 1–6.
- [32] L. A. Levin, “Universal sequential search problems,” *Problemy Peredachi Inf.*, vol. 9, no. 3, pp. 115–116, 1973.
- [33] L. A. Levin, “Randomness conservation inequalities; information and independence in mathematical theories,” *Inf. Control*, vol. 61, no. 1, pp. 15–37, 1984.
- [34] D. S. Levine and M. Aparicio, IV, Eds., *Neural Networks for Knowledge Representation and Inference*. Hillsdale, NJ, USA: Psychology Press, 1994.
- [35] M. Li and P. M. B. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed. New York, NY, USA: Springer, 2014.
- [36] G. Marcus, *Kluge: The Haphazard Evolution of the Human Mind*. London, U.K.: Faber and Faber, 2008.
- [37] D. Marr and T. Poggio, “A computational theory of human stereo vision,” *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 204, no. 1156, pp. 301–328, 1979.
- [38] P. McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, 2nd ed. Natick, MA, USA: A K Peters, Ltd., 2004.
- [39] P. M. Milner, “Neural representations: Some old problems revisited,” *J. Cognit. Neurosci.*, vol. 8, no. 1, pp. 69–77, 1996.
- [40] M. Minsky, *The Society of Mind*. New York, NY, USA: Simon and Schuster, 1986.
- [41] A. Mordvintsev, C. Olah, and M. Tyka, “Inceptionism: Going deeper into neural networks,” Google Inc., Mountain View, CA, USA, Tech. Rep. 2015-07-13, 2015. [Online]. Available: <http://bit.ly/1BkXP09>
- [42] U. Neisser, *Cognitive Psychology*. New York, NY, USA: Appleton-Century-Crofts, 1967.
- [43] A. Newell, “You can’t play 20 questions with nature and win: Projective comments on the papers in this symposium,” in *Visual Information Processing*, W. G. Chase, Ed. New York, NY, USA: Academic, 1973, pp. 283–308.
- [44] A. Newell, *Unified Theories of Cognition*. Cambridge, MA, USA: Harvard Univ. Press, 1990.
- [45] A. Newell and H. A. Simon, *Human Problem Solving*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1972.
- [46] A. Nguyen, J. Yosinski, and J. Clune, “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 427–436. DOI: 10.1109/CVPR.2015.7298640
- [47] E. L. Post, “Formal reductions of the general combinatorial decision problem,” *Amer. J. Math.*, vol. 65, no. 2, pp. 197–215, 1943.
- [48] J. Rissanen, “Modeling by shortest data description,” *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
- [49] S. Russell, “Unifying logic and probability,” *Commun. ACM*, vol. 58, no. 7, pp. 88–97, 2015.
- [50] A. V. Samsonovich, “Toward a unified catalog of implemented cognitive architectures,” *Biol. Inspired Cognit. Archit.*, vol. 221, pp. 195–244, Aug. 2010.
- [51] J. Schmidhuber, “Ultimate cognition à la Gödel,” *Cognit. Comput.*, vol. 1, no. 2, pp. 177–193, 2009.
- [52] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

- [53] H. Schwenk, "Continuous space translation models for phrase-based statistical machine translation," in *Proc. 24th Int. Conf. Comput. Linguistics (COLING)*, 2012, pp. 1071–1080.
- [54] T. Simonite, "Teaching machines to understand us," *MIT Technol. Rev.*, vol. 118, no. 5, pp. 70–77, 2015.
- [55] R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng, "Parsing with compositional vector grammars," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, 2013, pp. 455–465.
- [56] R. J. Solomonoff, "A formal theory of inductive inference. Parts I and II," *Inf. Control*, vol. 7, pp. 1–22 and 224–254, 1964.
- [57] R. J. Solomonoff, "The application of algorithmic probability to problems in artificial intelligence," in *Uncertainty in Artificial Intelligence*, L. N. Kanal and J. F. Lemmer, Eds. Amsterdam, The Netherlands: Elsevier, 1986, pp. 473–491.
- [58] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014.
- [59] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 27, 2014, pp. 3104–3112.
- [60] C. Szegedy *et al.*, "Intriguing properties of neural networks," Google Inc., Mountain View, CA, USA, Tech. Rep., Feb. 2014. [Online]. Available: <http://bit.ly/1elzRGM>
- [61] A. M. Turing, "On computable numbers, with an application to the entscheidungsproblem," *Proc. London Math. Soc.*, vol. 42, no. 2, pp. 230–265 and 544–546, 1936.
- [62] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, Oct. 1950.
- [63] A. M. Turing, "Intelligent machinery," in *The Collected Works of AM Turing: Mechanical Intelligence*, D. C. Ince, Ed. Amsterdam, The Netherlands: Elsevier, 1992. [Online]. Available: <http://bit.ly/1SeLmhO>
- [64] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proc. 31st Int. Conf. Mach. Learn. (ICML)*, 2014, pp. 3156–3164. DOI: 10.1109/CVPR.2015.7298935
- [65] C. S. Wallace and D. M. Boulton, "An information measure for classification," *Comput. J.*, vol. 11, no. 2, pp. 185–195, 1968.
- [66] S. Wiesler, A. Richard, R. Schlüter, and H. Ney, "Mean-normalized stochastic gradient for large-scale deep learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 180–184.
- [67] T. Winograd, "Understanding natural language," *Cognit. Psychol.*, vol. 3, no. 1, pp. 1–191, 1972.
- [68] J. G. Wolff, "Learning syntax and meanings through optimization and distributional analysis," in *Categories and Processes in Language Acquisition*, Y. Levy, I. M. Schlesinger, and M. D. S. Braine, Eds. Hillsdale, NJ, USA: Lawrence Erlbaum, 1988, pp. 179–215. [Online]. Available: <http://bit.ly/ZIGjyc>
- [69] J. G. Wolff, "Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, unification and search," *Decision Support Syst.*, vol. 42, pp. 608–625, Nov. 2006. [Online]. Available: <http://bit.ly/XE7pRG>
- [70] J. G. Wolff, *Unifying Computing and Cognition: The SP Theory and Its Applications*. Menai Bridge, U.K.: CognitionResearch.org, 2006. [Online]. Available: <http://bit.ly/WmB1rs>
- [71] J. G. Wolff, "Towards an intelligent database system founded on the SP theory of computing and cognition," *Data Knowl. Eng.*, vol. 60, no. 3, pp. 596–624, 2007. [Online]. Available: <http://bit.ly/Yg2onp>
- [72] J. G. Wolff, "The SP theory of intelligence: An overview," *Information*, vol. 4, no. 3, pp. 283–341, 2013. [Online]. Available: <http://bit.ly/1hz0lFE>
- [73] J. G. Wolff, "Application of the SP theory of intelligence to the understanding of natural vision and the development of computer vision," *SpringerPlus*, vol. 3, no. 1, pp. 552–570, 2014. [Online]. Available: <http://bit.ly/1scmpV9>
- [74] J. G. Wolff, "Autonomous robots and the SP theory of intelligence," *IEEE Access*, vol. 2, no. 1, pp. 1629–1651, Jan. 2014. [Online]. Available: <http://bit.ly/1zrSemu>
- [75] J. G. Wolff, "Big data and the SP theory of intelligence," *IEEE Access*, vol. 2, no. 4, pp. 301–315, Apr. 2014. [Online]. Available: <http://bit.ly/1jGWXDH>
- [76] J. G. Wolff, "Information compression, intelligence, computing, and mathematics," CognitionResearch.org, Menai Bridge, U.K., Tech. Rep. 2015-06-04, 2015. [Online]. Available: <http://bit.ly/1jEoECH>
- [77] J. G. Wolff, "The SP theory of intelligence: Benefits and applications," *Information*, vol. 5, no. 1, pp. 1–27, 2014. [Online]. Available: <http://bit.ly/1lcquWF>
- [78] J. G. Wolff and V. Palade, "Proposal for the creation of a research facility for the development of the SP machine," CognitionResearch.org, Menai Bridge, U.K., Tech. Rep. 2015-11-16, 2015. [Online]. Available: <http://bit.ly/1zzZjls>
- [79] Q. Yu, Y. Yang, Y. Song, T. Xiang, and T. Hospedales, "Sketch-a-net that beats humans," Queen Mary College, Univ. London, London, U.K., Tech. Rep., May 2015. [Online]. Available: <http://arxiv.org/pdf/1501.07873v2.pdf>
- [80] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
- [81] X. Zeng, W. Ouyang, and X. Wang, "Multi-stage contextual deep learning for pedestrian detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 180–184.
- [82] X. Zhang and Y. LeCun, "Text understanding from scratch," Dept. Comput. Sci., New York Univ., New York, NY, USA, Tech. Rep., Jun. 2015. [Online]. Available: <http://arxiv.org/pdf/1502.01710v3.pdf>
- [83] Y. Zhu, Y. Tian, D. Mexatas, and P. Dollár, "Semantic amodal segmentation," Dept. Comput. Sci., Rutgers Univ., New Brunswick, NJ, USA, 2015.



JAMES GERARD WOLFF (M'13) received the Natural Sciences Tripos degree from Cambridge University, Cambridge, U.K., specializing in experimental psychology, and the Ph.D. degree in psychology from the University of Wales, Cardiff. He has held academic positions with the School of Informatics, University of Bangor, U.K., the Department of Psychology, University of Dundee, U.K., and the University Hospital of Wales, Cardiff, U.K. He has also held a Research Fellowship with IBM, Winchester, U.K., and has been a Software Engineer with Praxis Systems plc, Bath, U.K. Since 1987, his research has been focused on the development of the SP theory of intelligence. He is currently the Director of CognitionResearch.org., Menai Bridge, U.K. He has numerous publications in a wide range of academic journals, collected papers, and conference proceedings. Previously, his main research interests were in developing computer models of language learning. He is also a Chartered Engineer and member of the British Computer Society (Chartered IT Professional).

• • •