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*

June 10, 2019

Abstract

The book Architects of Intelligence by Martin Ford presents conversations about AI between the author and influential researchers. Unsolved problems in AI research that are described in the book are considered in relation to features of the SP System, meaning the SP Theory of Intelligence and its realisation in the SP Computer Model, both outlined in an appendix. The SP System has the potential to solve most of the problems in AI described in the book, and some others. Strengths and potential of the SP System, which in many cases contrast with weaknesses of deep neural networks (DNNs), include the following: a top-down research strategy has yielded a system with a favourable combination of conceptual Simplicity with descriptive or explanatory Power; the SP System has strengths and potential with both symbolic and non-symbolic kinds of knowledge and processing; the system has strengths and long-term potential in pattern recognition; it is free of the tendency of DNNs to make large and unexpected errors in recognition; the system has strengths and potential in unsupervised learning, including grammatical inference; the SP Theory of Intelligence provides a theoretically coherent basis for generalisation and the avoidance of under- or over-generalisations; that theory of generalisation may help improve the safety of driverless; the SP System

*Dr Gerry Wolff, BA (Cantab), PhD (Wales), CEng, MIEEE; ORCID ID 0000-0002-4624-8904. CognitionResearch.org, Menai Bridge, UK; jgw@cognitionresearch.org; +44 (0) 1248 712962; +44 (0) 7746 290775; Skype: gerry.wolff; Web: www.cognitionresearch.org.
system, unlike DNNs, can achieve useful learning from a single occurrence or experience; it has relatively tiny demands for computational resources and volumes of data, with potential for much higher speeds in learning; the system, unlike most DNNs, has strengths in transfer learning; unlike DNNs, it provides transparency in the representation of knowledge and an audit trail for all its processing; the system has strengths and potential in the processing of natural language; it exhibits several different kinds of probabilistic reasoning; the system has strengths and potential in commonsense reasoning and the representation of commonsense knowledge; other strengths include information compression, biological validity, scope for adaptation, and freedom from catastrophic forgetting. Despite the importance of motivations and emotions, no attempt has been made in the SP research to investigate these areas.

Keywords: Artificial intelligence; SP Theory of Intelligence; deep neural networks; unsupervised learning; generalisation; information compression.

1 Introduction

The book *Architects of Intelligence* by Martin Ford [7], presents conversations about AI between the author and people who are influential in the field, and yields some fascinating insights into current thinking of leading AI researchers.

This paper is not a review of the book as such. Rather, it 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.

The SP System has the potential to solve most of the problems in AI research described in [7], and some others.

After a ‘background’ section, next, the main sections that follow describe problems in AI with quotes from [7], and describe how the SP System may help solve those problems.

2 Background

This ‘background’ section firstly outlines some key features of the SP System and how it has been developed. And, secondly, because ‘deep neural networks’ (DNNs) are discussed in many parts of [7], the following subsection makes some introductory remarks about them.
2.1 Key features of the SP System and its development

This section summarises some key features of the SP System and how it has been developed. There is more detail in Appendix A.

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”.

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” (an expression which may be abbreviated as ‘SIIKAC’) — see also Appendix A.9.4.

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 an extended programme of research from about 1987 to now, with a break between early 2006 and late 2012. The main product has been the concept of SP-multiple-alignment (Appendix A.4 Section 2.1.4), within the SP System, developed via the creation and testing of many versions of the SP Computer Model. 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 may be read without recourse to those other sources but, inevitably, it cannot provide a comprehensive account of the SP System without being very much longer.

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. As far as possible in this paper, 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 and neuroscience: that much of HLPC may be understood as information compression (IC). Evidence for that principle is described in [52].

As noted in Appendix A.2, the principle accords with concepts of Minimum Length Encoding (MLE) (see, for example, [15]), 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 B). 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”.

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 17, 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 SIIKAC (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 which are described in [46, Section V], with some discussion of how those shortcomings may be overcome, but 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 AI, we should not put all our eggs in one basket. With relatively small resources, the SP System may be developed quite well alongside further research on DNNs.

3 A top-down strategy

“... one of [the] stepping stones [towards 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, trying to work from low-level concepts to over-arching principles.

There has been and still is research with top-down strategies: the quest for ‘unified theories of cognition’ (see, for example, [19] and later research such as [12]); and research aiming for ‘artificial general intelligence’ (AGI,
But it appears to be fair to say that, out of the several systems that have been developed, the SP System provides the most favourable combination of Simplicity and Power, achieving versatility as “an inelegant kluge of different capabilities”, as mentioned in Section 2.1.1.

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 many of the insights into the
nature of those symbolic systems that have been gained 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. 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 human learning, perception, and cognition:
  
  - 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’ [18] 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.

– The key difference between the SP System and other systems, symbolic or non-symbolic, is its 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’ [26, 27], 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 localised cluster of neurons or, less likely, a single neuron. Whether or not there are enough neurons in the brain to support this manner of representing knowledge is discussed in [37, Section 10.4.9].

- 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 [20].

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].

Naturally, there is research going on to try to solve these problems. But difficulties are that DNNs represent and process knowledge in a way that is not transparent (Section 13).

Experience with the SP System to date, and its transparency in both the representation and processing of knowledge (Section 13), 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.” Joshua 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 overgeneralisation (‘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 [16, 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. The two larger envelopes represents sets that are 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 $L$.

The difference in size between the smallest envelope and the middle-sized envelope represents correct generalisations. If a learning system creates a grammar that generates $L$ and some other utterances, then it overgeneralises, 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...
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.

things that are in \( L \). In the figure, the ‘illegal’ utterances that children hear are marked as ‘dirty data’.

A challenge in understanding how young children learn their first language is to explain:

- 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 [4], 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 its size 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? Probably the best answer at present is 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].

Figure 2 shows how the SP Computer Model may achieve a ‘correct’ analysis of the sentence ‘fortune favours the brave’ in much the same way as in Figure 5 despite errors of omission, commission, and substitution.

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 ‘fortune favours the brave’, there is the New SP-pattern ‘fortunefavourtresthebravete’ containing an error of omission (‘fortune’ instead of ‘fortune’), an addition (‘favours’ instead of ‘favours’), and a substitution (‘brave’ instead of ‘brave’).

This kind of recognition in the face of errors may be seen as a kind of generalisation. In this example: ‘fortune’ has been ‘generalised’ to ‘fortunefavourtresthebravete’, ‘favours’ has been ‘generalised’ to ‘favours’ and ‘brave’ has been ‘generalised’ to ‘bravete’. 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 na"ıve 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 streetlights 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 ‘generalisation via unsupervised learning’ in the SP Theory of Intelligence, as described in Section 8.1 is needed.
The chief merit of that theory of generalisation is that it is underpinned by a coherent body of theory which is supported by much evidence. It seems likely that, over reasonable amounts of time and across a range of driving conditions, generalisations that are made in accordance with the SP Theory would minimise the risk of accidents.

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.

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 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 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.

There is an example in Figure 5 where the New SP-pattern ‘fort unefavoursthebrave’ is analysed immediately in terms of pre-established Old SP-patterns; with the addition of semantic analysis (needing more work in the SP System), the meanings may enter immediately into reasoning, planning, problem solving, and the like. 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).

3. And:

   - The creation of a grammar (a collection of SP-patterns) that can achieve economical encoding of any small instance of N, can be done without undue computational demands (Appendix A.5);
   - 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 a grammar may be developed for much
larger bodies of information, including ‘all New information received by a given robot or other kind of AI since it was first created’; and because of this incremental development, the computational demands at any one stage may be relatively modest.

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. 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 12), 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 observation and with a little interaction with the world.” Yann LeCun [7, p. 124].

As mentioned above, this kind of transfer learning (Section 12) 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 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 19), SP-patterns that have been learned at any stage, will be available in the system’s repository of Old SP-patterns for use later:

- **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 ‘that girl runs’ and the Old SP-pattern ‘that boy runs’ would, with some simplifications, lead to the creation of SP-patterns like ‘X x0 that #X’, ‘Y y0 girl #Y’, ‘Y y1 boy #Y’, ‘Z z0 runs #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 on the table” and inferring that “blah” is a noun that means some kind of object that can be put on a table.

13 **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 more.

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.

It is true that similar things can be said about other symbolic systems. But, as mentioned in Section 4, the key difference between the SP System and other systems is its combination of conceptual Simplicity with explanatory and descriptive Power.

14 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

---

2See, for example, “Inside DARPA’s push to make artificial intelligence explain itself”, CET US News, 2017-08-10, bit.ly/2FQMoAr
equivalences that has been built up via human mark up and pattern matching, with English 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 of interest 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 Appendix A.4 (Figure 5), 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 syntactic dependencies in natural language [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 ([49], [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 SIIKAC. This feature of the system is likely to facilitate the smooth integration of syntax with semantics.

In general, as noted in Section 4, the key difference between the SP System and other systems is its combination of conceptual Simplicity with explanatory and descriptive Power.

15 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.

Much as with the processing of natural language (Section 14), a strength of the SP System is that there can be SIIKAC. And more generally, as noted in Section 4, the key difference between the SP System and other systems is its combination of conceptual Simplicity with explanatory and descriptive Power.

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

16 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 15 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 [28], 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 [51, Sections 4.1 to 4.3], 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 [51, Sections 5 to 7]: 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; and 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 [51, 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 [49], 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” [49].

17 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 non-linearity.” Yoshua Bengio [7, p. 26].

As noted in Section 2.1.3, IC 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, 16, 23, 24, 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,” [28, Section 5.10], IC gets only brief mentions in his review of research about DNNs [28, Sections 4.4, 5.6, 5.7, 5.10].
- Given the importance of IC in the SP System, in its theoretical foundations [46, 52], 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 [25] 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 encoding/decoding complexity? 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).

18 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 Hebb’s concept of a ‘cell assembly’ [10, 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” [13, 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 diverse agents [15] or a kluge deriving from the haphazard nature of evolution [17]: or 2) the product of some unifying
principle, as in the SP Theory; or 3) that it might be some combination of diverse agents with some unifying principle. Although the main focus in presenting the SP System is on the second 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):

- Demonstrates the fundamental connection between IC and human-like intelligence.
- Reflects the way in which IC is a unifying principle in HLPC [52].

By comparison, and notwithstanding the use of autoencoders in DNN systems (Sections 2.1.3 and 17), 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.

19 Catastrophic forgetting

Catastrophic forgetting—which is a problem for at least some DNNs—is when a given DNN has learned one thing and then it learns something else, there is a tendency for the new learning to wipe out the earlier learning (see, for example, [9]). A related problem is that, to be practical, a learning system, like a person, should be able to learn continuously from its environment without old knowledge being disturbed by new knowledge.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 adapted in that way.

As in Section 13, similar things can be said about other symbolic systems. But the key difference between the SP System and other systems is its combination of conceptual Simplicity with explanatory and descriptive Power.

3It 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
20 Scope for adaptation

In the SP System, the concept of SP-multiple-alignment, with the concept of SP-pattern, provide a much greater scope for modelling the world than the relatively constrained framework of DNNs. This is because 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, with unlimited SP-patterns, there is no limit to the number of ways in which SP-patterns may be connected within the framework of SP-multiple-alignments. By contrast, the layers in a DNN, and the potential connections amongst them, are finite and pre-defined. It is true that the connections can vary in strength but only within pre-defined limits.

21 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 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 ‘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.

22 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 reference 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 which combines simplicity with a versatility which is much wider than that of other AI system.

- **A central role for IC.** With its central role for IC, the SP System adopts a key principle from human cognitive psychology and neuroscience, for which there is much evidence.
• **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, and for its potential for SIIKAC.

• **SP-Neural.** Abstract concepts in the SP Theory map quite well into 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 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 the recognition of images and speech, its long-term potential appears to be greater.

• **Deep neural networks 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;

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

### 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 quite fully in [37]. Distinctive features and advantages of the SP System are described in [40]. Other papers in this programme of research are detailed, with download links, on [www.cognitionresearch.org/sp.htm](http://www.cognitionresearch.org/sp.htm).

In broad terms, the SP System is a brain-like system that takes in *New* information through its senses and stores some or all of it as *Old* information, as shown schematically in Figure 3.

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. 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,
Figure 3: Schematic representation of the SP System from an ‘input’ perspective. Reproduced, with permission, from Figure 1 in [39].
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 IC 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 [52].

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 [52, 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, [15]), 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 (Appendix B).

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 (Appendix B), 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 Appendix B 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 [52, 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.

```
| G G A | G | C A G | G | G | G | G | G | G | A |
| G | G | G | C | C | C | A | G | G | G | G | A |
| A | G | A | C | T | G | C | C | A | G | G | G | G | G | G | A |
| G | G | A | A | A | C | A | G | G | G | A | G | G | G | G | A |
```

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’—which produce results that are good enough for practical purposes.

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 information. A ‘good’ SP-multiple-alignment is one which allows the New SP-pattern to be encoded economically in terms of the Old SP-patterns.

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 (‘fortune favours the brave’) into its grammatical parts and sub-parts.

Figure 5: The best SP-multiple-alignment produced by the SP Computer Model with a New SP-pattern, ‘fortune favours the brave’, 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 (‘fortune favours the brave’) into its grammatical parts and sub-parts.

---

39

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

Although the search space for good SP-multiple-alignments is normally very large, 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. [8]).

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.

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 IC 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) 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: 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.

   (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 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 be realised as a neural symbol comprising a single neuron or, more likely, a small cluster of neurons. An SP-pattern maps quite well onto the concept of a pattern assembly comprising a group of inter-connected 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 [28](#) 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 sequences of two-dimensional ‘frames’ from films or videos.
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], [40], and in other sources referenced in the subsections that follow.

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 B), 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 (\textit{ibid.}, Section 3), and concepts in mathematics, logic, and computing, such as ‘function’, ‘variable’, ‘value’, ‘set’, and ‘type definition’ \[37\] Chapter 10], \[13\] Section 6.6.1], \[48\] 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 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 seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination. (SIKAC) 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 seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination, 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. 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 and more; the analysis of streaming data; compression of big data; model-based coding for 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 [10];

- **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; applications of natural language processing; best-match and semantic forms of information retrieval; software engineering [45]; 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 [47]. 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 [50], 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

48
A roadmap for the development of the SP Machine is described in [21].

B Information Compression, Inference, and Probability

Before we get on to discussions of issues relating to IC and probability, a preliminary subsection is needed to clarify the view of probability that has been adopted in the SP programme of research.

B.1 Frequentist and Bayesian views of probability

Perhaps because of the prominence of uncertainties in the way people think, much research in AI and cognitive science is based on concepts such as those in Bayes’ Theorem.

In brief, Bayes’ Theorem may summarized with the equation \( P(h|D) = \frac{P(D|h)P(h)}{P(D)} \), where \( P(h) \) = prior probability of hypothesis \( h \), \( P(D) \) = prior probability of training data \( D \), \( P(h|D) \) = probability of \( h \) given \( D \), and \( P(D|h) \) = probability of \( D \) given \( h \).
To be clear, the concept of probability in the SP programme of research differs sharply from the concept of probability in Bayesian theory because it depends entirely on the frequencies of occurrence of entities or events—it takes a ‘frequentist’ view of probability—whereas Bayesian theory adopts the view that probability is a degree of belief in an entity or event, and includes the concept of prior probability, meaning the degree of belief in an entity or event prior to the receipt of new data.

Of course, the SP theory, like other frequentist theories, may recognise probabilities that are prior to the receipt of new data, but such probabilities do not have any special theoretical status.

B.2 Asymmetry in the relationships between ICMUP and concepts of probability

It has been recognized for some time that there is an intimate connection between IC and concepts of probability [29, 30, 31, 15]. In the SP System, the very close connection between IC and concepts of probability makes sense in terms of ICMUP because:

- **IC via unification of patterns.** The unification of SP-patterns achieves compression of information.

- **Absolute probability.** Absolute probabilities may be derived from the frequencies of patterns, and, for a given pattern, its frequency of occurrence may be derived from the number of original patterns that have been unified to create that pattern.

- **Inference and conditional probability.** Inference may be achieved via partial matching as, for example, how seeing black clouds allows us to make the inference that rain is likely, via a partial match between ‘black clouds’ and the pre-established pattern ‘black clouds rain’.

- **(Deriving absolute and relative probabilities from frequencies of occurrence)** In general, absolute and relative probabilities in the SP System may be derived from the frequencies of occurrence of patterns ([39, Section 4.4], [37, Section 3.7]).

On the strength of evidence like this, we may infer that there is nothing to choose between IC and concepts of probability as a foundation for the development of the SP System, or any other artificial system that aspires to human-like intelligence. But for reasons outlined in the following subsections, there is an advantage in putting the main focus on IC.
B.2.1 Loss of information in the derivation of probabilities

As may be seen from points made above, there is an asymmetry between ICMUP and concepts of probability: absolute and conditional probabilities may be derived from the matching and unification of patterns, but the reverse is not true. This is partly because, arguably, the matching and unification of patterns is more primitive than concepts of probability. But more to the point, values for probability, in themselves, have lost information about the matches and unifications that led to their creation.

Because probabilities may be derived from unifications but not the other way round, and because the matching and unification of patterns is prominent in HLPC [52], any artificial system that aspires to the generality of human intelligence should be founded on ICMUP, not concepts of probability.

B.2.2 Potential for creating new structures via ICMUP

Much can be done with Bayesian and other probabilistic approach to AI, but something is missing: it is assumed that all of the conceptual entities in a probabilistic analysis have been created already, and there is nothing about how they may be formed. By contrast, the matching and unification of patterns in the SP System opens up the possibility of isolating words as discrete entities in speech [52, Section 15.1], and likewise for phrases [52, Section 15.2]. And it can provide a basis for the building of three-dimensional models of entities, as outlined in [40, Sections 6.1 and 6.2].

B.2.3 The scope of frequency information may be extended via ICMUP

It is often assumed that, when the frequency of occurrence of entities or events is used as the basis of probability measures, high frequencies are needed to ensure that results are statistically significant. Thus, for example:

“There is a definition of probability in terms of frequency that is sometimes usable. It tells us that a good estimate of the probability of an event is the frequency with which it has occurred in the past. This simple definition is fine in many situations, but breaks down when we need it most; i.e., its precision decreases markedly as the number of events in the past (the ‘sample size’) decreases. For sample sizes of 1 or 2 or none, the method is essentially useless.” [31 pp. 74–75].

By contrast, in searching for repeating patterns that may be unified to yield compression, the sizes of repeating patterns are as important as their
Maximising the amount of redundancy found means maximising $R$ where:

$$R = \sum_{i=1}^{n} (f_i - 1) \cdot s_i,$$

and where $f_i$ is the frequency of the $i$th member of a set of $n$ patterns and $s$ is its size in bits. Patterns that are both big and frequent are best.

But there is a trade-off between the sizes of patterns and the minimum frequency that is needed for IC. With small patterns, high frequencies are required. But with large patterns, useful compression can be achieved when frequencies that are as low as 2 or 3 [37, Sections 2.2.8.3 and 2.2.8.4].

As an example, one can, very often, recognize with high confidence a previously-heard song or other piece of music from hearing, only once, a smallish sample of the piece. Accordingly, we assign a mental probability of 1.0 to the identification we have made, a probability which corresponds to a frequency of 2, because the first learning-by-hearing of the music yields a frequency of 1, and the second recognition-by-hearing of the music yields a frequency of 2.

**B.2.4 Probability, causation, and structure**

Another apparent shortcoming of relying too much on concepts of probability arises if we wish to know about causation. For example:

“The answer [to difficulties in solving causal problems with statistics ...] has to do with the official language of statistics—namely the language of probability. This may come as a surprise to some [people] but the word cause is not in the vocabulary of probability theory; we cannot express in the language of probabilities the sentence mud does not cause rain—all we can say is that the two are mutually correlated or dependent—meaning that if we find one, we can expect the other. Naturally, if we lack a language to express a certain concept explicitly, we can’t expect to develop scientific activity around that concept.” [22, p. 342].

“[An engineering diagram] is, in fact, one of the greatest marvels of science. It is capable of conveying more information than millions of algebraic equations or probability functions or logical expressions. What makes [such a diagram] so much more powerful is the ability to predict not merely how the [system] behaves under normal conditions but also how [it] will behave under millions of abnormal conditions.” [22, p. 344].
In brief: any satisfactory account of causation requires a description of relevant structures, not merely numbers and equations.

Although no serious attempt has yet been made to examine issues in causality in terms of the SP Theory (but see [39] Section 10.5 and [37] Section 7.9), there are reasons to think that it may be more successful than classical statistics. This is because, in accordance with the second quote above, the SP System has the potential represent and to learn the kinds of structures that are needed for a comprehensive causal analysis.

Conflict of interest

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

References


56


