Unsolved problems in AI, described in the book "Architects of Intelligence" by Martin Ford, and how they may be solved via the SP System

J Gerard Wolff[*,†]

Abstract

The book *Architects of Intelligence* by Martin Ford presents conversations about AI between the author and influential researchers. In this paper, unsolved problems in AI research that are described in the book are considered in relation to features of the *SP System*, outlined in an appendix with pointers to where fuller information may be found. This analysis shows that *The SP System has clear potential to help solve at least 10 problems in AI described in the book, and some others.* Strengths and potential of the SP System include: the representation and processing of both symbolic and non-symbolic kinds of knowledge, and their integration; strengths and long-term potential in pattern recognition; freedom from the tendency of deep neural networks (DNNs) to make large and unexpected errors in recognition; strengths and potential in unsupervised learning; a theoretically coherent basis for generalisation; potential to help improve the safety of driverless cars; the ability to achieve useful learning from a single occurrence or experience; relatively tiny demands for computational resources and volumes of data, with potential for high speeds of learning; strengths in transfer learning; transparency in the representation of knowledge and an audit trail for all its processing; strengths and potential in the processing of natural language; several different kinds of probabilistic reasoning; strengths and potential in commonsense reasoning and the representation of commonsense knowledge; information compression; biological validity; scope for adaptation; and freedom from catastrophic forgetting. Together, the strengths and potential of the SP System suggest that *the SP System provides a firmer foundation for the development of human-level general AI than any alternative.* In developing the SP System, no attempt has yet been made to model motivations and emotions, an area of interest that is mentioned by more than one of the researchers that Ford consulted.

[*]CognitionResearch.org, UK

[†]Corresponding author. *Email address:* jgw@cognitionresearch.org (Gerry Wolff)

1 Introduction

The book *Architects of Intelligence* by Martin Ford [7], presents interviews that Ford has held with people who are influential in AI research, and yields some fascinating insights into current thinking of leaders in the field.

This paper discusses problems in AI described in the book, and some others, in relation to features of the *SP System*, meaning the *SP Theory of Intelligence* and its realisation in the *SP Computer Model*, both described in outline in Appendix A and in more detail in sources referenced there.

Thus the paper is not a review of Ford's book as such, and its organisation is unusual: the paper is mainly an extended response to the very interesting material that Ford has gathered in his book; that is what dictates the form of the paper. Although the form of the paper is unusual, it attempts to perform an important task in any field of science: the evaluation of problems in the field and how they may be solved.

After a 'background' section, next, the main sections that follow describe a range of problems in AI described by those leading researchers in AI as reported in [7], and some others. For each problem, the paper describes how the SP System may help solve it. There is clear potential for the SP System solve most and perhaps all of those problems. This is good evidence that *the SP System provides a firmer foundation for the development of human-level general AI than any alternative.*

2 Background

This 'background' section first outlines some key features of the SP System and how it has been developed. And, second, because 'deep neural networks' (DNNs) are discussed in many parts of [7], there are some introductory remarks about them.

2.1 Key features of the SP System and its development

This section introduces the main features of the SP System and its origins. There is more detail in Appendix A, with pointers to where fuller information may be found.

2.1.1 Conceptual *Simplicity* and descriptive or explanatory *Power*

As described in Appendix A.1, *the SP programme of research is a unique attempt to simplify and integrate observations and concepts across artificial intelligence,*

mainstream computing, mathematics, and human learning, perception, and cognition (HLPC). In accordance with Ockham's razor, this means trying to develop a system which combines conceptual *Simplicity* with descriptive or explanatory *Power*. Hence the name "SP". A second reason for that name is given in Section 2.1.3.

Despite its ambition, this strategy has been largely successful, leading to the creation of a system that combines relative simplicity with strengths and potential in diverse aspects of intelligence (Appendix A.9.1), including several kinds of reasoning (Appendix A.9.2), and in the representation of diverse kinds of knowledge (Appendix A.9.3). It also facilitates the *seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination* (Appendix A.9.4). 'Seamless integration ...' may be abbreviated as 'SI'.

Because the SP System is the product of research aiming to simplify and integrate observations and concepts across a broad canvass, it will inevitably have points of similarity with existing systems. But any such similarities should not distract attention from how the SP System combines conceptual Simplicity with descriptive and explanatory Power, meaning that one relatively simple system can do many different things. This is one of the most important advantages of the SP System compared with: 1) systems that can do only one or two kinds of things; or 2) other systems which achieve versatility via an inelegant kluge of different capabilities.

2.1.2 Development of the SP System

The SP System has not been dreamt up overnight. It is the product of a lengthy programme of research, from about 1987 to now with a break between early 2006 and late 2012. This programme of research has included the creation and testing of many versions of the SP Computer Model. A major discovery has been the concept of *SP-multiple-alignment* and its versatility in many aspects of intelligence (Section 2.1.4, Appendix A.4).

Within the SP System, the SP-multiple-alignment concept is largely responsible for the versatility of the SP System in many aspects of intelligence and the representation of diverse kinds of knowledge, and potential for their seamless integration (Appendix A.9.1). Potential applications of the SP System have also been examined. Many of these are described in papers that may be accessed via links in www.cognitionresearch.org/sp.htm.

It must be stressed that Appendix A is only an outline of the SP System and its potential applications. There is much more detail in [39], and even more detail in the book *Unifying Computing and Cognition* [37]. This paper has been written with the aim of making it freestanding, so that, without undue length or repetition of what has been published already, it may be understood as it stands.

As with any research, some results are solid and others are more conjectural. In this paper, aspects of the SP System which are solidly established are described as "strengths", while conjectural features—with some foundation but less certain than "strengths"—are described as "potential". As with other research, conjectural or potential features are important as a guide for future investigations and development. Pure speculations, without any foundation, have been omitted.

2.1.3 A central role for information compression

The SP System incorporates a key principle from human cognitive psychology: that much of HLPC may be understood as information compression (IC). Evidence for that principle is described in [50]. It seems likely that the same principle may also apply in neuroscience [45].

As noted in Appendix A.2, the principle accords with concepts of Minimum Length Encoding (MLE) (see, for example, [16]), although there are important differences between the SP Theory and MLE concepts.

It has been recognised for some time that IC is closely related to concepts of probability (Appendix A.3). This makes it relatively straightforward to calculate absolute and relative probabilities of inferences that are made by the SP Computer Model.

In the SP Computer Model, IC is achieved via the building of SP-multiple-alignments (Appendix A.4 and Section 2.1.4), and via unsupervised learning (Appendix A.5 and Section 7).

IC may be seen to be a process of maximising *Simplicity* in a body of information **I**, by reducing redundancy in **I**, whilst at the same time retaining as much as possible of its non-redundant expressive *Power*. This is a second reason for the name 'SP', additional to that mentioned near the beginning of Section 2.1.1.

Before leaving this subsection, it is pertinent to note that IC is an important feature of a type of artificial neural network known as an *autoencoder* (see, for example, [5, pp. 115–117]). As noted in Section 18, with comments, some research with autoencoders in DNNs is continuing today.

2.1.4 SP-multiple-alignment

The third main feature of the SP System to be mentioned here is the simple but powerful concept of *SP-multiple-alignment*, outlined in Appendix A.4.

The concept of SP-multiple-alignment is largely responsible for the SP System's versatility in aspects of intelligence including several kinds of reasoning, its versatility in the representation of diverse kinds of knowledge, and for its potential for SI (all summarised in Appendix A.9).

Bearing in mind that, in a presentation like this, it is just as bad to underplay the strengths and potential of a system as it is to oversell its strengths and potential, it seems fair to say that *the concept of SP-multiple-alignment may prove to be as significant for an understanding of 'intelligence' as is DNA for biological sciences: it may prove to be the 'double helix' of intelligence.*

2.1.5 SP-Neural

Abstract concepts in the SP Theory of Intelligence map quite well into structures and processes represented in terms of neurons and their interconnections (Appendix A.7 and [45]). It is anticipated that, as with the non-neural SP Theory, SP-Neural will be developed via the creation and testing of computer models—to reduce vagueness in the SP-Neural concepts, to weed out ideas that do not work, and to demonstrate what can be done with the SP-Neural version of the SP Theory.

2.2 Deep learning and its strengths and weaknesses

> "The vast majority of the dramatic advances we've seen over the past decade or so—everything from image and facial recognition, to language translation, to AlphaGo's conquest of the ancient game of Go—are powered by a technology known as deep learning, or deep neural networks." Martin Ford [7, p. 3].
>
> "The remaining conversations in this book are generally with people who might be characterized as deep learning agnostics, or perhaps even critics. All would acknowledge the remarkable achievements of deep neural networks over the past decade, but they would likely argue that deep learning is just 'one tool in the toolbox' and that continued progress will require integrating ideas from other spheres of artificial intelligence." Martin Ford [7, pp. 4–5].

The first of these two quotes acknowledges the undoubted successes of DNNs while the other suggests that there are "deep learning agnostics, or perhaps even critics." That seems to be a reasonably accurate reflection of the range of views about deep learning amongst AI researchers today.

But, bearing in mind that there is quite a variety of DNNs and that there is research going on which is aiming to improve DNNs, there are several shortcomings in at least some of the kinds of DNN. Some of those shortcomings are outlined in [46, Section V], with some brief discussion of how those shortcomings may be overcome. Those discussions are less well developed than in this paper.

Because of the dominance of DNNs in AI research today, DNNs are the implied or explicit comparison in describing the strengths and potential of the SP System. As will be seen in the main sections that follow, *the SP System has many advantages compared with DNNs*.

That said, there is no suggestion here that all the resources that are currently devoted to the development of DNNs should be switched to the development of the SP System. If we are to solve a difficult problem like the development of human-level general AI, we should not put all our eggs in one basket. With only a small part of the budget devoted to DNNs, the SP System may be developed quite well alongside further research on DNNs.

3 A top-down strategy

This section, and all the sections that follow before the Conclusion (Section 24), discuss problems in the development of AI as it is today and the clear potential of the SP System to solve them.

This section, about the need for a top-down strategy in AI research, begins with a quote as follows:

> "... one of [the] stepping stones [towards Artificial General Intelligence (AGI)] would be an AI program that can really handle multiple, very different tasks. An AI program that's able to both do language and vision, it's able to play board games and cross the street, it's able to walk and chew gum. Yes, that is a joke, but I think it is important for AI to have the ability to do much more complex things." Oren Etzioni [7, p. 502].

An implication of this quote is that researchers should be placing more emphasis on the search for mechanisms and processes that can combine conceptual Simplicity with descriptive or explanatory Power, as in Section 2.1.1 and Appendix A.1. This in turn suggests that there may be merits in adopting a top-down strategy, working from overarching principles to lower-level concepts, and putting less effort into the popular bottom-up strategies which try to work from low-level concepts to overarching principles.

There has been and still is research with top-down strategies: the quest for 'unified theories of cognition' (see, for example, [21] and later research such as [13]); and research aiming for 'artificial general intelligence' (AGI, see, for example, [12]).

But it appears to be fair to say that, out of the several such systems that have been developed:

- The SP System is unique in seeking to achieve a favourable combination of Simplicity and Power (Appendix A.1).
- More specifically, the SP System is unique in its quest for a framework that would simplify and integrate observations and concepts across an exceptionally broad canvas: across AI, mainstream computing, mathematics, and human learning, perception, and cognition (Section 2.1.1).
- And the SP System has achieved relative success in its versatility across diverse aspects of intelligence, the representation of diverse kinds of knowledged, and in the seamless integration of diverse aspects of knowledge and the representation of diverse kinds of knowledge in any combination (Appendix A.9.1).

Top-down and bottom-up strategies are both difficult: there is no royal road to success. But in AI research today there appears to be a case for giving more attention to top-down strategies.

4 Bridging symbolic and non-symbolic kinds of representation and processing

> "Some people still believe in symbolic AI, and they think there's potentially a need for a hybrid approach that incorporates both deep learning and more traditional approaches." Martin Ford [7, p. 84].
>
> "In my view, we need to bring together symbol manipulation, which has a strong history in AI, with deep learning. They have been treated separately for too long, and it's time to bring them together." Gary Marcus [7, p. 318].
>
> "Many people will tell a story that in the early days of AI we thought intelligence was symbolic, but then we learned that was a terrible idea. It didn't work, because it was too brittle, couldn't handle noise and couldn't learn from experience. So we had to get statistical, and then we had to get neural. I think that's very much a false narrative. The early ideas that emphasize the power of symbolic reasoning and abstract languages expressed in formal systems were incredibly important and deeply right ideas. I think it's only now that we're in the position, as a field, and as a community, to try to understand how to bring together the best insights and the power of these different paradigms." Josh Tenenbaum [7, pp. 476–477].

With regard to the kinds of issues mentioned in the quotes above, the SP System is consistent with the following views:

- Given that people can and do learn and use symbolic systems like natural languages, mathematics, and logic, it is inconceivable that our brains do not have the capacity to represent and process those kinds of symbolic systems. It seems likely that the kinds of structures that are recognised by theoretical and computational linguists, mathematicians, and logicians will have some counterpart in our brains.

- At the same time, any theory of human-level AI should be true to the kinds of things that can be done by: babies before they learn any natural language, or mathematics or logic; or adults in performing putatively 'non-symbolic' skills such as recognising things, playing tennis, and so on; or animals when they are engaged in putatively 'non-symbolic' activities such as foraging for food, hunting, swinging from branch to branch through trees, and so on.

- These two points suggest that some kind of hybrid of symbolic and non-symbolic kinds of system is the way to go, perhaps a hybrid of DNNs with something more symbolic. But DNNs suffer from the several shortcomings that are detailed in this paper (Section 2.2). Something else is needed that bridges symbolic and non-symbolic kinds of knowledge representation and processing.

- Although the SP System is far from providing a comprehensive account of all these kinds of things, it provides a framework that is showing promise in bridging symbolic and non-symbolic aspects of HLPC:

  - The concepts of 'SP-symbol' in the SP System (Appendix A) can represent a relatively large 'symbolic' kind of thing such as a word, or it can represent a relatively fine-grained kind of thing such as a pixel in an image.
  - The concept of 'SP-multiple-alignment' (Appendix A.4), with the concept of SP-pattern (Appendix A) provides a versatile framework for the representation of diverse kinds of knowledge (Appendix A.9.3) and for the processing of that knowledge in diverse aspects of intelligence (Appendices A.9.1 and A.9.2).
  - The concept of SP-multiple-alignment also facilitates the seamless integration of diverse kinds of knowledge and diverse aspects of intelligence, in any combination (Appendix A.9.4), a kind of integration that appears to be essential in any system that aspires to human-level AI.

– In the 'neural' version of the SP Theory called 'SP-Neural' (Section 2.1.5, Appendix A.7, and [45]), abstract constructs in the SP System—SP-symbols, SP-patterns, and, within SP-multiple-alignments, connections amongst them—may be mapped into corresponding structures expressed in terms of neurons and their inter-connections. Each such structure is, in some respects, similar to Donald Hebb's concept of a 'cell assembly' 19 but, as noted in Section 10, unsupervised learning in the SP System is quite different from the gradualist kinds of learning which, with variations, are prominent in learning by DNNs.

– As noted in Appendix A.3, the SP System is fundamentally probabilistic so in that respect it sits comfortably with the probabilistic nature of most of HLPC. But, when probabilities are at or near 0 or 1, the SP System has potential to imitate the all-or-nothing nature of much of mathematics and logic.

– Unlike other systems, the SP System has compression of information as its unifying principle (Section 18), a principle which applies to both symbolic and non-symbolic knowledge. In accordance with Ockham's razor, IC equates with a search for a favourable combination of conceptual Simplicity with descriptive and explanatory Power (Section 2.1.1 and Appendix A.1).

4.1 Parallel distributed processing?

An issue that is associated with the symbolic / non-symbolic distinction is whether or not the brain is fundamentally a system for 'parallel distributed processing' [27, 28], meaning that, much like DNNs, the processing of information in the brain is highly parallel and that the knowledge of any particular concept such as 'house' or 'tree' is widely distributed in the brain, not localised in a small part of the brain.

In the SP research, it accepted that processing in the brain is highly parallel but that:

- Knowledge is localised in the sense that a given concept, such as 'car', is represented by a single neuron or, more likely, a localised cluster of neurons. Whether or not there are enough neurons in the brain to support this manner of representing knowledge is discussed in [37, Section 11.4.9] and [45, Section 4.4].

- Knowledge is distributed in the sense that, with our 'car' example, each of the many components of a car would each be represented in its own location

(and connected directly or indirectly via nerve fibres to the main node for 'car') so that, overall, the concept of car would be widely distributed.

There is more discussion of this issue in [45, Section 5.8].

5 Recognition of images and speech

> "The vast majority of the dramatic advances we've seen over the past decade or so—everything from image and facial recognition, to language translation, to AlphaGo's conquest of the ancient game of Go—are powered by a technology known as deep learning, or deep neural networks." Martin Ford [7, p. 3].

With some qualification (Section 6), DNNs do well in tasks such as the recognition of images and speech. But there is potential for the SP System in both areas with the completion of some 'unfinished business' (Appendix A.11). How the SP System may be developed for computer vision and scene analysis is described in [40].

For the achievement of human-like abilities in perception, an attractive feature of the SP System is that it has robust abilities to achieve what are intuitively correct analyses of incoming information despite errors of omission, commission and substitution (Section 8.2, [39, Section 4.2.2], [38, Section 2.2.2]).

Another attractive feature of the SP System which chimes with human psychology is that, in recognising a particular person (say "John"), we are instantly aware without conscious effort that John, as a person, is likely to have the characteristics of people, and if we were trained in zoology, our knowledge would probably extend to the characteristics of mammals, vertebrates, and so on.

The way in which the SP System may achieve recognition at several different levels of abstraction is described, with examples, in [39, Section 9.1].

6 Deep neural networks are easily fooled

> "In [a recent] paper [2], [the authors] show how you can fool a deep learning system by adding a sticker to an image. They take a photo of a banana that is recognized with great confidence by a deep learning system and then add a sticker that looks like a psychedelic toaster next to the banana in the photo. Any human looking at it would say it was a banana with a funny looking sticker next to it, but the deep learning system immediately says, with great confidence, that it's now a picture of a toaster." Gary Marcus [7, p. 318].

There are several reports describing how DNNs can make surprisingly big and unexpected errors in recognition. For example, a DNN 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 [33]. It has been reported that a DNN may assign an image with near certainty to a 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 [22].

More recently:

> "... it is relatively easy to force [DNNs] to make mistakes that seem ridiculous, but with potentially catastrophic results. Recent tests have shown autonomous vehicles could be made to ignore stop signs, and smart speakers could turn seemingly benign phrases into malware. ... tiny changes to many of the pixels in an image could cause DNNs to change their decisions radically; a bright yellow school bus became, to the automated classifier, an ostrich. But the changes made were imperceptible to humans." [6, p. 13].

and

> "Convolutional Neural Networks (CNNs) are commonly thought to recognise objects by learning increasingly complex representations of object shapes. Some recent studies suggest a more important role of image textures. ..." [8, Abstract].

Of course, there is potential for improvement in the performance of DNNs. But it seems that there are limits to what can be achieved because of fundamental weaknesses in DNNs: that they represent and process knowledge in a way that is not transparent (Section 14), and that their organisation and workings are not psychologically valid.

Experience with the SP System to date, and its transparency in both the representation and processing of knowledge (Section 14), suggests that it would not be vulnerable to the kinds of mistakes made by DNNs.

7 The importance of unsupervised learning

> "Unsupervised learning represents one of the most promising avenues for progress in AI. ... However, it is also one of the most difficult challenges facing the field. *A breakthrough that allowed machines to efficiently learn in a truly unsupervised way would likely be considered one*

> *of the biggest events in AI so far, and an important waypoint on the road to human-level AI.*" Martin Ford [7, pp. 11–12], emphasis of the last sentence added.
>
> "Until we figure out how to do this unsupervised/self-supervised/predictive learning, we're not going to make significant progress because I think that's the key to learning enough background knowledge about the world so that common sense will emerge." Yann Lecun [7, p. 130].
>
> "Unsupervised learning is hugely important, and we're working on that." Demis Hassabis [7, p. 170].

Judging by these remarks, the development of unsupervised learning is important in AI today. So it should be of interest to researchers in AI that learning by the SP Computer Model (Appendix A.5) is entirely unsupervised.

Although there are (at least) two shortcomings in how the model learns [39, Section 3.3], it appears that these problems are soluble and that "solving them will greatly enhance the capabilities of the system for the unsupervised learning of structure in data." (*ibid.*).

Unlike DNNs, the SP System has strengths and potential in unsupervised grammatical inference. The SP computer model ([37, Chapter 9], [39, Section 5]) has already demonstrated an ability to discover generative grammars from unsegmented samples of English-like artificial languages, including segmental structures, classes of structure, and abstract patterns.

In the SP programme of research, unsupervised grammatical inference is seen as a foundation for all kinds of unsupervised learning, not merely the learning of syntax. It may, for example, provide for the learning of non-syntactic semantic structures and for learning the integration of syntax with semantics.

And in the SP programme of research, it is recognised that unsupervised learning is likely to be a good foundation for other kinds of learning, such as learning by being told, learning by imitation, learning via rewards and punishments (reinforcement learning), learning via labelled examples (supervised learning), learning via reading, and more [41, Section V-A.1].

8 Generalisation, under-generalisation, and over-generalisation

> "Many of us think that we are ... missing the basic ingredients needed [for generalization], such as the ability to understand causal relationships in data—an ability that actually enables us to generalize and to

come up with the right answers in settings that are very different from those we've been trained in." Yoshua Bengio [7, p. 18].

"... we might have a photograph, where we've got all the pixels in the image, and then we have a label saying that this is a photograph of a boat, or of a Dalmatian dog, or of a bowl of cherries. In supervised learning for this task, the goal is to find a predictor, or a hypothesis, for how to classify images in general." Stuart J. Russell [7, p. 41].

"The theory [worked on by Roger Shepard and Joshua Tenenbaum] was of how humans, and many other organisms, solve the basic problem of generalization, which turned out to be an incredibly deep problem. ... The basic problem is, how do we go beyond specific experiences to general truths? Or from the past to the future?" Joshua Tenenbaum [7, p. 468].

An important issue in learning is how to generalise 'correctly' from the specific information which provides the basis for learning, without over-generalisation ('under-fitting') or under-generalisation ('over-fitting'). This issue is discussed quite fully in [39, Section 5.3] (see also [37, Section 9.5.3] and [46, Section V-H]) but because it is an important issue, the main elements of the solution proposed in the SP Theory of Intelligence are described here.

Generalisation may be seen to occur in two aspects of AI: as part of the process of unsupervised learning; and as part of the process of recognition. Those two aspects are considered in the following two subsections.

8.1 Generalisation via unsupervised learning

The generalisation issue arises most clearly in considering how children learn their native language or languages, as illustrated in Figure 1.

Each child learns a language **L** from a sample of things that they hear being said by people around them, a sample which is normally large but, nevertheless, finite. This is shown as the smallest envelope in the figure.

The variety of 'legal' utterances in the language **L** is represented by the next largest envelope. The largest envelope represents the variety possible utterances, both those in **L** and everything else including grunts, gurgles, and so on.

Each of the two larger envelopes represents a set that is infinite in size but, in accordance with principles pioneered by Georg Cantor, the set of all possible utterances is larger than the set of utterances in **L**. This is much like the way that the set of all integers is larger than the set of even integers, or the set of odd integers, although each of those sets is infinite in size.

The difference in size between the smallest envelope and the middle-sized envelope represents correct generalisations. If a learning system creates a grammar

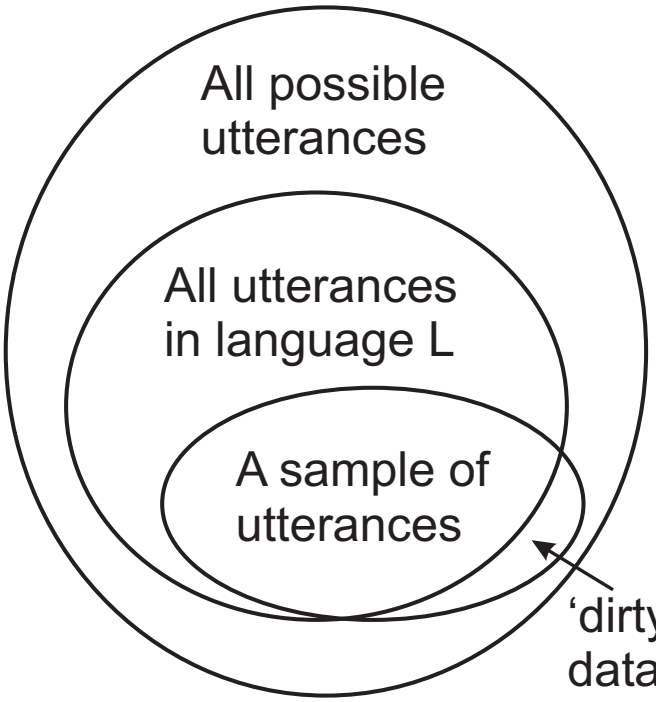

Figure 1: Categories of utterances involved in the learning of a first language, **L**. In ascending order of size, they are: the finite sample of utterances from which a child learns; the (infinite) set of utterances in **L**; and the larger (infinite) set of all possible utterances. Adapted from Figure 7.1 in [35], with permission.

that generates **L** and some other utterances, then it over-generalises, and if the grammar generates a subset of the utterances in **L**, then it under-generalises.

An interesting feature of the learning of a first language by children is that their finite sample of what people are saying includes some things that are *not* in **L** (because of slips of the tongue and the like) as well as many things that are in **L**. In the figure, the 'illegal' utterances that children hear are marked as 'dirty data'.

Challenges in understanding how young children learn their first language are:

- How they generalise correctly without over- or under-generalisation. In this connection, it is interesting that young children often over-generalise (saying things like 'mouses' instead of 'mice'—applying a pluralisation rule too widely—or saying things like 'hitted' instead of 'hit'—applying a past tense rule too widely)—but they normally weed out such over-generalisations when they are older.

- And also how they learn **L** without their learning being distorted or corrupted by dirty data.

Judging by the quotations above and elsewhere in [7], and judging by other writings about language learning and other kinds of learning, there is no coherent theory of generalisation, and over- or under-generalisation, that is widely recognised by researchers in AI or psychology.

For reasons that would take us too far afield to explain, 'nativist' theories of first language learning, such as that proposed by Noam Chomsky [3] and others, will not suffice.

What follows is a summary of what is proposed in the SP Theory of Intelligence, which appears to be robust and well-founded:

1. Unsupervised learning in the SP Theory of Intelligence may be seen as a process of compressing a body of information, **I**, to achieve lossless compression of **I** in a structure **T**, where the size of **T** (represented by t) is as small as can be achieved with the available computational resources.

2. **T** may be divided into two parts: a *grammar*, **G** of size g; and an *encoding* of **I** in terms of **G**, where the encoding is **E** and the size of **E** is e. Clearly, $t = g + e$.

3. Discard **E** and retain **G**.

4. Provided the compression of **I** has been done quite thoroughly, **G** may be seen to be a theory of **I** which generalises 'correctly' beyond **I** without either over- or under-generalisations.

Why should we have more confidence in **G** as a source of 'correct' generalisations than anything else? The best answer at present seems to be that **G** may be seen as a distillation of redundancies in **I**, whereas **E** is largely the non-redundant aspects of **I**. Since redundancy equates largely with repetition, and since repetition provides the basis for inductive inference, **G** may be seen to be the most promising source of generalisations. Each typo, cough, or similar kind of dirty data is normally rare in a given context and will normally be recorded in **E**, not **G**—and so it will be discarded.

Informal tests with a program for unsupervised learning, 'SNPR' [35], and with the SP Computer Model [37, Chapter 9], suggest that both models may learn what are intuitively 'correct' structures, in spite of being supplied with data that is incomplete in the sense that generalisations are needed to produce a 'correct' result, and in spite of being supplied, on other occasions, with data which contains errors that may be seen as 'dirty data'.

In an application like switching a thermostat on and off, or controlling an automatic washing machine, everything can be fully defined and there is no need for generalisations. But with complex activities like driving a car, playing football, or even playing chess, there are far too many possible situations for everything to be fully defined, which means that generalisations are needed.

Evidence to date suggests that the theory of generalisation outlined in this section, which is part of the SP Theory of Intelligence, is likely to provide generalisations that are as close to being optimum as may be achieved with the available computational resources.

8.2 Generalisation via perception

The SP Computer Model has a robust ability to recognise things or to parse natural language despite errors of omission, commission, or substitution in what is being recognised or parsed. Incidentally, it is assumed here that recognition in any sensory modality may be understood largely as parsing, as described in [40].

The example here makes reference, first, to Figure 5 in Appendix A.4, which shows how an SP-multiple-alignment may achieve the effect of parsing the sentence '`f o r t u n e f a v o u r s t h e b r a v e`' in terms of grammatical categories, including words.

To illustrate generalisation via perception, Figure 2 in this section shows how the SP Computer Model, with the New SP-pattern with errors, '`f o t u n e f a v x t o u r s t h e b r y n v e`', may achieve a 'correct' analysis of the sentence despite the errors.

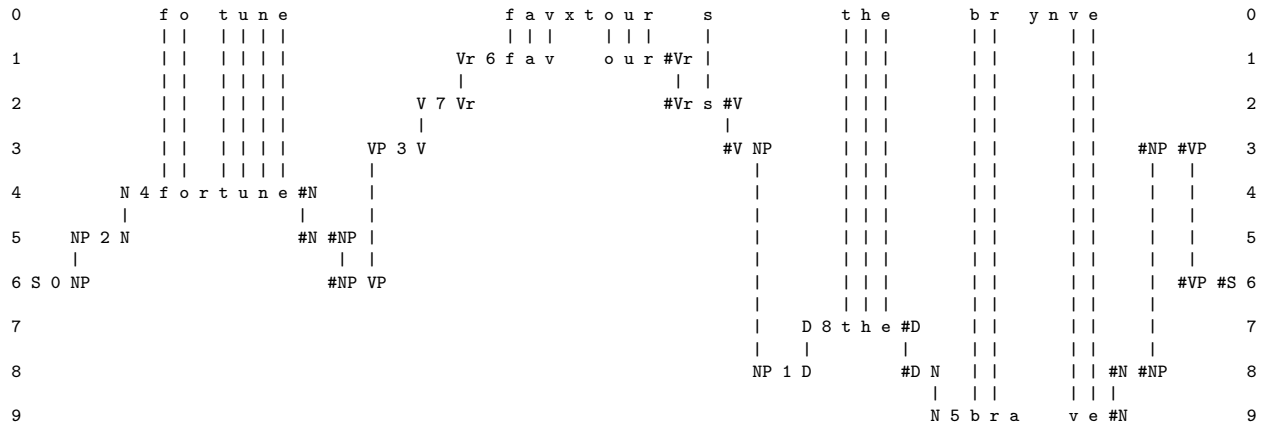

Figure 2: The best SP-multiple-alignment created by the SP Computer Model in the same way as in Figure 5 but with errors in the New SP-pattern representing the sentence to be parsed (row 0). Instead of the New SP-pattern '`f o r t u n e f a v o u r s t h e b r a v e`', there is the New SP-pattern '`f o t u n e f a v x t o u r s t h e b r y n v e`' containing an error of omission ('`f o t u n e`' instead of '`f o r t u n e`'), an addition ('`f a v x t o u r s`' instead of '`f a v o u r s`'), and a substitution ('`b r y n v e`' instead of '`b r a v e`').

This kind of recognition in the face of errors may be seen as a kind of generalisation, where an incorrect form is generalised to the correct form. In the figure: the incorrect form '`f o t u n e`' has been 'generalised' to the correct form '`f o r t u n e`'; the incorrect form '`f a v x t o u r s`' has been 'generalised' to the correct form '`f a v o u r s`'; and the incorrect form '`b r y n v e`' has been 'generalised' to the correct form '`b r a v e`'.

There is relevant discussion in [40, Sections 4.1 and 4.2].

9 Minimising the risk of accidents with driverless cars

> "In the early versions of Google's [driverless] car, ... the problem was that every day, Google found themselves adding new rules. Perhaps they would go into a traffic circle ... and there would be a little girl riding her bicycle the wrong way around the traffic circle. They didn't have a rule for that circumstance. So, then they have to add a new one, and so on, and so on." Stuart J. Russell [7, p. 47].

> "... the principal reason [for pessimism about the early introduction of driverless cars for all situations is] that if you're talking about driving in a very heavy metropolitan location like Manhattan or Mumbai, then the AI will face a lot of unpredictability. It's one thing to have a driverless car in Phoenix, where the weather is good and the population is a lot less densely packed. The problem in Manhattan is that anything goes at any moment, nobody is particularly well-behaved and everybody is aggressive, the chance of having unpredictable things occur is much higher." Gary Marcus [7, p. 321].

A naïve approach to the avoidance of accidents with driverless cars would be to specify stimulus-response pairs, where the stimulus would be a picture of the road in front (perhaps including sounds), and the response would be a set of actions with the steering wheel, brakes, and so on. Of course, driving is far too complex for anything like that to be adequate.

It seems that, for any kind of driver, either human or artificial, some kind of generalisation from experience to date is essential. In that connection, people will have the benefit of all their visual experience prior to their driving lessons, but the same principles apply.

If a person or a driverless car has learned to apply the brakes when a child runs out in front, that learning should be indifferent to the multitude of images that may be seen: the child may be fat or thin; tall or short; running, skipping,

or jumping; in a skirt or wearing trousers; facing towards the car or away from it; seen through rain or not; lit by street lights or by the sun; and so on.

There may be some assistance from 'generalisation via perception' (Section 8.2) but that in itself is unlikely to be sufficient. It seems that something like 'generalisation via unsupervised learning' in the SP Theory of Intelligence (Section 8.1) is needed.

With generalisation via unsupervised learning, it seems possible that, with reasonable amounts of driving experience across a range of driving conditions, generalisations may be made that would minimise the risk of accidents. Capabilities in that area may be strengthened by 'generalisations via perception' (Section 8.2).

As with human drivers, there would still be errors made by the artificial driver—because the generalisations would be probabilistic. But there is potential for the artificial driver to do substantially better than most human drivers—by inheriting the experience of many other artificial drivers, by not suffering from such things as falling asleep at the wheel, and by not being tempted to consume alcohol before driving.

10 Unsupervised learning from a single exposure or experience

> "How do humans learn concepts not from hundreds or thousands of examples, as machine learning systems have always been built for, but from just one example? ... Children can often learn a new word from seeing just one example of that word used in the right context, ... You can show a young child their first giraffe, and now they know what a giraffe looks like; you can show them a new gesture or dance move, or how you use a new tool, and right away they've got it ..." Joshua Tenenbaum [7, p. 471].
>
> "There's also 'zero-shot learning,' where people are trying to build programs that can learn when they see something even for the first time. And there is 'one-shot learning' where a program sees a single example, and they're able to do things." Oren Etzioni [7, p. 500].

As noted in Appendix A.5, most DNNs incorporate some variant of the idea that neural connections are gradually strengthened or weakened. Although this seems to reflect the way that it takes time to learn a complex skill such as playing the piano well, or competition-winning abilities in pool, billiards, or snooker, this feature of DNNs conflicts with the undoubted fact that people can and often do learn things from a single occurrence or experience.

Unsupervised learning in the SP System (Appendix A.5), which is quite different from gradualist kinds of learning in DNNs, will accommodate both learning from a single exposure or experience and the slow learning of complex skills:

- The SP System may exhibit learning from a single occurrence or experience because all learning in the system starts with the direct taking in of new information from the environment, much like an electronic recording machine. But unlike an electronic recording machine, all new information from the environment is interpreted in terms of existing knowledge (if any), so that newly-acquired information may immediately slot into a more-or-less rich interpretive structure;
- The SP system will, like a person, be relatively slow in the learning of a complex skill because that kind of learning requires a time-consuming search through a large abstract space of ways in which the knowledge may be structured in order to compress it and thus arrive at an efficient configuration.

It is true that DNNs begin their learning by taking in information supplied by the user, and may thus be said to achieve learning from a single exposure or experience. But this is quite different from the way a person, typically a child, may learn to be careful with hot things from one experience of getting burned.

In this area, the key difference between a DNN and the SP System is that, with a DNN, that first taking in of information is not useful until a lengthy process of learning has been applied. But with the SP System, newly-acquired information may serve immediately in processes of recognition or reasoning.

Figure 5 provides an example where the New SP-pattern '`f o r t u n e f a v o u r s t h e b r a v e`' enters directly into a process of analysis in terms of pre-established Old SP-patterns. To the extent that the encoding that results from that analysis is stored, the figure may be seen as an example of learning from one trial or experience. In a similar way, New information may be used immediately in processes of reasoning like those described in [39, Section 10] and [37, Chapter 7].

11 Computational resources, speeds of learning, and volumes of data

> "We can imagine systems that can learn by themselves without the need for huge volumes of labeled training data." Martin Ford [7, p. 12].
>
> "... the first time you train a convolutional network you train it with thousands, possibly even millions of images of various categories." Yann LeCun [7, p. 124].

> "People can learn from very few examples and generalize. We don't know how to build machines that can do that." Cynthia Breazeal [7, p. 456].
>
> "[A] stepping stone [towards AGI] is that it's very important that [AI] systems be a lot more data-efficient. So, how many examples do you need to learn from? If you have an AI program that can really learn from a single example, that feels meaningful. For example, I can show you a new object, and you look at it, you're going to hold it in your hand, and you're thinking, 'I've got it.' Now, I can show you lots of different pictures of that object, or different versions of that object in different lighting conditions, partially obscured by something, and you'd still be able to say, 'Yep, that's the same object.' But machines can't do that off of a single example yet. That would be a real stepping stone to AGI for me." Oren Etzioni [7, p. 502].

In connection with these issues, it has been discovered by Emma Strubell and colleagues [32] that the process of training a large AI model can emit more than 626,000 pounds of carbon dioxide, which is equivalent to nearly five times the lifetime emissions of the average American car, including the manufacture of the car itself.

The large computational resources, slow speeds of learning, and large volumes of data that are often associated with the training of DNNs seem to conflict with how people can learn fast with relatively little data and a brain that consumes only about 20 watts. A possible explanation, in terms of the workings of the SP System, has three elements:

1. Part of the explanation may be in the way that all learning in the SP System starts with the direct taking in of information from the system's environment, like an electronic recording system (Section 10).

2. Also, the processing of a small body of New information **N** to detect redundancies within **N**, and between **N** and pre-established Old SP-patterns, can be done efficiently via the building of SP-multiple-alignments (Appendix A.4), with corresponding savings in energy.

3. The creation of a grammar (a collection of SP-patterns) that can achieve economical encoding of any instance of **N**, can be done relatively efficiently (Appendix A.5), with corresponding savings in energy.

4. And, although the necessary research has not yet been done, it seems likely that this kind of unsupervised learning may be done incrementally, so that the computational demands at any one stage may be relatively modest. Thus

a grammar may be developed for all New information received by a given robot or other kind of AI since it was first created, regardless of the size of that body of information.

With regard to what Oren Etzioni says in the quote above, the SP System has potential as follows:

- *Learning three-dimensional structures.* It is unlikely that anyone would have fully 'got' an object in their hand until it has been viewed from two or more different angles. In that connection, there are applications already on the market that can construct a three-dimensional computational model of an object from photographs of the object taken from different angles.[1] It is envisaged that the SP System will be developed for the learning of 3D structures in a similar way [40, Sections 6.1 and 6.2];

- *Other aspects of learning.* It is likely that anyone older than a toddler would have learned a lot about objects that one can hold in one's hand, and, in the SP System, that kind of learning would facilitate unsupervised learning from a single exposure or experience (Section 10), via transfer learning (Section 13), and via generalisation in learning (Section 8.1). This may explain why an adult may look at a new object once and think 'I've got it.';

- *Generalisation in perception.* With knowledge of that object in the SP System, generalisation in perception (Section 8.2) may explain how that person may say 'Yep, that's the same object' despite variations in the object, variations in lighting conditions, or partial obscuring of the object by something else.

In connection with the large amounts of data that DNNs often need for learning, Yann LeCun says:

> "If [after training a convolutional network on one category] you ... want to add a new category, for example if the machine has never seen a cat and you want to train it to recognize cats, then it only requires a few samples of cats. That is because it has already been trained to recognize images of any type and it knows how to represent images; it knows what an object is, and it knows a lot of things about various objects. ... In the first few months of life, babies learn a huge amount by observation without having any notion of language. They learn an enormous amount of knowledge about how the world works just by

[1] See, for example, 'BigObjectBase', http//:bit.ly/2V1Koh5 and 'PhotoModeller', http//:bit.ly/2rNIPG9

> observation and with a little interaction with the world." Yann LeCun [7, p. 124].

As mentioned above, this kind of transfer learning (Section 13) may help to explain how the SP System may learn from relatively small amounts of data. Also, the last two sentences in the quote are references to unsupervised learning which is one of the strengths of the SP System (Section 7).

12 Strong compositionality

> "By the end of the '90s and through the early 2000s, neural networks were not trendy, and very few groups were involved with them. I had a strong intuition that by throwing out neural networks, we were throwing out something really important.
>
> "Part of that was because of something that we now call compositionality: The ability of these systems to represent very rich information about the data in a compositional way, where you compose many building blocks that correspond to the neurons and the layers." Yoshua Bengio [7, p. 25].

One can view the neurons and layers of a DNN as building blocks for a concept, and may thus be seen as an example of compositionality. But otherwise it seems that DNNs fail to capture the way in which we conceptualise a complex thing like a car in terms of smaller things (engine, wheels, etc), and these in terms of still smaller things (pistons, valves, etc), and so on. This kind of hierarchical representation of concepts, which is prominent in the way people conceptualise things, we may call 'strong' compositionality.

It appears that here, the SP System has a striking advantage compared with DNNs. Any SP-pattern may contain SP-symbols that serve as references to other SP-patterns, a mechanism which allows hierarchical structures to be built up through as many levels as are required.

This can be seen in Figure 5, where the SP-pattern '`D 8 t h e #D`' connects with the SP-pattern '`NP 1 D #D N #N #NP`', which connects with the SP-pattern '`VP 3 V #V NP #NP #VP`', which connects with the SP-pattern '`S 0 NP #NP VP #VP #S`'.

As noted in Section 13, there is a close relation between strong compositionality and transfer learning.

13 Transfer learning

> "Transfer learning is where you usefully transfer your knowledge from one domain to a new domain that you've never seen before, it's something humans are amazing at. If you give me a new task, I won't be terrible at it out of the box because I'll bring some knowledge from similar things or structural things, and I can start dealing with it straight away. That's something that computer systems are pretty terrible at because they require lots of data and they're very inefficient." Demis Hassabis [7, p. 174].
>
> "Humans can learn from much less data because we engage in transfer learning, using learning from situations which may be fairly different from what we are trying to learn." Ray Kurzweil [7, p. 230].
>
> "We need to figure out how to think about problems like transfer learning, because one of the things that humans do extraordinarily well is being able to learn something, over here, and then to be able to apply that learning in totally new environments or on a previously unencountered problem, over there." James Manyika [7, p. 276].

Transfer learning—meaning the use of old learning to facilitate later tasks—is fundamental in the SP System. Because the system does not suffer from catastrophic forgetting (Section 20), SP-patterns that have been learned at any stage, will be available in the system's repository of Old SP-patterns for use later:

- *Strong compositionality.* In unsupervised learning in the SP System, an SP-pattern that has been learned at any stage may, via 'references' between SP-patterns, become part of any SP-pattern that is learned later. This may yield strong compositionality in concepts, as outlined in Section 12.

- *Analysis or parsing.* Figure 5 shows an SP-multiple-alignment in which the New SP-pattern '`f o r t u n e f a v o u r s t h e b r a v e`' has been analysed in terms of the pre-existing Old SP-patterns that appear in rows 1 to 9. That old learning facilitates the new task of analysing a new sentence.

- *Partial matching in unsupervised learning.* Partial matching between, for example, the New SP-pattern '`t h a t g i r l r u n s`' and the Old SP-pattern '`t h a t b o y r u n s`' would, with some simplifications, lead to the creation of SP-patterns like '`X x0 t h a t #X`', '`Y y0 g i r l #Y`', '`Y y1 b o y #Y`', '`Z z0 r u n s #Z`', and '`A a0 X #X Y #Y Z #Z`', all of which are added to the repository of Old SP-patterns. In this case, the Old

SP-pattern has fed into the analysis of the New SP-pattern, leading to the creation of five additional SP-patterns.

- *Facilitation of the learning of new words.* When knowledge of a target language is relatively mature, it should facilitate the learning of new words or other structures. This would be like someone who is quite advanced in learning English as a foreign language hearing "The blah is good to eat" and inferring that "blah" is a noun that means something like a cake that is good to eat.

14 Transparency in the representation and processing of knowledge

> "... if regulation is intended to think about questions of safety, questions of privacy, questions of transparency, questions around the wide availability of these techniques so that everybody can benefit from them—then I think those are the right things that AI regulation should be thinking about." James Manyika [7, p. 283].
>
> "Although Bayesian updating is one of the major components in machine learning today, there has been a shift from Bayesian networks to deep learning, which is less transparent." Judea Pearl [7, p. 363].
>
> "The current machine learning concentration on deep learning and its non-transparent structures is such a hang-up." Judea Pearl [7, p. 369].

It is now widely recognised that: a major problem with DNNs is that the way in which learned knowledge is represented in such systems is far from being comprehensible by people; and that the way in which DNNs arrive at their conclusions is difficult or impossible for people to understand. These deficiencies are of concern for reasons of safety, legal liability, and perhaps more.[2]

By contrast, knowledge in the SP System is represented in a manner that is familiar to people, using such devices as class-inclusion hierarchies, part-whole hierarchies, and others. And there is an audit trail for all processing in the SP System, so that it is explicit and comprehensible by people.

As noted in Section 6, DNNs can make errors in recognition that may have serious consequences. And transparency in the SP System (in both the representation of knowledge and in processing), and experience with the SP Computer Model, suggests that it would not make those kinds of errors.

[2]See, for example, "Inside DARPA's push to make artificial intelligence explain itself", *CET US News*, 2017-08-10, bit.ly/2FQMoAr.

15 The representation and processing of natural language

> "... I think that many of the conceptual building blocks needed for AGI or human-level intelligence are already here. But there are some missing pieces. One of them is a clear approach to how natural language can be understood to produce knowledge structures upon which reasoning processes can operate." Stuart J. Russell [7, p. 51],
>
> "... a successful AI system needs some key abilities, including perception, vision, speech recognition, and action. These abilities help us to define artificial intelligence. We're talking about the ability to control robot manipulators, and everything that happens in robotics. We're talking about the ability to make decisions, to plan, and to problem-solve. We're talking about the ability to communicate, *and so natural language understanding also becomes extremely important to AI*." Stuart J. Russell [7, p. 169], emphasis added.

DNNs can do well in recognising speech. Also, they can produce impressive results in the translation of natural languages using a database of equivalences between surface structures that has been built up via human mark up and pattern matching, with English in some cases as a bridge between languages that are not English.

But DNNs are relatively weak in processing natural languages via the kinds of syntactic/semantic structures that are recognised in theoretical linguistics—and it appears that structures of that kind are important in the way that people process natural languages. It seems likely that, without those kinds of abilities, AI systems will not achieve human levels of language understanding, or production, and it seems likely that AI systems will not reach the accuracy of translations between natural languages that can be achieved by human experts.

As can be seen from the example in Figure 5 (Appendix A.4), the SP System lends itself well to the representation of syntactic knowledge and to its application in such tasks as parsing a natural language sentence. The SP System may also represent and process discontinuous syntactic dependencies in natural language ([39, Section 8.1] and [37, Section 5.4]). And the system has robust abilities to arrive at an intuitively 'correct' parsing, despite errors of omission, commission, and substitution in the sentence to be parsed (Section 8.2, Figure 2).

The SP System also lends itself to the representation of semantic structures (see, for example, [39, Section 9.1]) and it lends itself to the integration of syntax with semantics ([48], [37, Section 5.7]), and the understanding and production of natural language [37, Section 5.7].

A neat feature of the SP System is that the production of natural language is achieved by exactly the same mechanisms as are used for the parsing or understanding of natural language ([39, Section 4.5]), [37, Section 5.7.1]).

With the processing of natural language and the representation of its structures, as with other aspects of AI, a key feature of the SP System is its potential for SI. This feature of the system is likely to facilitate the smooth integration of syntax with semantics.

16 Several forms of probabilistic reasoning

> "What's going on now in the deep learning field is that people are building on top of these deep learning concepts and starting to try to solve *the classical AI problems of reasoning* and being able to understand, program, or plan." Yoshua Bengio [7, p. 21], emphasis added.
>
> "A lot of people might [say]: 'Deep learning systems are fine, but we don't know how to store knowledge, *or how to do reasoning*, or how to build more expressive kinds of models, because deep learning systems are just circuits, and circuits are not very expressive after all.'" Stuart J. Russell [7, p. 49], emphasis added.
>
> "I think there's a presupposition that the way AIs can develop is by making individuals that are general-purpose robots like you see on Star Trek. ... I ... think, *in terms of general reasoning capacity, it's not going to happen for quite a long time.*" Geoffrey Hinton [7, p. 88], emphasis added.

As potential foundations for AGI, DNNs appear to be unsuitable for performing anything but the most rudimentary kind of reasoning. By contrast, a strength of the SP System is that, via the SP Computer Model, several different kinds of reasoning can be demonstrated, without any special provision or adaptation (Appendix A.9.2).

Because of the probabilistic nature of the SP System (Appendix A.3), all the kinds of reasoning that may be modelled in the SP System are fundamentally probabilistic, although it is possible to simulate the all-or-nothing nature of much classical logic when when probabilities are at or near 0 or 1 (Section 4).

Much as with the processing of natural language (Section 15), a strength of the SP System is that there can be SI.

The foregoing remarks apply to the non-neural version of the SP System, expressed in the SP Computer Model. It is anticipated that, when the 'neural' version of the SP System (SP-Neural, Appendix A.7) is more mature, it will inherit the features and capabilities of the non-neural version of the SP System.

17 Commonsense reasoning and commonsense knowledge

> "We don't know how to build machines that have human-level common sense. We can build machines that can have knowledge and information within domains, but we don't know how to do the kind of common sense we all take for granted." Cynthia Breazeal [7, p. 456].
>
> "We still don't have any real AI in the sense of the original vision of the founders of the field, of what I think you might refer to as AGI—machines that have that same kind of flexible, general-purpose, common sense intelligence that every human uses to solve problems for themselves." Joshua Tenenbaum [7, p. 472].

Although 'commonsense reasoning' (CSR) is a kind of reasoning, it is discussed here, with 'commonsense knowledge' (CSK), in a section that is separate from Section 16 because of the way CSR and CSK (which, together, may be referred to as 'CSRK') have been developing as a discrete subfield of AI (see, for example, [4]).

Judging by the nature of DNNs and the paucity of research on how they might be applied in the CSRK area [29], it seems that DNNs are not well suited to this aspect of AI. By contrast, the SP System shows promise in this area:

- Several features of the SP System suggest that it is likely to be useful with CSRK [49, Section 3], and more so when 'unfinished business' in the development of the SP Computer Model has been completed (Appendix A.11).

- Three aspects of CSRK may be modelled with the SP Computer Model [49, Sections 4 to 6]: how to interpret a noun phrase like "water bird"; how, under various scenarios, to assess the strength of evidence that a given person committed a murder; how to interpret the horse's head scene in *The Godfather* film.

  A fourth problem—how to model the process of cracking an egg into a bowl—is beyond what can be done with the SP System as it is now [49, Section 9], but fixing the problems mentioned under the previous bullet point may make it feasible.

- With the SP Computer Model, it is possible to determine the referent of an ambiguous pronoun in a 'Winograd schema' type of sentence [48], where a Winograd schema is a pair of sentences like *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*, and the ambiguous pronoun in each sentence is "they" [15].

18 The central role of IC in the SP System compared with other AI systems

> "There are two parts to an autoencoder, an encoder and a decoder. The idea is that the encoder part takes an image, for example, and tries to represent it in a compressed way, such as a verbal description. The decoder then takes that representation and tries to recover the original image. The autoencoder is trained to do this compression and decompression so that it is as faithful as possible to the original.
>
> "Autoencoders have changed quite a bit since that original vision. Now, we think of them in terms of taking raw information, like an image, and transforming it into a more abstract space where the important, semantic aspect of it will be easier to read. That's the encoder part. The decoder works backwards, taking those high-level quantities—that you don't have to define by hand—and transforming them into an image. That was the early deep learning work.
>
> "Then a few years later, we discovered that we didn't need these approaches to train deep networks, we could just change the nonlinearity." Yoshua Bengio [7, p. 26].

As noted in Section 2.1.3, IC may be seen as a unifying theme in HLPC [50], and it is fundamental in all aspects of the SP System. The remarks quoted above suggest that IC might be equally important in research with autoencoders. But this seems to be far from the case:

- The third of the remarks quoted above, suggests that research with autoencoders has now been dropped.
- However, a search of the literature shows that research with autoencoders is still going on (see, for example, [1, 17, 24, 25, 34], but it appears not to be in the mainstream of current research with DNNs.
- Although Jürgen Schmidhuber suggests that "much of machine learning is essentially about compression," [29, Section 5.10], IC gets only brief mentions in his review of research about DNNs [29, Sections 4.4, 5.6, 5.7, 5.10].
- Given the importance of IC in the SP System, in its theoretical foundations [46, 50], and in what it can do (Appendix A.9), autoencoders are puzzling: compression of information by an autoencoder sounds OK, but then all the good work is undone when the economical code created by the encoder is decompressed by the decoder.

- An explanation for that puzzling feature of autoencoders is provided by [1, Section 1]: "Autoencoders were first introduced in the 1980s by Hinton and the PDP group [26] to address the problem of 'backpropagation without a teacher', by using the input data as the teacher."

- This explanation for the curious design of autoencoders makes a certain amount of sense but it raises a bigger problem than it solves: why not simply use the input data directly as the teacher, without the complexity of encoding and decoding? Using the input data directly at the beginning of unsupervised learning, is how the SP System learns (Section 7, Appendix A.5), and evidence to date provides no reason to change that feature of the system.

- The way in which the SP System uses input data directly at the beginning of learning is the reason that the SP System is able, like people and unlike DNNs, to learn effectively from a single exposure or experience (Section 10).

19 Biological validity

> "A convolutional network is a particular way of connecting the neurons with each other in such a way that the processing that takes place is appropriate for things like images. I should add that we don't normally call them neurons because they're not really an accurate reflection of biological neurons. Yann LeCun [7, p. 122].
>
> "Deep learning will do some things, but biological systems rely on hundreds of algorithms, not just one algorithm. We will need hundreds more algorithms before we can make that progress, and we cannot predict when they will pop." Rodney Brooks [7, p. 427].

With regard to the first quote, above, it is generally recognised that most neural networks are only vaguely related to biological systems. By contrast, the concepts of SP-pattern and SP-multiple-alignment in the SP System have the benefit of Hebb's studies in neuroscience: in SP-Neural (Appendix A.7), the concept of an SP-pattern maps quite well on to a variant of Hebb's concept of a 'cell assembly' [11, Location 142], and connections amongst cell assemblies may be understood in terms of the SP-multiple-alignment construct.

Although 'convolutional' neural networks "are directly inspired by the classic notions of simple cells and complex cells in visual neuroscience" [14, p. 439], the concepts of 'neural symbol' and 'pattern assembly' in SP-Neural (Appendix A.7) may be seen to be configured in a similar way.

The second quote, above, prejudges the issue of whether human intelligence might be: 1) the product of some unifying principle, as in the SP Theory; or 2) some non-unifying perspective, as in Marvin Minsky's concept of diverse agents [20], or Gary Marcus's concept of a kluge deriving from the haphazard nature of evolution [18]; or 3) that it might be some combination of a unifying principle with some elements of a kluge. Although the main focus in presenting the SP System is on the first possibility, it is difficult to deny the kluge-like nature of much human thinking which, with evidence for the SP concepts, suggests that the third possibility is most likely.

Apart from those issues, the central role for IC in the workings of the SP System (Section 2.1.3 and Appendix A.2), coupled with the SP System's versatility in AI-related functions (Appendix A.9) reflects the way in which IC may be seen as a unifying principle in HLPC [50].

By comparison, and notwithstanding the use of autoencoders in DNN systems (Sections 2.1.3 and 18), there appears to be little or no recognition, in research on DNNs, of the fundamental importance of IC in understanding the nature of intelligence and in HLPC.

20 Catastrophic forgetting

This and the following two sections describe problems in AI that are not apparently condidered in [7] but are problems which the SP System may help to solve.

Catastrophic forgetting—which is a problem for at least some DNNs—is the way in which, when a given DNN has learned one thing and then it learns something else, the new learning normally wipes out the earlier learning (see, for example, [10]). This is quite different from human learning, where new learning often builds on earlier learning, although of course we all have a tendency to forget things.

A related problem is that, to be practical, a learning system, like a person, should be able to learn continuously from its environment and not always in discrete batches.[3]

The SP System is entirely free of the problem of catastrophic forgetting. Although the SP Computer Model is not currently configured for continuous learning, it is likely that it could be be adapted in that way.

The reasons that, in general, DNNs suffer from catastrophic forgetting and that the SP System does not, are that:

[3]It appears that this problem is a matter of concern to military planners as described, for example, in "DARPA seeking AI that learns all the time", *IEEE Spectrum*, 2017-11-21, bit.ly/2BdERfZ.

- In DNNs there is a single structure for the learning and storage of new knowledge, a concept like 'my grandmother' is encoded in the strengths of connections between artificial neurons in that single structure (*cf.* the discussion of 'grandmother' cells in [45, Section 5.8]), so that the later learning of a concept like 'my house' is likely to disturb the strengths of connections for 'my grandmother';

- By contrast, the SP System has an SP-pattern for each concept in its repository of knowledge, there is no limit to the number of such SP-patterns that may be stored (apart from the limit imposed by the available storage space in the computer), and there is no interference between any one SP-pattern and any other.

It is true that, in the SP System, a concept like 'person' may be composed of subsidiary concepts like 'head', 'body' and 'legs' and that corruption of those subsidiary concepts would corrupt the concept of 'person'. But, in accordance with our ordinary experience, it is entirely feasible to provide an SP-pattern to record that 'John' has an injury to his 'head' without disturbing the SP-pattern that records the fact that the 'head' of a typical 'person' is not injured.

21 Scope for adaptation

> "What's missing from AI today—and likely to stay missing, until and unless the field takes a fresh approach—is broad (or 'general') intelligence. AI needs to be able to deal not only with specific situations for which there is an enormous amount of cheaply obtained relevant data, but also problems that are novel, and variations that have not been seen before.
>
> "Broad intelligence, where progress has been much slower, is about being able to adapt flexibly to a world that is fundamentally open-ended—which is the one thing humans have, in spades, that machines haven't yet touched. But that's where the field needs to go, if we are to take AI to the next level." Gary Marcus and Ernest Davis [19, p. 15].

A problem which is closely related to catastrophic forgetting (Section 20) and also to the distinction between 'broad' and 'narrow' AI (Section 22) is that any one DNN is really only designed to learn a single concept. It is true that one could provide multiple DNNs for the learning of multiple concepts but, since a DNN has multiple layers and multiple connections between layers (which is what makes it 'deep'), the provision of a DNN for each of the many concepts that people can learn would be expensive.

In the SP System, the concept of SP-multiple-alignment, with the concept of SP-pattern, provides a much greater scope for modelling the world than the relatively constrained framework of DNNs. This is because each concept is represented by one SP-pattern, there is no limit to the number of SP-patterns that may be formed (apart from the memory that is available in the host computer), and there is no limit to the number of ways in which a given SP-pattern may be connected to other SP-patterns within the framework of SP-multiple-alignments (in much the same way that there is no limit to the number of ways in which a given web page may be connected to other web pages).

By contrast, the layers in a DNN, and the potential connections amongst them, are finite and pre-defined (Section 20). It is true that the connections can vary in strength but only within pre-defined limits.

22 Broad versus narrow AI

Last but one in this paper, but by no means least, is a problem described by Gary Marcus and Ernest Davis in their book *Rebooting AI* [19]. Writing about what they see as the shortcomings of most AI systems today, they say:

> "The central problem, in a word: current AI is *narrow*; it works for particular tasks that it is programmed for, provided that what it encounters isn't too different from what it has experienced before. That's fine for a board game like Go—the rules haven't changed in 2,500 years—but less promising in most real-world situations. Taking AI to the next level will require us to invent machines with substantially more flexibility." ([19, pp. 12–13], emphasis in the original).

And later:

> "To be sure, ... narrow AI is certainly getting better by leaps and bounds, and undoubtedly there will be more breakthroughs in the years to come. But it's also telling: AI could and should be about so much more than getting your digital assistant to book a restaurant reservation." [19, p. 14].

Writing about possible ways forward, Marcus and Davis say:

> "What's missing from AI today—and likely to stay missing, until and unless the field takes a fresh approach—is *broad* (or 'general') intelligence. AI needs to be able to deal not only with specific situations for which there is an enormous amount of cheaply obtained relevant data,

but also problems that are novel, and variations that have not been seen before.

"Broad intelligence, where progress has been much slower, is about being able to adapt flexibly to a world that is fundamentally open-ended—which is the one thing humans have, in spades, that machines haven't yet touched. But that's where the field needs to go, if we are to take AI to the next level." ([19, p. 15], emphasis in the original).

"We call this book *Rebooting AI* because we believe that the current approach isn't on a path to get us to AI that is safe, smart, or reliable. A short-term obsession with narrow AI and the easily achievable 'low-hanging fruit' of big data has distracted too much attention away from a longer-term and much more challenging problem that AI needs to solve if it is to progress: the problem of how to endow machines with a deeper understanding of the world. Without that deeper understanding, we will never get to truly trustworthy AI. In the technical lingo, we may be stuck at a local maximum, an approach that is better than anything similar that's been tried, but nowhere good enough to get us where we want to go.

"For now, there is an enormous gap—we call it 'the AI Chasm'—between ambition and reality." [19, pp. 17–18].

Of course, the SP System is nowhere near delivering 'broad AI' as described by Marcus and Davis. But, for reasons summarised in the following subsections, it may with some justice claim to be the kind of "fresh approach", with breadth and generality, that they call for. It has strengths that put it on a much firmer foundation than narrow approaches to the development of 'broad AI' or 'artificial general intelligence'.

22.1 Generality via simplification and integration across a broad canvass

As noted in Appendix A.1, "The SP programme of research is a unique attempt to simplify and integrate observations and concepts across artificial intelligence, mainstream computing, mathematics, and human learning, perception, and cognition," aiming for a favourable combination of conceptual Simplicity with descriptive or explanatory Power. And, as noted in the same section, the simplicity-with-power objective has been largely met via the development of the SP-multiple-alignment concept.

This overarching goal in the SP programme of research, and success in meeting that goal, contrasts sharply with the kind of narrow AI described by Marcus and

Davis which dominates AI research today. That overarching goal chimes well with the need for the development of "*broad* (or 'general') intelligence."

It is true that there has been research for many years inspired by Allen Newell's book about *Unified Theories of Cognition* [21], and there is also a significant strand of research aiming to develop 'Artificial General Intelligence' (see, for example, [12]). But despite the welcome aims of researchers in these areas, it appears safe to say that, on the strength of [19] and other sources such as [12], nothing, apart from the SP System, has yet been developed that demonstrates any useful degree of simplification and integration across diverse aspects of intelligence.

22.2 Generality via information compression, SP-multiple-alignment, and unsupervised learning

Another reason to be optimistic about the potential of the SP System to provide a relatively firm foundation for the development of 'broad' or 'general' AI is: 1) that all processing in the SP System is achieved via the compression of information; 2) that IC lies at the heart of how the system represents its knowledge; 3) that IC is an extremely general concept that can in principle be applied to any kind of processing or the representation of any kind of knowledge; and 4) that evidence to date shows the wide range of kinds of processing and kinds knowledge where IC may be applied to good effect.

To be more specific, IC in the SP System is achieved exclusively via the SP-multiple-alignment version of ICMUP (Appendices A.2 and A.4), which is itself a major part of unsupervised learning within the SP System. These processes provide the key to the existing and potential versatility of the SP System (Section 22.4).

22.3 Generality via generalisation and the correction of over- and under-generalisations

A potentially valuable bonus from a theory of HLPC and AI which is founded on IC for both the processing and representation of knowledge is that it provides a robust, coherent theory of generalisation and the correction of both over- and under-generalisations, which has some empirical support (Section 8).

As described in Section 9, such a theory of generalisation may help to overcome problems of adaptability in driverless cars, described by Marcus and Davis like this:

> "Take driverless cars. It's comparatively easy to create a demo of a driverless car that keeps to a lane correctly on a quiet road; people have been able to do that for years. It appears to be vastly harder

> to make them work under circumstances that are challenging or unexpected. As Missy Cummings, director of Duke University's Humans and Autonomy Laboratory (and former U.S. Navy fighter pilot), put it to us in an email, the issue isn't even how many miles a given driverless car might go without an accident, it's how adaptable those cars are. In her words, today's semi-autonomous vehicles 'typically perform only under extremely narrow conditions which tell you nothing about how they might perform [under] different operating environments and conditions.' Being almost perfectly reliable across millions of test miles in Phoenix doesn't mean it is going to function well during a monsoon in Bombay." [19, p. 21].

It is an ability to generalise in accordance with a theory of generalisation based on well-founded principles which is perhaps the most important means of bridging the 'AI Chasm', taking us from 'narrow' AI to 'broad' AI as described in the quote above which is repeated here:

> ""Broad intelligence ... is about being able to adapt flexibly to a world that is fundamentally open-ended—which is the one thing humans have, in spades, that machines haven't yet touched. But that's where the field needs to go, if we are to take AI to the next level." [19, p. 15].

22.4 Generality via versatility, seamless integration, and Simplicity with Power

As described in Appendix A.9, the SP System has strengths and potential in a variety of aspects of intelligence (Appendix A.9.1), including several kinds of reasoning (Appendix A.9.2), and in the representation of several different kinds of knowledge (Appendix A.9.3). It also has clear potential for the seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination (Appendix A.9.4).

That these things all flow from relatively simple structures and mechanisms in the SP System, indicates a favourable ratio in the SP System of conceptual Simplicity with descriptive and explanatory Power. A favourable S/P ratio is a key feature of any broad AI.

22.5 Generality via potential benefits and applications of the SP System

Another means of assessing the generality of the SP System is via potential benefits and applications of the system which are credible in the sense that they are not

mere speculations but have supporting evidence via the workings and performance of the system. In this respect, the SP System scores well, as can be seen from the range of potential benefits and applications outlined in Section A.10.

23 Motivations and emotions

> "How much prior structure do we need to build into those systems for them to actually work appropriately and be stable, and for them to have intrinsic motivations so that they behave properly around humans? There's a whole lot of problems that will absolutely pop up, so AGI might take 50 years, it might take 100 years, I'm not too sure." Yann LeCun [7, p. 130].
>
> "Machine learning needs a lot of data, and so I borrowed [a] dataset [from Cambridge Autism Research Center] to train the algorithms I was creating, on how to read different emotions, something that showed some really promising results. This data opened up an opportunity to focus not just on the happy/sad emotions, but also on the many nuanced emotions that we see in everyday life, such as confusion, interest, anxiety or boredom." Rana el Kaliouby [7, p. 209].
>
> "[A] subtle question is that of relating emotionally to other beings. I'm not sure that's even well defined, because as a human you can fake it. There are people who fake an emotional connection to others. So, the question is, if you can get a computer to fake it well enough, how do you know that's not real?" Daphne Koller [7, p. 394].
>
> "If you look at human intelligence we have all these different kinds of intelligences, and social and emotional intelligence are a profoundly important, and of course underlies how we collaborate and how we live in social groups and how we coexist, empathize, and harmonize." Cynthia Breazeal [7, p. 450].
>
> "... why are we assuming the same evolutionary forces that drove the creation of our motivations and drives would be anything like those of [a] super intelligence?" Cynthia Breazeal [7, p. 457].

In developing human-like AI, motivations and emotions are clearly important, not least because of the possibility that super-intelligent AIs might come to regard people as dispensable. But, in the SP programme of research, there has, so far, been no attempt to give the SP Computer Model any kind of motivation (except the motivation 'compress information'), or any kind of emotion. This is because of the belief that, in relation to the SP concepts and their development, it would be

trying to run before we can walk. When the SP System is more mature, there may be a case for exploring how it may be applied to the complexities of motivations and emotions.

24 Conclusion

In the book *Architects of Intelligence* [7], Martin Ford's interviews with people who are influential in AI give very interesting and useful insights into current thinking in the field. As noted in the Introduction, this paper is not a review of *Architects of Intelligence* as such. Rather, it discusses problems in AI described in the book in relation to features of the SP System.

Those discussions help to throw into relief the substantial advantages of the *SP System* compared with deep neural networks ('DNNs')—which, because of their dominance in AI today, form a backdrop for much of this paper and, very often, the explicit or implied referent for comparison with the SP System.

Key features of the SP System are:

- *Conceptual Simplicity and descriptive or explanatory Power.* The SP System is the product of a programme of research seeking to simplify and integrate observations and concepts across a broad canvass. This has proved to be, to a large extent, successful with the creation of a system where the combination of conceptual Simplicity with descriptive and explanatory Power is, arguably, much more favourable than with other AI system.

- *A central role for IC.* With its central role for information compression (IC), the SP System adopts a key principle from human cognitive psychology and neuroscience, for which there is much evidence [50].

- *SP-multiple-alignment.* IC in the SP Computer Model is achieved largely via the powerful concept of *SP-multiple-alignment*. With the concept of SP-pattern, the SP-multiple-alignment construct is largely responsible for the relative simplicity of the SP System and for its versatility in diverse aspects of intelligence, including diverse kinds of reasoning, and the representation of diverse kinds of knowledge, and for its potential for the seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination.

- *SP-Neural.* Abstract concepts in the SP Theory map quite well into plausible structures and processes in neurons and their interconnections.

Apart from the key features, the main strengths of the SP System and the SP programme of research are:

- *Top-down strategy.* In the quest for a favourable combination of conceptual Simplicity with descriptive or explanatory Power, there are merits in adopting a top-down strategy, developing over-arching principles and working from them to lower-level concepts. It appears that the SP System, as a product of that top-down strategy, has achieved a much more favourable combination of Simplicity and Power than any of the alternatives.

- *Recognition of images and speech.* Although the SP System does not at present do as well as DNNs in such things as the recognition of images and speech, its long-term potential appears to be greater.

- *DNNs are easily fooled.* The SP System appears to be entirely free from the tendency of DNNs to make large and unpredictable errors in recognition.

- *The importance of unsupervised learning.* Unlike most DNNs, learning in the SP System is unsupervised, a form of learning which is prominent in the way people learn, which may be the foundation for other kinds of learning, and which is regarded as important by several of the interviewees for *Architects of Intelligence*.

- *Generalisation, under-generalisation and over-generalisation.* The central role of IC in the SP system provides the basis for generalisation in accordance with coherent principles. These principles provide definitions of over- or under-generalisation and a safeguard against them.

- *Minimising the risk of accidents with driverless cars.* The complexities of driving a car seem to require human-like intelligence or better. And that seems to require the development of generalisation in accordance with principles in the SP Theory of Intelligence.

- *Unsupervised learning from a single occurrence or experience.* Like people, and unlike DNNs, the SP System can learn useful information from a single occurrence or experience.

- *Computational resources, speed of learning, and volumes of data.* Like people, and unlike DNNs, the SP System can demonstrate useful learning with relatively small computational resources, relatively fast, and with relatively tiny volumes of data.

- *Transfer learning.* Transfer learning—meaning the use of old learning to facilitate later tasks—is prominent in human learning and is fundamental in the SP System. In this respect, the SP System contrasts sharply with DNNs.

- *Transparency in the representation and processing of knowledge.* Unlike DNNs, the SP System is entirely transparent in how it represents knowledge, and it provides a full audit trail for all its processing.

- *The representation and processing of natural language.* Unlike DNNs, the SP System can represent and process natural language with the kinds of structures that are recognised by linguists which, arguably, have psychological validity, and which appear to be needed ultimately for human-like capabilities with natural languages.

- *Several forms of probabilistic reasoning.* As a by-product of its design, the SP System exhibits several forms of probabilistic reasoning.

- *Commonsense reasoning and commonsense knowledge.* Because commonsense reasoning and commonsense knowledge (CSRK) have developed as a distinct field within AI, it is discussed in a separate section, although the SP System's strengths in probabilistic reasoning are part of its potential with CSRK. Other strengths of the SP System in that area have been described.

- *The central role for IC in the SP System compared with other AI systems.* The SP System appears to be unique in employing IC as the basis for all aspects of intelligence;

- *Biological validity.* Arguably, the SP System, with *SP-Neural*, has greater validity in terms of biology than DNNs, both in its organisation and in the central role of IC in how it works;

- *Catastrophic forgetting.* Unlike most DNNs, the SP System is entirely free from catastrophic forgetting;

- *Scope for adaptation.* The representation of knowledge with SP-patterns, with the SP-multiple-alignment construct, provides for much greater scope for adaptation than the layers of a DNN;

- *Broad versus narrow AI.* Owing to its combination of conceptual Simplicity with descriptive and explanatory Power, and some other features, the SP System provides a firmer foundation that other AI systems for the development of 'broad' or 'general' AI.

Despite the importance of motivations and emotions, no attempt has yet been made to study them in the SP programme of research.

Acknowledgements

I am grateful to anonymous reviewers for constructive comments on earlier drafts of this paper.

A Outline of the SP Theory of Intelligence and its realisation in the SP Computer Model

The *SP System*—meaning the *SP Theory of Intelligence* and its realisation in the *SP Computer Model*—is a system that has been under development since about 1987, with a break between early 2006 and late 2012.

The SP System is described in outline here, in more detail in [39], and much more fully in [37]. Distinctive features and advantages of the SP System are described in [46]. Other papers in this programme of research are detailed, with download links, on www.cognitionresearch.org/sp.htm.

In broad terms, the SP System is a brain-like system that takes in *New* and not compressed information through its senses and stores some or all of it as *Old* information that is compressed, as shown schematically in Figure 3.

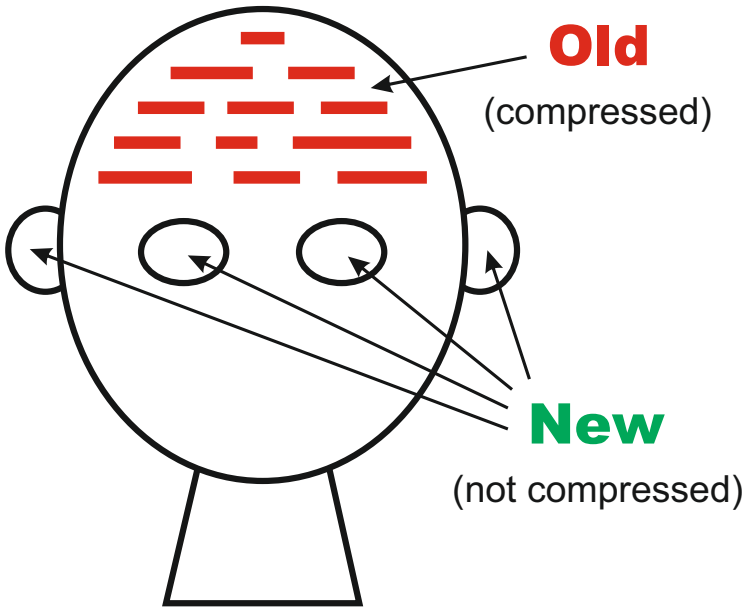

Figure 3: Schematic representation of the SP System from an 'input' perspective. Reproduced, with permission, from Figure 1 in [39].

In the SP System, all kinds of knowledge is represented with *SP-patterns*, where each such SP-pattern is an array of atomic *SP-symbols* in one or two dimensions. At present, the SP Computer Model works only with one-dimensional SP-patterns but it is envisaged that it will be generalised to work with two-dimensional SP-patterns as well.

A.1 Aiming for a favourable combination of conceptual *Simplicity* with descriptive or explanatory *Power*

The SP programme of research is a unique attempt to simplify and integrate observations and concepts across artificial intelligence, mainstream computing, mathematics, and human learning, perception, and cognition (HLPC). This may be seen to be a process of developing concepts that combine conceptual *Simplicity* with high levels of descriptive or explanatory *Power*.

The main justification for this strategy is Ockham's razor: a theory should be simple but not so simple that it becomes trivial (ie, it loses Power). Also, President Eisenhower is reputed to have said: "If you can't solve a problem, enlarge it," meaning that putting a problem in a broader context may make it easier to solve. Good solutions to a problem may be hard to see when the problem is viewed through a keyhole, but become visible when the door is opened.

This top-down approach to the development of concepts contrasts with the more popular bottom-up approach which seeks to develop ideas in one area such as computer vision and then integrate it with other areas such as reasoning, and to repeat that kind of integration to create groupings of progressively increasing size.

Despite its ambition, the simplicity-with-power objective has been largely met. This is because the SP System, which is largely the simple but powerful concept of SP-multiple-alignment (Appendix A.4), has strengths and potential across diverse aspects of intelligence and the representation of knowledge (Appendix A.9).

A.2 Information compression via the matching and unification of patterns

A central idea in the SP System is that all kinds of processing would be achieved via information compression (IC). Evidence for the importance of IC in HLPC is described in [50].

In the development of the SP System, it has proved useful to understand IC in terms of the discovery of patterns that match each other and the merging or 'unification' of patterns that are the same. The expression 'IC via the matching and unification of patterns' may be shortened to 'ICMUP'. Compression of information

in the SP System is achieved via ICMUP and, more specifically via the creation of SP-multiple-alignments (Appendix A.4).

Seven variants of ICMUP are described in [50, Section 2.1]. SP-multiple-alignment is the seventh variant of ICMUP which may be seen to be a generalisation of the other six variants.

In terms of theory, the emphasis on IC in the SP System accords with research in the tradition of Minimum Length Encoding (see, for example, [16]), with the qualification that most research relating to MLE assumes that the concept of a universal Turing machine provides the foundation for theorising, whereas the SP System is founded on concepts of ICMUP and SP-multiple-alignment [37,].

A.3 The probabilistic nature of the SP System

Owing to the intimate relation that is known to exist between IC and concepts of probability [30, 31], and owing to the fundamental role of IC in the workings of the SP System, the system is inherently probabilistic.

That said, it appears to be possible to imitate the all-nothing-nature of conventional computing systems via the use of data where all the probabilities yielded by the system are at or close to 0 or 1.

Because of the probabilistic nature of the SP System, it lends itself to the modelling of HLPC because of the prevalence of uncertainties in that domain. Also, the SP System sits comfortably within AI because of the probabilistic nature of most systems in AI, at least in more recent work in that area.

An advantage of the SP System in those areas is that it is relatively straightforward to calculate absolute or conditional probabilities for results obtained in, for example, different kinds of reasoning (Appendix A.9.2).

The very close connection that exists between IC and concepts of probability may suggest that there is nothing to choose between them. But [51, Section 8.2] argues that, in research on aspects of AI and HLPC, there are reasons to regard IC as more fundamental than probability and a better starting point for theorising.

A.4 SP-multiple-alignment

A central idea in the SP System, is the simple but powerful concept of *SP-multiple-alignment*, borrowed and adapted from the concept of 'multiple sequence alignment' in bioinformatics. As mentioned in Appendix A.2, SP-multiple-alignment is the seventh variant of ICMUP described in [50, Section 2.1] and may be seen as a generalised version of the other six variants.

Probably the best way to explain the idea is by way of examples. Figure 4 shows an example of multiple sequence alignment in bioinformatics. Here, there are five DNA sequences which have been arranged alongside each other, and then,

by judicious 'stretching' of one or more of the sequences in a computer, symbols that match each other across two or more sequences have been brought into line.

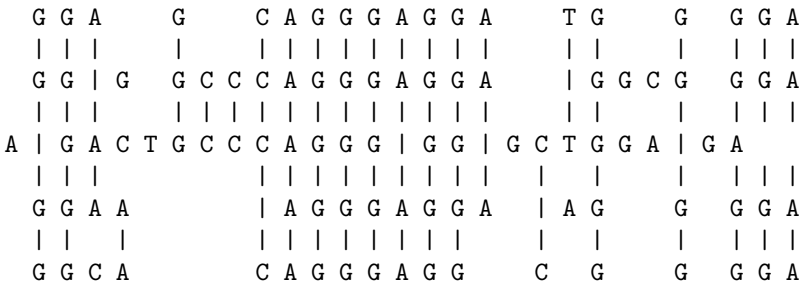

Figure 4: A 'good' multiple alignment amongst five DNA sequences.

A 'good' multiple sequence alignment, like the one shown, is one with a relatively large number of matching symbols from row to row. The process of discovering a good multiple sequence alignment is normally too complex to be done by exhaustive search, so heuristic methods are needed, building multiple sequence alignments in stages and, at each stage, selecting the best partial structures for further processing.

Some people may argue that the combinational explosion with this kind of problem, and the corresponding computational complexity, is so large that there is no practical way of dealing with it. In answer to that objection, there are several multiple sequence alignment programs used in bioinformatics—such as 'Clustal Omega', 'Kalign', and 'MAFFT'[4]—which produce results that are good enough for practical purposes.

This relative success is achieved via the use of heuristic methods that conduct the search for good structures in stages, discarding all but the best alignments at the end of each stage. With these kinds of methods, reasonably good results may be achieved but normally they cannot guarantee that the best possible result has been found.

Figure 5 shows an example of an SP-multiple-alignment, superficially similar to the one in Figure 4, except that the sequences are called *SP-patterns*, the SP-pattern in row 0 is New information and the remaining SP-patterns, one per row, are Old SP-pattern, selected from a relatively large pool of such SP-patterns. A 'good' SP-multiple-alignment is one which allows the New SP-pattern to be encoded economically in terms of the Old SP-patterns.

[4]Provided as online services by the European Bioinformatics Institute (see https://www.ebi.ac.uk/Tools/msa/).

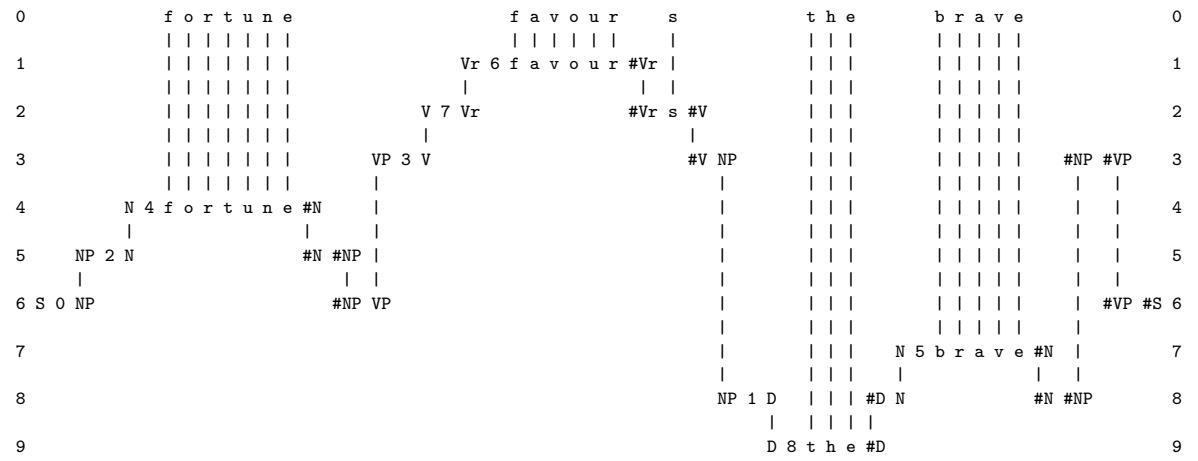

Figure 5: The best SP-multiple-alignment produced by the SP Computer Model with a New SP-pattern, '`f o r t u n e f a v o u r s t h e b r a v e`', representing a sentence to be parsed and a repository of user-supplied Old SP-patterns representing grammatical categories, including morphemes and words.

In this example, the New SP-pattern (in row 0) is a sentence and each of the remaining SP-patterns represents a grammatical category, where 'grammatical categories' include words. The overall effect of SP-multiple-alignment in this example is the parsing a sentence ('`f o r t u n e f a v o u r s t h e b r a v e`') into its grammatical parts and sub-parts.

Contrary to the impression that may be given by Figure 5, the SP-multiple-alignment concept is very versatile and, as described in Appendix A.6 and Appendix A.9, it may serve to model several different aspects of intelligence, including several kinds of reasoning, it may serve in the representation of several different kinds of knowledge, and it facilitates the seamless integration of diverse aspects of intelligence and the diverse kinds of knowledge in any combination (SI).

As with multiple sequence alignments, it is almost always necessary to use heuristic methods to achieve useful results without undue computational demands. The use of heuristic methods helps to ensure that computational complexities in the SP System are within reasonable bounds [37, Sections A.4, 3.10.6 and 9.3.1].

In the SP Computer Model, the size of the memory available for searching may be varied, which means in effect that the scope for backtracking can be varied. When the scope for backtracking is increased, the chance of the program getting stuck on a 'local peak' (or 'local minimum') in the search space is reduced.

A.5 Unsupervised learning in the SP System

In the SP System, learning is 'unsupervised', deriving structures from incoming sensory information without the need for any kind of 'teacher', or anything equivalent (*cf.* [9]).

Unsupervised learning in the SP System is quite unlike learning via the gradual strengthening or weakening of neural connections, variants of which are the mainstay of learning in DNNs. In the SP System, unsupervised learning incorporates the building of SP-multiple-alignments but there are other processes as well.

In brief, the system creates Old SP-patterns from complete New SP-patterns and also from partial matches between New and Old SP-patterns. With a given body of New SP-patterns, the system processes them as just sketched, and then searches for one or two 'good' *SP-grammars*, where an SP-grammar is a collection of Old SP-patterns, and it is 'good' if it is effective in the economical encoding of the original set of New SP-patterns, where that economical encoding is achieved via SP-multiple-alignment.

As with the building of SP-multiple-alignments, the process of creating good grammars is normally too complex to be done by exhaustive search so heuristic methods are needed. This means that the system builds SP-grammars incrementally and, at each stage, it discards all but the best SP-grammars. As with the building of SP-multiple-alignments, the use of heuristic methods helps to ensure

that computational complexities in the SP System are within reasonable bounds [37, Sections A.4, 3.10.6 and 9.3.1].

The SP Computer Model has already demonstrated an ability to learn generative grammars from unsegmented samples of English-like artificial languages, including segmental structures, classes of structure, and abstract patterns, and to do this in an 'unsupervised' manner ([39, Section 5], [37, Chapter 9]).

But there are (at least) two shortcomings in the system [39, Section 3.3]: it cannot learn intermediate levels of structure or discontinuous dependencies in grammar, although the SP-multiple-alignment framework can accommodate structures of those kinds. It appears that those two problems may be overcome and that their solution would greatly enhance the capabilities of the SP Computer Model in unsupervised learning.

A.6 Two main mechanisms for information compression in the SP System, and their functions

The two main mechanisms for IC in the SP System are as follows, each one with details of its function or functions:

1. *The building of SP-multiple-alignments.* The process of building SP-multiple-alignments achieves compression of New information. At the same time it may achieve any or all of the following functions described in [37, Chapters 5 to 8] and [39, Sections 7 to 12], with potential for more:

   (a) The parsing of natural language (which is quite well developed); and understanding of natural language (which is only at a preliminary stage of development).

   (b) Pattern recognition which is robust in the face of errors of omission, commission, or substitution; and pattern recognition at multiple levels of abstraction.

   (c) Information retrieval which is robust in the face of errors of omission, commission, or substitution.

   (d) Several kinds of probabilistic reasoning, as summarised in Section A.9.2.

   (e) Planning such as, for example, finding a flying route between London and Beijing.

   (f) Problem solving such as solving the kinds of puzzle that are popular in IQ tests.

   The building of SP-multiple-alignments is also part of the process of unsupervised learning, next.

2. *Unsupervised learning.* Unsupervised learning, outlined in Appendix A.5, means the creation of one or two *SP-grammars* which are collections of SP-patterns which are effective in the economical encoding of a given set of New SP-patterns.

A.7 SP-Neural

A potentially useful feature of the SP System is that it is possible to see how abstract constructs and processes in the system may be realised in terms of neurons and their interconnections. This is the basis for *SP-Neural*, a 'neural' version of the SP System, described in [45].

The concept of an SP-symbol may realised as a *neural symbol* comprising a single neuron or, more likely, a small cluster of neurons. An SP-pattern maps quite well on to the concept of a *pattern assembly* comprising a group of interconnected SP-symbols. And an SP-multiple-alignment may be realised in terms of pattern assemblies and their interconnections, as illustrated in Figure 6.

In this connection, it is relevant to mention that the SP System, in both its abstract and neural forms, is quite different from DNNs [29] and has substantial advantages compared with such systems, as described in Section 2.2 and [46, Section V].

A.8 Generalising the SP System for two-dimensional SP-patterns, both static and moving

This brief description of the SP System and how it works may have given the impression that it is intended to work entirely with sequences of SP-symbols, like multiple sequence alignments in bioinformatics. But it is envisaged that, in future development of the system, two-dimensional SP-patterns will be introduced, with potential to represent and process such things as photographs and diagrams, and structures in three dimensions as described in [40, Section 6.1 and 6,2], and procedures that work in parallel as described in [41, Sections V-G, V-H, and V-I, and C].

It is envisaged that, at some stage, the SP System will be generalised to work with two-dimensional 'frames' from films or videos, and the sequencing needed to represent motion, and eventually the information needed to represent 3D bodies in motion, as in a 3D film.

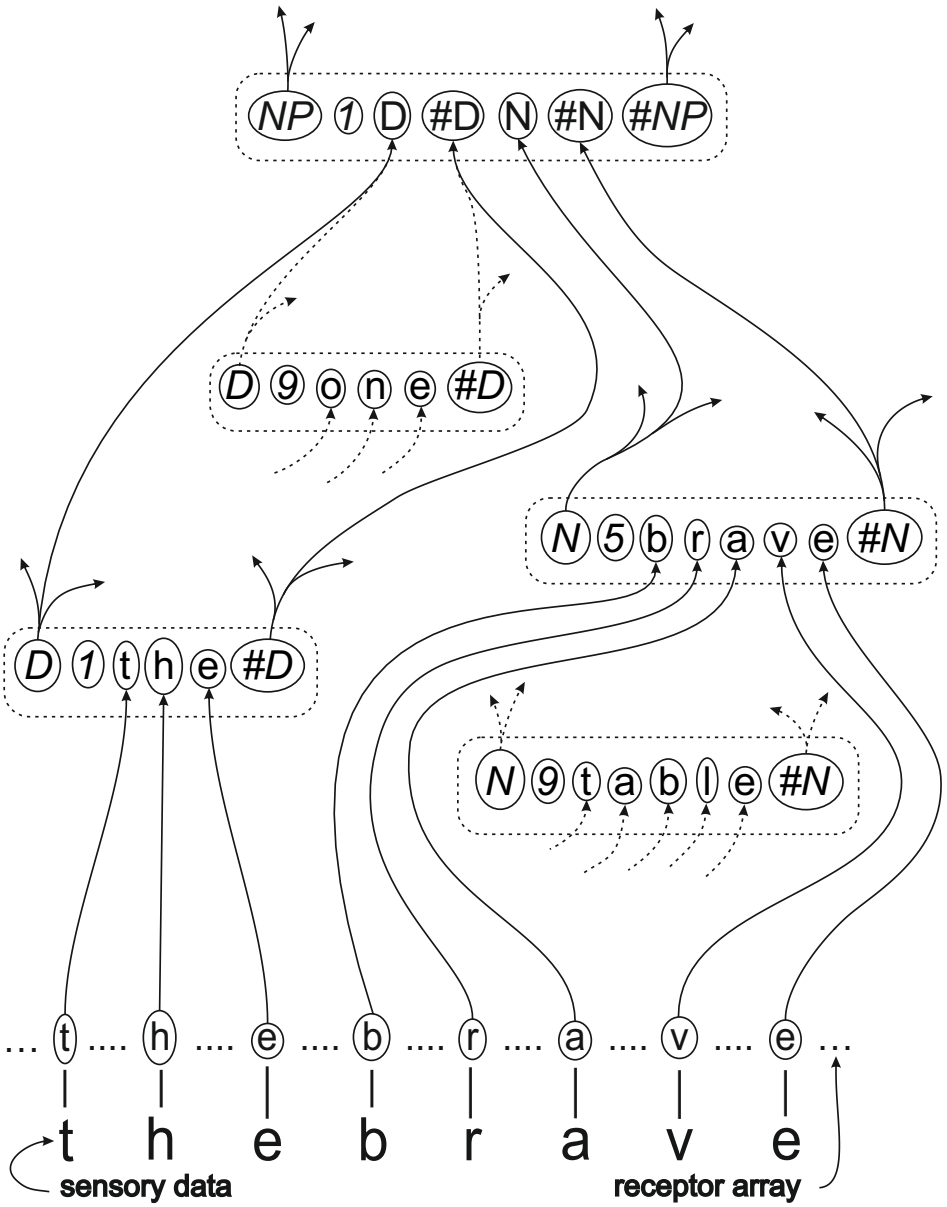

Figure 6: A schematic representation of a partial SP-multiple-alignment in SP-Neural, as discussed in [45, Section 4]. Each broken-line rectangle with rounded corners represents a *pattern assembly*—corresponding to an SP-pattern in the main SP Theory of Intelligence; each character or group of characters enclosed in a solid-line ellipse represents a *neural symbol* corresponding to an SP-symbol in the main SP Theory of Intelligence; the lines between pattern assemblies represent nerve fibres with arrows showing the direction in which impulses travel; neural symbols are mainly symbols from linguistics such as '`NP`' meaning 'noun phrase, '`D`' meaning a 'determiner', '`#D`' meaning the end of a determiner, '`#NP`' meaning the end of a noun phrase, and so on.

A.9 Strengths and potential of the SP System in AI-related functions

The strengths and potential of the SP System are summarised in the subsections that follow. Further information may be found in [39, Sections 5 to 12], [37, Chapters 5 to 9], [46], and in other sources referenced in the subsections that follow.

In view of the relative Simplicity of the SP System, the strengths and potential of the system summarised here mean that the system combines relative Simplicity with relatively high levels of descriptive and explanatory Power (Section 2.1.1 and Appendix A.1).

A.9.1 Versatility in aspects of intelligence

The SP System has strengths and potential in the 'unsupervised' learning of new knowledge. As noted in Appendix A.5, this is an aspect of intelligence in the SP System that is different from others because it is not a by-product of the building of multiple alignments but is, instead, achieved via the creation of *grammars*, drawing on information within SP-multiple-alignments.

Other aspects of intelligence where the SP System has strengths or potential are modelled via the building of SP-multiple-alignments. These other aspects of intelligence include: the analysis and production of natural language; pattern recognition that is robust in the face of errors in data; pattern recognition at multiple levels of abstraction; computer vision [40]; best-match and semantic kinds of information retrieval; several kinds of reasoning (next subsection); planning; and problem solving.

A.9.2 Versatility in reasoning

Kinds of reasoning exhibited by the SP System include: one-step 'deductive' reasoning; chains of reasoning; abductive reasoning; reasoning with probabilistic networks and trees; reasoning with 'rules'; nonmonotonic reasoning and reasoning with default values; Bayesian reasoning with 'explaining away'; causal reasoning; reasoning that is not supported by evidence; the inheritance of attributes in class hierarchies; and inheritance of contexts in part-whole hierarchies. Where it is appropriate, probabilities for inferences may be calculated in a straightforward manner ([37, Section 3.7], [39, Section 4.4]).

There is also potential in the system for spatial reasoning [41, Section IV-F.1], and for what-if reasoning [41, Section IV-F.2].

It seems unlikely that the features of intelligence mentioned above are the full extent of the SP System's potential to imitate what people can do. The close

connection that is known to exist between IC and concepts of probability (Appendix A.3), the central role of IC in the SP-multiple-alignment framework, and the versatility of the SP-multiple-alignment framework in aspects of intelligence suggest that there are more insights to come.

As noted in Appendix A.3, the probabilistic nature of the SP System makes it relatively straightforward to calculate absolute or conditional probabilities for results from the system, as for example in its several kinds of reasoning, most of which would naturally be classed as probabilistic.

A.9.3 Versatility in the representation of knowledge

Although SP-patterns are not very expressive in themselves, they come to life in the SP-multiple-alignment framework. Within that framework, they may serve in the representation of several different kinds of knowledge, including: the syntax of natural languages; class-inclusion hierarchies (with or without cross classification); part-whole hierarchies; discrimination networks and trees; if-then rules; entity-relationship structures [38, Sections 3 and 4]; relational tuples (*ibid.*, Section 3), and concepts in mathematics, logic, and computing, such as 'function', 'variable', 'value', 'set', and 'type definition' ([37, Chapter 10], [43, Section 6.6.1], [47, Section 2]).

As previously noted, the addition of two-dimensional SP patterns to the SP Computer Model is likely to expand the representational repertoire of the SP System to structures in two-dimensions and three-dimensions, and the representation of procedural knowledge with parallel processing.

As with the SP System's generality in aspects of intelligence, it seems likely that the SP System is not constrained to represent only the forms of knowledge that have been mentioned. The generality of IC as a means of representing knowledge in a succinct manner, the central role of IC in the SP-multiple-alignment framework, and the versatility of that framework in the representation of knowledge, suggest that the SP System may prove to be a means of representing *all* the kinds of knowledge that people may work with.

A.9.4 The seamless integration of diverse aspects of intelligence, and diverse kinds of knowledge, in any combination

An important third feature of the SP System, alongside its versatility in aspects of intelligence and its versatility in the representation of diverse kinds of knowledge, is that *there is clear potential for the SP System to provide SI.* This is because diverse aspects of intelligence and diverse kinds of knowledge all flow from a single coherent and relatively simple source: the SP-multiple-alignment framework.

It appears that SI is *essential* in any artificial system that aspires to the fluidity, versatility and adaptability of the human mind.

Figure 7 shows schematically how the SP System, with SP-multiple-alignment centre stage, exhibits versatility and integration. The figure is intended to emphasise how development of the SP System has been and is aiming for versatility and integration in the system.

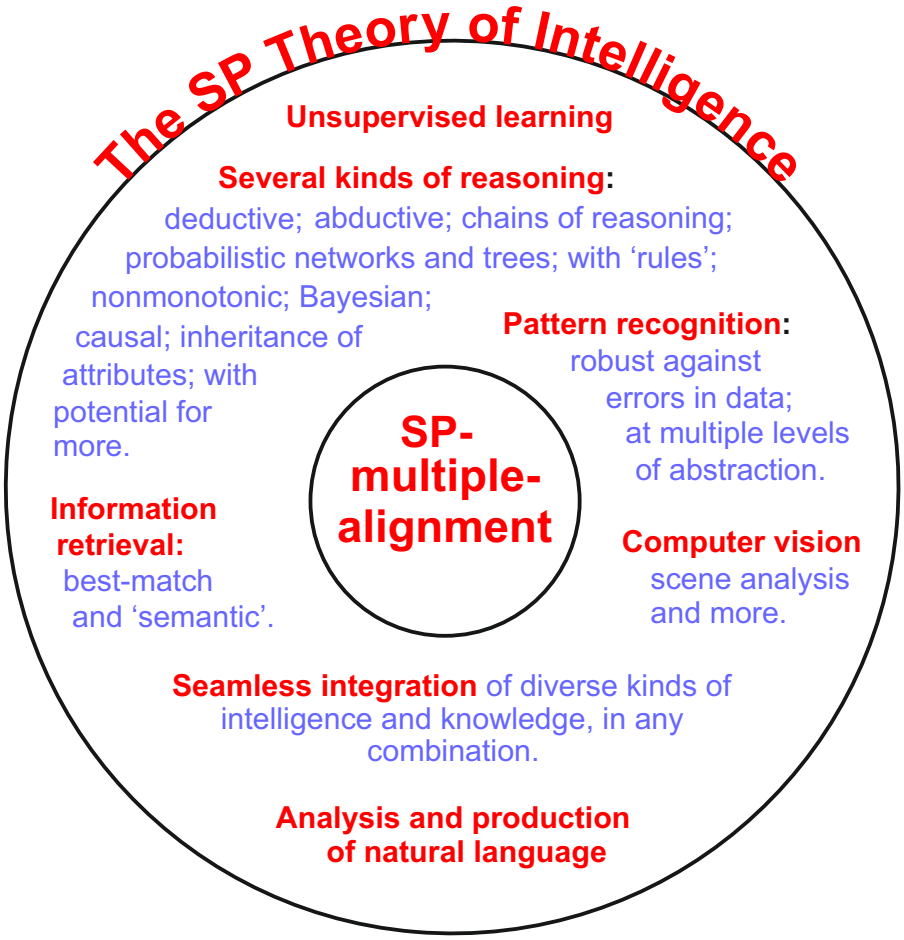

Figure 7: A schematic representation of versatility and integration in the SP System, with SP-multiple-alignment centre stage.

A.10 Potential benefits and applications of the SP System

Apart from its strengths and potential in modelling AI-related functions (Appendix A.9), it appears that, in more humdrum terms, the SP System has several potential benefits and applications, several of them described in peer-reviewed papers. These include:

- *Big data*. Somewhat unexpectedly, it has been discovered that the SP System has potential to help solve nine significant problems associated with big data [42]. These are: overcoming the problem of variety in big data; the unsupervised learning of structures and relationships in big data; interpretation of big data via pattern recognition, natural language processing; the analysis of streaming data; compression of big data; model-based coding for the efficient transmission of big data; potential gains in computational and energy efficiency in the analysis of big data; managing errors and uncertainties in data; and visualisation of structure in big data and providing an audit trail in the processing of big data.

- *Autonomous robots*. The SP System opens up a radically new approach to the development of intelligence in autonomous robots [41];

- *An intelligent database system*. The SP System has potential in the development of an intelligent database system with several advantages compared with traditional database systems [38]. In this connection, the SP System has potential to add several kinds of reasoning and other aspects of intelligence to the 'database' represented by the World Wide Web, especially if the SP Machine were to be supercharged by replacing the search mechanisms in the foundations of the SP Machine with the high-parallel search mechanisms of any of the leading search engines.

- *Medical diagnosis*. The SP System may serve as a vehicle for medical knowledge and to assist practitioners in medical diagnosis, with potential for the automatic or semi-automatic learning of new knowledge [36];

- *Computer vision and natural vision*. The SP System opens up a new approach to the development of computer vision and its integration with other aspects of intelligence. It also throws light on several aspects of natural vision [40];

- *Neuroscience*. As outlined in Appendix A.7, abstract concepts in the SP Theory of Intelligence map quite well into concepts expressed in terms of neurons and their interconnections in a version of the theory called *SP-Neural* ([45], [37, Chapter 11]). This has potential to illuminate aspects of neuroscience and to suggest new avenues for investigation.

- *Commonsense reasoning.* In addition to the previously-described strengths of the SP System in several kinds of reasoning, the SP System has strengths in the surprisingly challenging area of "commonsense reasoning", as described by Ernest Davis and Gary Marcus [4]. How the SP System may meet the several challenges in this area is described in [44].

- *Other areas of application.* The SP System has potential in several other areas of application including [43]: the simplification and integration of computing systems; best-match and semantic forms of information retrieval; software engineering [47]; the representation of knowledge, reasoning, and the semantic web; information compression; bioinformatics; the detection of computer viruses; and data fusion.

- *Mathematics.* The concept of IC via the matching and unification of patterns provides an entirely novel interpretation of mathematics [51]. This interpretation is quite unlike anything described in existing writings about the philosophy of mathematics or its application in science. There are potential benefits in science from this new interpretation of mathematics.

A.11 Unfinished business and the SP Machine

Like most theories, the SP Theory is not complete. Four pieces of 'unfinished business' are described in [39, Section 3.3]: the SP Computer Model needs to be generalised to include SP-patterns in two dimensions, with associated processing; research is needed to discover whether or how the SP concepts may be applied to the identification of low-level perceptual features in speech and images; more work is needed on the development of unsupervised learning in the SP Computer Model; and although the SP Theory has led to the proposal that much of mathematics, perhaps all of it, may be understood as IC [51], research is needed to discover whether or how the SP concepts may be applied in the representation of numbers. A better understanding is also needed of how quantitative concepts such as time, speed, distance, and so on, may be represented in the SP System.

It appears that these problems are soluble and it is anticipated that, with some further research, they can be remedied.

More generally, a programme of research is envisaged, with one or more teams of researchers, or individual researchers, to create a more mature *SP Machine*, based on the SP Computer Model, and shown schematically in Figure 8. A roadmap for the development of the SP Machine is described in [23].

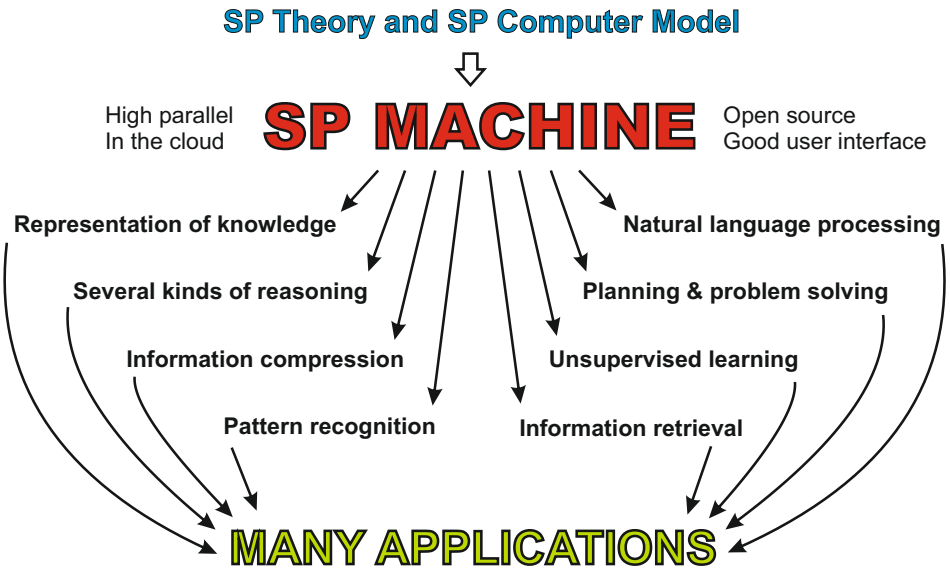

Figure 8: Schematic representation of the development and application of the SP Machine. Reproduced from Figure 2 in [39], with permission.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] P. Baldi. Autoencoders, unsupervised learning, and deep architectures. In *JMLR: Workshop and Conference Proceedings: Workshop on Unsupervised and Transfer Learning*, volume 27, pages 37–50, 2012.

[2] T. B. Brown, D. Mané, A. Roy, M. Abadi, and J. Gilmer. Adversarial patch. In *In the Proceedings of 31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017.

[3] N. Chomsky. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA, 1965.

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

[5] P. Domingos. *The Master Algorithm*. Allen Lane, London, Kindle edition, 2015.

[6] C. Edwards. Hidden messages fool AI. *Communications of the ACM*, 62(1):13–14, 2019.

[7] M. Ford. *Architects of Intelligence: the Truth About AI From the People Building It*. Packt Publishing, Birmingham, UK, Kindle edition, 2018.

[8] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations, 2019*.

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

[10] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio. An empirical investigation of catastrophic forgetting in gradient-based neural networks. Technical report, Département d'informatique et de recherche opérationnelle, Université de Montréal, 2015. arXiv:1312.6211v3.

[11] D. O. Hebb. *The Organization of Behaviour*. John Wiley & Sons, New York, kindle edition.

[12] M. Iklé, A. Franz, R. Rzepka, and B. Goertzel, editors. *Artificial General Intelligence*, volume 10999 of *Lecture Notes in Computer Science*, Heidelberg, 2018. Springer.

[13] J. E. Laird, C. Lebiere, and P. S. Rosenbloom. A standard model of the mind: toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, 38(4):13–26, 2017.

[14] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[15] H. J. Levesque. The Winograd Schema Challenge. In *Proceedings of the Tenth International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense-2011)*, 2011. Part of the AAAI Spring Symposium Series at Stanford University, March 21-23, 2011.

[16] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, New York, 4th edition, 2019.

[17] Y. Lv, Y. Duan, W. Kang, Z. Li, and F-Y. Wang. Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):865–873, 2015.

[18] G. Marcus. *Kluge: the Hapharzard Construction of the Human Mind*. Faber and Faber, London, paperback edition, 2008.

[19] G. F. Marcus and E. Davis. *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon Books, New York, Kindle edition, 2019.

[20] M. Minsky, editor. *The Society of Mind*. Simon & Schuster, New York, 1986.

[21] A. Newell, editor. *Unified Theories of Cognition*. Harvard University Press, Cambridge, Mass., 1990.

[22] A. Nguyen, J. Yosinski, and J. Clune. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In *Proceedings of the IEEE confernce on computer vision and pattern recognition (CVPR 2015)*, pages 427–436, 2015.

[23] V. Palade and J. G. Wolff. A roadmap for the development of the 'SP Machine' for artificial intelligence. *The Computer Journal*, 62:1584–1604, 2019. https://doi.org/10.1093/comjnl/bxy126, arXiv:1707.00614, bit.ly/2tWb88M.

[24] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin. Variational autoencoder for deep learning of images, labels and captions. In *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016)*, 2016.

[25] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G-Z. Yang. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21, 2017.

[26] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. The MIT Press, Cambridge, Mass., 1986.

[27] D. E. Rumelhart, J. L. McClelland, and PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1: Foundations. MIT Press, Cambridge Mass., 1986.

[28] D. E. Rumelhart, J. L. McClelland, and PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 2: Psychological and Biological Models. MIT Press, Cambridge Mass., 1995.

[29] J. Schmidhuber. Deep learning in neural networks: an overview. *Neural Networks*, 61:85–117, 2015.

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

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

[32] E. Strubell, A. Ganesh, and A. McCallum. Energy and policy considerations for deep learning in NLP. In *The 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy. July 2019*, 2019. arXiv:1906.02243v1 [cs.CL].

[33] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. Technical report, Google Inc. and others, 2014. arXiv:1312.6199v4 [cs.CV] 19 Feb 2014, bit.ly/1elzRGM (PDF).

[34] M. Tschannen, O. Bachem, and M. Lucic. Recent advances in autoencoder-based representation learning. In *The third workshop on Bayesian Deep Learning (NeurIPS 2018), Montréal, Canada*, 2018.

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

[36] J. G. Wolff. Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, unification and search. *Decision Support Systems*, 42:608–625, 2006. arXiv:1409.8053 [cs.AI], bit.ly/1F366o7.

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

[38] J. G. Wolff. Towards an intelligent database system founded on the SP theory of computing and cognition. *Data & Knowledge Engineering*, 60:596–624, 2007. arXiv:cs/0311031 [cs.DB], bit.ly/1CUldR6.

[39] J. G. Wolff. The SP Theory of Intelligence: an overview. *Information*, 4(3):283–341, 2013. arXiv:1306.3888 [cs.AI], bit.ly/1NOMJ6l.

[40] J. G. Wolff. Application of the SP Theory of Intelligence to the understanding of natural vision and the development of computer vision. *SpringerPlus*, 3(1):552–570, 2014. arXiv:1303.2071 [cs.CV], bit.ly/2oIpZB6.

[41] J. G. Wolff. Autonomous robots and the SP Theory of Intelligence. *IEEE Access*, 2:1629–1651, 2014. arXiv:1409.8027 [cs.AI], bit.ly/18DxU5K.

[42] J. G. Wolff. Big data and the SP Theory of Intelligence. *IEEE Access*, 2:301–315, 2014. arXiv:1306.3890 [cs.DB], bit.ly/2qfSR3G. This paper, with minor revisions, is reproduced in Fei Hu (Ed.), *Big Data: Storage, Sharing, and Security*, Taylor & Francis LLC, CRC Press, 2016, Chapter 6, pp. 143–170.

[43] J. G. Wolff. The SP Theory of Intelligence: benefits and applications. *Information*, 5(1):1–27, 2014. arXiv:1307.0845 [cs.AI], bit.ly/1FRYwew.

[44] J. G. Wolff. Commonsense reasoning, commonsense knowledge, and the SP Theory of Intelligence. Technical report, CognitionResearch.org, 2016. Submitted for publication, arXiv:1609.07772 [cs.AI], HAL: hal-01970147 v2, bit.ly/2eBoE9E.

[45] J. G. Wolff. Information compression, multiple alignment, and the representation and processing of knowledge in the brain. *Frontiers in Psychology*, 7:1584, 2016. arXiv:1604.05535 [cs.AI], bit.ly/2esmYyt.

[46] J. G. Wolff. The SP Theory of Intelligence: its distinctive features and advantages. *IEEE Access*, 4:216–246, 2016. arXiv:1508.04087 [cs.AI], bit.ly/2qgq5QF.

[47] J. G. Wolff. Software engineering and the SP Theory of Intelligence. Technical report, CognitionResearch.org, 2017. Submitted for publication. arXiv:1708.06665 [cs.SE], bit.ly/2w99Wzq.

[48] J. G. Wolff. Interpreting Winograd Schemas via the SP Theory of Intelligence and its realisation in the SP Computer Model. Technical report, CognitionResearch.org, 2018. Submitted for publication. bit.ly/2ME8DOA.

[49] J. G. Wolff. Commonsense reasoning, commonsense knowledge, and the sp theory of intelligence. Technical report, CognitionResearch.org, 2019. In preparation. viXra:1901.0051v2, hal-01970147 version 3, bit.ly/2RESeut.

[50] J. G. Wolff. Information compression as a unifying principle in human learning, perception, and cognition. *Complexity*, 2019:38 pages, February 2019. Article ID 1879746. viXra:1707.0161v3, hal-01624595 v2.

[51] J. G. Wolff. Mathematics as information compression via the matching and unification of patterns. *Complexity*, 2019:25, 2019. Article ID 6427493, Archives: vixra.org/abs/1912.0100 and hal.archives-ouvertes.fr/hal-02395680.