Problems in AI research and how the SP System may help to solve them

J Gerard Wolff*,†

September 2, 2020

Abstract

This paper describes problems in AI research and how the SP System (described in sources referenced in the paper) may help to solve them. Most of the problems considered in the paper are described by leading researchers in AI in interviews with science writer Martin Ford, and reported by him in his book *Architects of Intelligence*. These problems, each with potential solutions via SP, are: the divide between symbolic and non-symbolic kinds of knowledge and processing, and how the SP System may bridge the divide; the tendency of deep neural networks (DNNs) to make large and unexpected errors in recognition, something that does not happen with the SP System; in most AI research, unsupervised learning is regarded as a challenge, but unsupervised learning is central in how SP learns; in other AI research, generalisation, with under- and over-generalisation is seen as a problem, but it is a problem that has a coherent solution in the SP System ; learning usable knowledge from a single exposure or experience is widely regarded as a problem, but it is a problem that is already solved in the SP System; transfer learning (incorporating old knowledge in new) is seen as an unsolved problem, but it is bedrock in how the SP System learns; there is clear potential for the SP System to solve problems that are prevalent in most AI systems: learning that is slow and greedy for large volumes of data and large computational resources; the SP System provides solutions to problems of transparency in DNNs, where it is difficult to interpret stored knowledge and how it is processed; although there have been successes with DNNs in the processing of natural language, the SP System has strengths in the representation and processing of natural languages which appear to be more in accord with how people process natural language, and these strengths in the SP System are well-integrated with other strengths of the system in aspects of intelligence; by contrast with DNNs, SP has strengths and potential in human-like probabilistic reasoning, and these are well integrated with strengths in other aspects of intelligence; unlike most DNNs, the SP System eliminates the problem of catastrophic forgetting (where new learning wipes out old learning); the SP System provides much of the generality across several aspects of AI which is missing from much research in AI. The strengths and potential of the SP System in comparison with alternatives suggest that *the SP System provides a relatively promising foundation for the development of artificial general intelligence*.

*CognitionResearch.org, UK

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

1 Introduction

This paper describes some problems in artificial intelligence (AI) research and how the *SP System* may help to solve them.

An introduction to the SP System, with pointers to where fuller information may be found is in Section 1.2, below.

The relevance of this research to the study of complexity is outlined in Section 1.3, below.

Some other problems and potential solutions may be found in the technical report [32], with a fairly full outline description of the SP System [32, Appendix A].

1.1 Preliminaries

Most of the problems in AI research considered in this paper are described by leading researchers in AI in interviews with science writer Martin Ford, and reported by him in his book *Architects of Intelligence* [7].

This paper is not a review of Ford's book. Rather, it is an extended response to the very interesting views about problems in AI research that Ford has gathered in his book, and some other problems in AI research described elsewhere. Owing to the considerable influence in AI research of deep neural networks (DNNs), most of the problems considered are problems with DNNs.

In this paper, each problem in AI research is described in its own section, with a description of how SP System may help to solve it. The paper performs an important task that is needed in any field of science: the evaluation of problems in the field and how they may be solved.

1.2 Information about the SP System

The *SP System*, meaning the *SP Theory of Intelligence* and its realisation in the *SP Computer Model, is the product of a lengthy programme of research, seeking to simplify and integrate observations and concepts across AI, mainstream computing, mathematics, and human learning, perception, and cognition* (HLPC). The SP System has approached AI via the modelling of HLPC, and may be seen to be largely a model of HLPC.

Since since computing and mathematics (and logic) are the products of human thinking and have been designed as aids to human thinking, they may be seen to fall within the scope of the SP System as a model of HLPC (more later).

Information compression (IC) is fundamental in the workings of the SP System [32, Appendices A.2 and A.3]. More specifically, the powerful concept of SP-multiple-alignment (SPMA) [32, Appendix A.4], a generalisation of six other techniques for IC [31, Section 5.7], is bedrock in the workings of the SP System.

The SP System is described in outline in [32, Appendix A], in more detail in [24], and even more fully in [22]. There are also outline descriptions of the SP System in [31, Section 3] and [30, Section 2.2].

1.3 Relevance of this research to the study of complexity

Since the human brain may be seen as a complex system "characterized by interactions between [its] components that produce new information—present in neither the initial nor boundary conditions—which limit [its] predictability",[1] the SP System as a model of HLPC (Section 1.2), and IC as fundamental in the workings of the SP System (*ibid.*), may be seen as a means of taming some of the complexity in the workings of the human brain.

That conclusion is strengthened by evidence, presented in this paper, in support of the SP System.

Similar things may be said about two other papers:

- The paper [30] describes an array of relatively direct evidence, largely independent of the SP System, for the importance of IC in HLPC.
- The paper [31] describes evidence that much of mathematics, perhaps all of it, may be seen as a set of techniques for IC, and their application. It also describes evidence that logic and computing may be understood in the same terms.

  Since, as noted in Section 1.2, computing, mathematics, and logic, may be seen as products of human thinking, and they are designed to aid human thinking, evidence presented in [31] may be seen as further evidence for the importance of IC in HLPC.

This paper, with the two papers just described, are mutually supportive, providing distinct streams of evidence for IC as a means of making sense of some of the complexity in HLPC, which itself represents some of the complexity in the workings of the human brain.

1.4 Abbreviations

Abbreviations used in this paper are detailed below. Since readers may wish to approach topics in various directions, all abbreviations are defined in this one place as well as where they are first used. Each of the following items shows the section in this paper where the abbreviation is first used.

- *Artificial General Intelligence* (Section 1): 'AGI'.
- *Artificial intelligence* (Section 1): 'AI'.
- *Deep Neural Network* (Section 1.1): 'DNN'.
- *Human Learning, Perception, and Cognition* (Section 1.2): 'HLPC'.
- *Information compression* (Section 1.2): 'IC'.

[1] From "About this Journal," "Aims and scope," for the journal *Complexity* at tinyurl.com/y4o3hwt3.

- *SP-multiple-alignment* (Section 1.2): 'SPMA'.

It is intended that 'SP' should be treated as a name. Reasons for that name are given in [32, Appendix A.1].

2 The need to bridge the devide between symbolic and non-symbolic kinds of knowledge and processing

This main section, and each of the main sections up to and including Section 13, discusses one significant problem in AI research, and describes potential solutions via the SP System. In most cases, as in this section, we begin with one or more quotes from leading researchers in AI, as reported in Martin Ford's book, *Architects of Intelligence* [7].

> "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:

- Given that people can and do learn and use symbolic systems like natural languages, mathematics, and logic, it is clear that our brains have the capacity to represent and to process those kinds of symbolic information.
- At the same time, any theory of AGI should be true to the kinds of things that can be done by: babies before they learn any symbolic system including natural language (eg, learning to walk); or adults in performing putatively 'non-symbolic' skills such as recognising things, playing tennis, making bread, and so on; or animals when they are engaged in such activities as foraging for edible plants, hunting prey, swinging from branch to branch through trees, and so on.

- In the light of those two points, something is needed that bridges symbolic and non-symbolic kinds of knowledge and processing.
- The SP System provides a framework that is showing promise in those two areas:
  - The concept of *SP-symbol* in the SP System [32, 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-pattern* [32, Appendix A], with the concept of SPMA [32, Appendix A.4], provides a versatile framework for diverse aspects of intelligence [32, Appendix A.9.1] including several forms of reasoning [32, Appendix A.9.2], and the representation of diverse kinds of knowledge [32, Appendix A.9.3].
  - The concept of SPMA also facilitates the seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination [32, Appendix A.9.4], a kind of integration that appears to be essential in any system that aspires to AGI.
  - As outlined in [32, Appendix A.6], 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 the potential to imitate the all-or-nothing nature of much of mathematics and logic [31, Section 8].
  - The SP System has compression of information as a unifying principle, a principle which applies to both symbolic and non-symbolic kinds of knowledge.
  - More specifically, the SP System conforms to the 'DONSVIC' principle (the 'Discovery of Natural Structures via Information Compression') [24, Section 5.2]. This means that the system can discover relatively 'symbolic' kinds of things such as words in otherwise non-symbolic information such as natural language without explicit markers for the beginnings and ends of words—as is the case with speech, and is also the case with text without punctuation or spaces between words (*ibid.*). It seems likely that similar principles apply to the discovery of 'objects' via vision [25, Sections 6.1 and 6.2].

3 The tendency of deep neural networks to make large and unexpected errors in recognition

> "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].

> "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].

Although deep neura networks (DNNs) often do well with images and speech (as described in the first quote above), they can make surprisingly big and unexpected errors in recognition (as described in the second quote). 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 [20]. 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 [15].

In a laboratory setting, errors like these may be interesting or even amusing. But to the extent that DNNs come into use in applications in, for example, commerce, administration, and defence, the consequences of errors can be expensive, or dangerous, or both those things:

> "... it is relatively easy to force [DNNs] to make mistakes that seem ridiculous, but with potentially catastrophic results. Recent tests have shown [how] 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].

Of course, there is potential for improvement in the performance of DNNs. But it seems that fixing problems in DNNs is difficult or impossible because of a fundamental weakness in them: that they represent and process knowledge in a way that is not transparent (Section 9). From an SP perspective, problems with DNNs arise because they do not have the tight focus on IC that is central in the workings of the SP System; and because the basic organisation of DNNs has long been recognised as only a guess at how real nervous systems work.

From experience with the SP System to date, and from its transparency in both the representation and processing of knowledge (Section 9), it seems very unlikely that the system would be vulnerable to the kinds of mistakes made by DNNs.

3.1 Identification of parsings that are good, despite errors of omission, commission, and substitution

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, as can be seen in Figure 2 (b) (see also [24, Section 4.2.2], [23, Section 2.2.2]).

This strength in the SP System arises from the way it calculates absolute and relative probabilities for each SPMA that it creates, and this allows it to identify the best (most probable) analysis out of many alternatives, even though the best analysis may not be perfect.

4 The challenges 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 AGI.*" Martin Ford [7, pp. 11–12], emphasis 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].

From these remarks it can be seen that the development of unsupervised learning is regarded as an important challenge for AI research today. So it should be of interest that the SP Computer Model 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 ([24, Section 5], [22, Chapter 9]).

There are shortcomings in how the SP Computer Model learns [32, Appendix A.6.3]. However, they appear to be soluble, and their solution would greatly enhance the capabilities of the SP Computer Model in unsupervised learning (*ibid.*).

In the SP programme of research, unsupervised grammatical inference is regarded as a paradigm or framework for other kinds of unsupervised learning, not merely the learning of syntax. In addition to the learning of syntactic structures [16, Section 12.3], it may, for example, provide a model for the unsupervised learning of non-syntactic semantic structures [16, Section 12.1], and for learning the integration of syntax with semantics [16, Section 12.3]. Unsupervised learning in the SP System may itself be a good foundation for 'supervised' kinds of learning, such as learning by being told, learning by imitation, learning via rewards and punishments, and so on.

5 The need for a coherent account of generalisation, under-generalisation, and over-generalisation

> "Many of us think that we are ... missing the basic ingredients needed [for generalization], such as the ability to understand causal relationships in data—an ability that actually enables us to generalize and to come up with the right answers in settings that are very different from those we've been trained in." Yoshua Bengio [7, p. 18].

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

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

An important issue in learning is how to generalise 'correctly' from the specific information which provides the basis for learning, without over-generalisation ('under-fitting') or under-generalisation ('over-fitting'). This issue is discussed quite fully in [24, Section 5.3] (see also [22, Section 9.5.3] and [28, Section V-H]). Because it is an important issue, the main elements of the solution proposed in the SP Theory of Intelligence are described here.

In the SP Theory of Intelligence, 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.

5.1 Generalisation via unsupervised learning

The generalisation issue arises quite 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.[2] That finite sample is shown as the smallest envelope in the figure.

[2]It is assumed here that the learning we are considering is unsupervised. This is because of evidence that, although correction of errors by adults may be helpful, children are capable of learning a first language without that kind of error correction ([12, 1]).

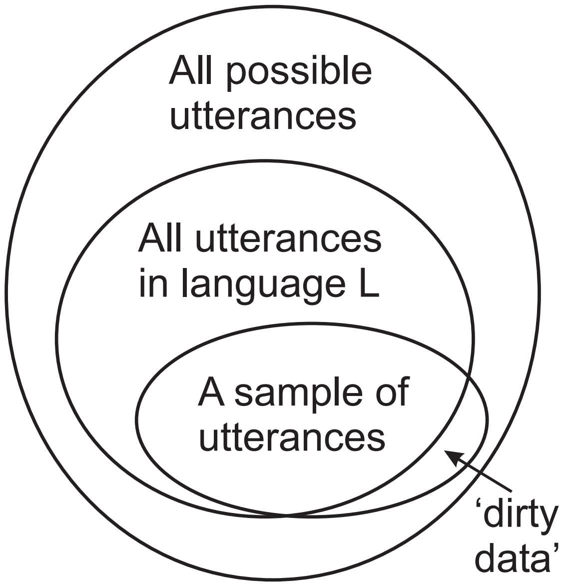

Figure 1: Categories of utterances involved in the learning of a first language, **L**. In ascending order of size, they are: the finite sample of utterances from which a child learns; the (infinite) set of utterances in **L**; and the larger (infinite) set of all possible utterances. Adapted from Figure 7.1 in [21], with permission.

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

Each of the two larger envelopes represents a set that is infinite in size but, in accordance with principles pioneered by Georg Cantor (see, for example, [5]), the set of all possible utterances is larger than the set of utterances in **L**. This is much like the way that the set of all integers is larger than the set of even integers, or the set of odd integers, although each of those sets is infinite in size.

The difference in size between the smallest envelope and the middle-sized envelope represents correct generalisations. If a learning system creates a grammar that generates **L** plus some other utterances, then it over-generalises, and if the grammar generates a subset of the utterances in **L**, then it under-generalises.

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

Challenges in understanding how young children learn their first language (or languages) are:

- How they generalise correctly without over-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.

- Young children may also under-generalise, but such under-generalisations would be difficult to distinguish from the fact that there is much of the language they are learning, because they are young.

- As with over-generalisations, it seems that children eventually learn **L** without their learning being distorted or corrupted by dirty data. Anything that is 'wrong' that they do learn would probably be regarded as dialect rather than errors in learning.

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

For reasons that would take us too far afield to explain, 'nativist' theories of first language learning, such as that proposed by Noam Chomsky [4] and others, will not suffice.

What follows is a summary of what is envisaged for the development in the SP Theory of Intelligence, and which is already largely incorporated in the SP Computer Model:

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** into 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*, **E**, of **I** in terms of **G**, where the size of **E** is e. Clearly, $t = g + e$.

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

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

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

Informal tests with a program for unsupervised learning, 'SNPR' [21], and with the SP Computer Model [22, 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 competence in speaking or writing a natural language, 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, and largely incorporated in the SP Computer Model, is likely to provide generalisations that are as close to being optimum as may be achieved with the available computational resources.

5.2 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 [25, Section 4].

The example here makes reference, first, to Figure 2 (a) below, which shows how an SPMA may achieve the effect of parsing the sentence '`t w o k i t t e n s p l a y`' in terms of grammatical categories, including words.

To illustrate generalisation via perception, Figure 2 (b) shows how the SP Computer Model, with the New SP-pattern with errors, '`t o k i t t e m s p l a x y`', may achieve what is intuitively a 'correct' analysis of the sentence despite the errors, which are described in section the caption for section (b) of the figure.

This kind of recognition in the face of errors may be seen as a kind of generalisation, where an incorrect form is generalised to the correct form.

There is relevant discussion in [25, Sections 4.1 and 4.2].

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

(a)

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

(b)

Figure 2: (a) The best multiple alignment created by the SP Computer Model with a store of Old patterns like those in rows 1 to 8 (representing grammatical structures, including words) and a New pattern, '`t w o k i t t e n s p l a y`', shown in row 0, representing a sentence to be parsed. (b) As in (a) but with errors in the sentence in row 0 ('`t o k i t t e m s p l a x y`'): an error of omission ('`t o`' instead of '`t w o`'), an error of substitution ('`k i t t e m s`' instead of '`k i t t e n s`'), and an error of addition ('`p l a x y`' instead of '`p l a y`'). Adapted from Figures 1 and 2 in [23], with permission.

6 How to learn usable knowledge 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].

Most DNNs incorporate some variant of the idea that, in learning, neural connections are gradually strengthened or weakened, a concept of learning which often draws its inspiration, directly or indirectly, from Donald Hebb's concept of learning:

> "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased." [9, Location 1496].

or, more briefly, "Neurons that fire together, wire together."

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 usable knowledge from a single occurrence or experience.

Unsupervised learning in the SP System ([32, Appendix A.4], [24, Section 5], [22, Chapter 9]) is quite different from the gradualist kinds of learning in DNNs. By contrast, with the SP model of learning, as noted in Section 8.1, unsupervised learning in the SP system accommodates both the learning of usable knowledge from a single exposure or experience, and the slow learning of complex knowledge or skills:

- *One-shot learning.* Any of the newly-received knowledge from the system's environment may serve immediately in any kind of thinking or activity. If we ask the way from someone, the guidance that we get can be put into effect at once, without the need for any kind of gradual strengthening of links. If a child touches something hot, he or she is likely to retain what they have learned for the rest of their lives, without the need for repetition. We may remember winning big on the lottery without the need for repeated experiences to drum it in (welcome as that may be!).

- *The relatively slow learning of complex knowledge or skills.* The SP system will, like a person, be relatively slow in the learning of a complex body knowledge or skill, partly because there is a lot of it, and partly because that kind of learning requires a time-consuming search through a large abstract space of ways in which the knowledge may be structured in order to compress it and thus arrive at an efficient configuration.

It is true that DNNs require the taking in of information supplied by the user, and may thus be said to have learned something from a single exposure or experience. But unlike the SP System, that knowledge cannot be used immediately, as illustrated by the parsing of a New sentence (fresh from the system's environment) in the parsing shown in Figure 2 (a). By contrast, any information taken in by a DNN at the beginning of a training session does not become usable until much more information has been taken in.

7 How to achieve 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 learning—is fundamental in the SP System.

Because the system does not suffer from catastrophic forgetting (Section 12), and because one-shot learning is possible and normal (Section 6), SP-patterns that have been learned at any stage, will be available in the system's repository of Old SP-patterns for use later. Some of the several ways in which previously-learned Old SP patterns may help with transfer learning is illustrated in the examples described here:

- *Partial matching in unsupervised learning.* In unsupervised learning in the SP System [32, A.6], partial matching between, for example, the New SP-pattern '`t h a t g i r l r u n s`' and the Old SP-pattern '`t h a t b o y r u n s`' would, with some simplifications, lead to the creation of SP-patterns like '`X x0 t h a t #X`', '`Y y0 g i r l #Y`', '`Y y1 b o y #Y`', '`Z z0 r u n s #Z`', and '`A a0 X #X Y #Y Z #Z`', all of which are added to the repository of Old SP-patterns. In this case, the Old SP-pattern has fed into the analysis of the New SP-pattern, leading to the creation of five additional SP-patterns.

- *Strong compositionality.* In unsupervised learning in the SP System, an SP-pattern that has been learned at any stage may, via 'references' between SP-patterns, become part of any SP-pattern that is learned later. This may yield strong compositionality in concepts, as outlined in [32, Section 6].

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

In DNNs, the reason why new learning fails to take advantage of old learning is simply that new learning normally wipes out old learning in 'catastrophic forgetting', described in Section 12, below.

8 How to increase the speed of learning in AI systems, and how to reduce the demands of AI learning for large volumes of data, and for large computational resources

> "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].
>
> "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].
>
> "[A] stepping stone [towards artificial general intelligence] 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 [artificial general intelligence] for me." Oren Etzioni [7, p. 502].

In connection with the large volumes of data and large computational resources that are often associated with the training of DNNs, it has been discovered by Emma Strubell and colleagues [19] 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.

Issues mentioned in the quotes above will considered in subsections that follow. They are: (1) The relatively slow speeds of learning with DNNs, (2) The relatively large volumes of data required by DNNs, and (3) The large computational resources that are often associated with the training of DNNs. These things seem to conflict with how people can learn fast with relatively little data, and a brain that runs with less than 20 watts [11].

8.1 Speeds of learning by people

The relatively high speed of learning by people compared with DNNs may be explained, at least in part, by the way that the SP Computer Model can learn things that are usable from a single exposure or experience (Section 6), whereas DNNs typically require many examples to learn something that is usable.

8.1.1 Fast learning via transfer learning

Relatively fast learning by people may also be explained by the SP System's good use of transfer learning (Section 7). If most of what you are learning is already known, there is relatively little to be learned with a new example.

It is probably also relevant to mention human capabilities in generalisation, both via unsupervised learning and via perception, as discussed in Section 5. If what you are learning is a generalisation of something you already know, there is, in principle, nothing more to be learned.

8.1.2 Slow learning

However, it seems that the kinds of fast learning just mentioned, only apply to relatively simple things (see Section 6). We should bear in mind that people can spend many years in school and college and later, building up their knowledge. There seem to be three main factors here:

- The volumes of data to be absorbed in school and beyond are large. But a computer can can upload and retain an electronic copy of whole text book in a few seconds or minutes, so there must be something else going on, as suggested in the next two points.
- Unlike a computer, people forget things. This may have a part to play in how people organise their knowledge (next).

- Unlike a computer, it seems that people put a lot of effort into organising their knowledge—perhaps during sleep, and perhaps by building structures like SP-grammars via IC in unsupervised learning ([32, Appendix A.6])—and that can slow things up, particularly with complex bodies of knowledge and complex skills.

Here, one would expect the SP System, like people, to be relatively slow.

8.2 Volumes of data, and computational resources

In the light of points made in Section 8.1, it seems that volumes of data and computational resources needed for learning may be considered together.

8.2.1 How to achieve fast learning of things that are simple

As noted in the previous subsection, the main problem with DNNs compared with people is to do with the learning of simple things: DNNs typically need a lot of examples to learn something like the concept of a cat, with correspondingly large demands for data and for computer power.

In the light of what is said in the previous subsection, it seems that the SP System provides the broad form of an answer to these problems: via one-shot learning (Section 6); via transfer learning (Section 7); and via generalisation (Section 5).

8.2.2 With complex information, how to reduce the processing power needed at any one time

But with complex bodies of knowledge or complex skills—such as learning the syntax and semantics of a natural language, learning to play pool, billiards, or snooker at expert levels, passing exams to be a medical doctor, and so on—the SP System, like people, may require significant capacity for the processing of information.

But in principle the SP System, like people, may use time as a partial alternative to capacity for the processing of data. If information is processed incrementally, then the computing power needed at any one time can be reduced. This is particularly relevant where the data to be processed is acquired incrementally, such as experience with weather forecasting, experience in farming, and so on.

9 The need for 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 for people to understand. These deficiencies are of concern for reasons of safety, legal liability, fixing problems in systems that use DNNs, and perhaps more.[3]

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 (see [31, Section 5]). And there is an audit trail for all processing in the SP System, so that it is explicit and comprehensible by people.

10 The need to strengthen the representation and processing of natural languages

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

DNNs can do well in recognising speech. Also, they can produce impressive results in the translation of natural languages using a database of equivalences between surface structures that has been built up via human mark up and pattern matching, with English in some cases as a bridge between languages that are not English. And impressive results have been achieved with 'zero-shot' learning[4] [18].

Despite successes like these, DNNs are otherwise weak in processing natural languages: they have no place for the kinds of syntactic structures that are recognised

[3] See, for example, "Inside DARPA's push to make artificial intelligence explain itself", *The Wall Street Journal*, 2017-08-10, tinyurl.com/y9vrov2k.

[4] A zero-shot learning method aims to solve a task without receiving any example of that task at training phase.

in theoretical linguistics, and which appear to be significant in human processing of language; they do not model the way in which people translate surface forms into meanings, and translate meanings into surface forms; and in such tasks as translation between languages, DNNs do not model the way in which, for people, meanings serve as an interlingua between different 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.

10.1 Parsing via SP-multiple-alignment

As can be seen from the example in Figure 2 (a) (Section 5.2), the SPMA construct in the SP Computer Model lends itself well to the representation of syntactic knowledge and to its application in the parsing of a natural language sentence.

There is a lot more detail about the processing of natural language in the SP System in [24, Section 8] and in [22, Chapter 5].

10.2 Discontinuous dependencies

SPMAs may also represent and process discontinuous syntactic dependencies in natural language such as the dependency between the 'number' (singular or plural) of the subject of a sentence, and the 'number' (singular or plural) of the main verb, and that dependency may bridge arbitrarily large amounts of intervening structure ([24, Section 8.1] and [22, Section 5.4]).

In Figure 2 (a), a number dependency (plural) is marked by the SP-pattern in row 8: '`Num PL ; Np Vp`'. Here, the SP-Symbol '`Np`' marks the noun phrase as plural, and the SP-symbol '`Vp`' marks the verb phrase as plural.

In a similar way, the SPMA construct may mark the gender dependencies (masculine or feminine) within a sentence in French, and, within the same sentence, those two kinds of dependency may overlap without interfering with each other, as can be seen in [22, Figure 5.8 in Section 5.4.1].

Arguably, this method of marking syntactic dependencies in natural language is simpler and better integrated with processes for the parsing and production of language than in other systems for computational linguistics. More to the point in the present discussion, modelling discontinuous dependencies in the syntax of natural languages appears to be well beyond anything that DNNs can do.

10.3 Other strengths of the SP System in the processing of natural language

In connection with the processing of natural language, it is appropriate to mention some other strengths of the SP Computer Model:

- *Robust against errors of omission, commission, and substitution.* As can be seen in Figure 2 (b), and described in Section 3.1, the SP System has robust abilities to arrive at an intuitively 'correct' parsing despite errors of omission, commission, and substitution, in the sentence being parsed. Naturally, there is a limit to how many errors can be tolerated. There is more detail in Section 5.2.

- *The generality of SP-patterns with SP-multiple-alignments.* In case the examples shown may have given the impression that the SP System is only good for the processing of natural language, it as well to emphasise that the system has strengths and potential in several different aspects of AI [24, 22], including several kinds of probabilistic reasoning ([24, Section 10], [22, Chapter 7]), and it lends itself to the representation and processing of a wide variety of other kinds of knowledge, and for the seamless integration of diverse aspects of intelligence, with diverse kinds of knowledge, in any combination. These strengths arise from the generality of SP-patterns, with SPMAs, as a means of representing diverse kinds of knowledge, and the generality of SPMAs for modelling diverse aspects of AI.

- *These strengths may help to promote the coordination of syntax with semantics.* Because of the versatility of SP-patterns with SPMAs, there is one framework which lends itself to the representation and processing of both the syntax and semantics of natural languages. It is likely that this will facilitate the coordination of syntax with semantics, so that surface forms may be translated into meanings, and *vice versa*. A preliminary example of how the SP Computer Model may support the coordination of syntax and semantics, in both the analysis and production of natural language, is shown in [22, Section 5.7]. There are more examples in [29].

- *One mechanism for both the parsing and production of natural language.* A neat feature of the SP System is that the production of natural language may be achieved by the application of IC, using *exactly* the same mechanisms as are used for the parsing or understanding of natural language. An explanation of how this is possible is given in [24, Section 4.5]) and [22, Section 5.7.1].

With the possible exception of the first item listed above, it seems that these kinds of capability are well beyond what can be done with DNNs.

11 How to achieve human-like capabilities in 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 probabilistic reasoning can be demonstrated, without any special provision or adaptation. ([24, Section 10], [22, Chapter 7]).

The kinds of probabilistic reasoning that can be demonstrated with the SP Computer Model are: one-step 'deductive' reasoning; abductive reasoning, reasoning with probabilistic decision networks and decision trees, reasoning with 'rules', nonmonotonic reasoning and reasoning with default values, causal diagnosis, and reasoning which is not supported by evidence. Also, the SP Computer Model can function as an alternative to reasoning in Bayesian Networks, as, for example, in modelling the phenomenon of "Explaining Away", described by Pearl in [17, pp. 7–9]. How that can be done is described in [24, Section 10.2] and [22, Section 7.8].

Because of the probabilistic nature of the SP System [32, Appendix A.5], all the kinds of reasoning that may be modelled in the SP System are fundamentally probabilistic, although it is possible to simulate the all-or-nothing nature of much classical logic when probabilities are at or near 0 or 1, as noted with references in Section 2.

As with the processing of natural language (Section 10), a major strength of the SP System is that there is smooth integration of diverse aspects of intelligence. This means that there can be smooth integration of one or more probabilistic kinds of reasoning with any or all of the SP System's strengths in several other aspects of AI and the representation of several kinds of knowledge.

12 Eliminating the problem of catastrophic forgetting

This and the next section describe problems in AI that are not apparently considered in [7] but are problems in AI research which the SP System may help to solve.

Catastrophic forgetting—which is a problem for at least some DNNs—is the way in which, when a given DNN has learned one thing and then it learns something else, the new learning wipes out the earlier learning (see, for example, [8]). This problem, mentioned at the end of the section about transfer learning (Section 7), is quite different from human learning, where new learning normally builds on earlier learning, as described in Section 7, although of course we all have a tendency to forget things.

The SP System is entirely free of the problem of catastrophic forgetting.

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

- In DNNs there is a single structure for the learning and storage of new knowledge, a concept like 'my house' is encoded in the strengths of connections between artificial neurons in that single structure, so that the later learning of a concept like 'my car' is likely to disturb the strengths of connections for 'my house' (see the discussion of 'grandmother' cells in [27, Section 5.8]);

  An apparent way round this problem is to provide an extremely large network for training, with multiple layers in all parts of the network, so that very large patterns such as bodies of text may be used in training. This seems to be employed in OpenAI's 'GPT-3' system [3]. There are implications here for the computational resources that are required (Section 8).

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

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

13 The need to develop 'broad AI' ('artificial general intelligence')

Last in this paper, but by no means least, is a problem described by Gary Marcus and Ernest Davis in their book *Rebooting AI* [13]. Writing about what they see as the shortcomings of most AI systems today, they say:

> "The central problem, in a word: current AI is *narrow*; it works for particular tasks that it is programmed for, provided that what it encounters isn't too different from what it has experienced before. That's fine for a board game like Go—the rules haven't changed in 2,500 years—but less promising in most real-world situations. Taking AI to the next level will require us to invent machines with substantially more flexibility. ... To be sure, ... narrow AI is certainly getting better by leaps and bounds, and undoubtedly there will be more breakthroughs in the years to come. But it's also telling: AI could and should be about so much more than getting your digital assistant to book a restaurant reservation." ([13, pp. 12–14], emphasis in the original).

Writing about possible ways forward, Marcus and Davis say:

> "What's missing from AI today—and likely to stay missing, until and unless the field takes a fresh approach—is *broad* (or 'general') intelligence. AI needs to be able to deal not only with specific situations for which there is an enormous amount of cheaply obtained relevant data, but also problems that are novel, and variations that have not been seen before.
>
> "Broad intelligence, where progress has been much slower, is about being able to adapt flexibly to a world that is fundamentally open-ended—which is the one thing humans have, in spades, that machines haven't yet touched. But that's where the field needs to go, if we are to take AI to the next level." ([13, p. 15], emphasis in the original).

And:

> "We call this book *Rebooting AI* because we believe that the current approach isn't on a path to get us to AI that is safe, smart, or reliable. A short-term obsession with narrow AI and the easily achievable 'low-hanging fruit' of big data has distracted too much attention away from a longer-term and much more challenging problem that AI needs to solve if it is to progress: the problem of how to endow machines with a deeper understanding of the world. Without that deeper understanding, we will never get to truly trustworthy AI. In the technical lingo, we may be stuck at a local maximum, an approach that is better than anything similar that's been tried, but nowhere good enough to get us where we want to go.
>
> "For now, there is an enormous gap—we call it 'the AI Chasm'—between ambition and reality." ([13, pp. 17–18]).

Of course, the SP System, like other AI systems, does not deliver 'broad AI' (which we'll assume means the same as AGI) as described by Marcus and Davis. But, for reasons summarised in the following subsections, it may with some justice claim to be the kind of "fresh approach", with the generality that they call for.

13.1 A fresh approach: a top-down strategy for the development of the SP System

As noted in [32, Appendix A.1], "The overarching goal of the SP programme of research is to simplify and integrate observations and concepts across AI, mainstream computing, mathematics, and HLPC."

The advantages of adopting a high-level goal, meaning essentially the same as a top-down strategy, are described in [32, Section 2], and more specifically in [32, Section 2.1]. They include: the potential to benefit in terms of broad scope and Ockam's razor (simplicity-with-power, next subsection); enlarging a problem can make it easier to solve; micro-theories rarely generalise well; and bottom-up strategies often lead to the fragmentation of research.

13.2 A fresh approach: aiming for a theory that combines conceptual ***Simplicity*** with descriptive and explanatory ***Power***

It seems that the simplicity-with-power objective mentioned in the previous subsection is largely met by the SP System, as described in the following subsections.

13.2.1 Simplicity

fresh-approach-simplicity-section

The SP Computer Model comprises mainly the processes for SPMA ([32, Appendix A.4]), with processes for unsupervised learning [32, Appendix A.5]. Overall, the model is remarkably simple, considering its versatility (next).

13.2.2 Power

fresh-approach-power-section

The descriptive and explanatory Power of the SP System lies in its versatility with aspects of intelligence [32, Appendix A.10.1], including diverse forms of reasoning [32, Appendix A.10.2], its versatility in the representation of diverse kinds of knowledge [32, Appendix A.10.3], and its facility for the seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination [32, Appendix A.10.4].

13.2.3 Related research

fresh-approach-related-research-section

It is true that there has been research for many years inspired by Allen Newell's book about *Unified Theories of Cognition* [14], and there is also a significant strand of research aiming to develop 'Artificial General Intelligence' (see, for example, [10]).

But despite the welcome aims of researchers in these areas, it appears fair to say that broad AI or AGI has not yet been achieved. Evidence to that effect includes what Marcus and Davis say in their book [13], and in the preface to the proceedings of a recent conference on AGI research:

> "Despite all the current enthusiasm in AI, the technologies involved still represent no more than advanced versions of classic statistics and machine learning. Behind the scenes, however, many breakthroughs are happening on multiple fronts: in unsupervised language and grammar learning, deep-learning, generative adversarial methods, vision systems, reinforcement learning, transfer learning, probabilistic programming, blockchain integration, causal networks, and many more." [10, Preface, Location 51].

Since, as noted above, the SP System has already largely met the objective of combining Simplicity with Power (Section 13.2), it seems fair to say that the SP System may claim to be part of a fresh approach to the development of broad AI.

13.3 A fresh approach: information compression as a unifying principle in the SP System

Another reason to be optimistic about the potential of the SP System to provide a relatively firm foundation for the development of broad AI is: 1) that all processing in the SP System is achieved via IC; 2) that IC lies at the heart of how the system represents its knowledge; 3) that IC is an extremely general concept that can in principle be applied to any kind of processing or the representation of any kind of knowledge; and 4) that evidence summarised in [32, Appendix A.10] shows the wide range of kinds of processing and kinds knowledge where IC may be applied to good effect.

To be more specific, IC in the SP System is achieved largely via SPMA ([32, Appendix A.4], and via unsupervised learning [32, Appendix A.5]). These two processes exemplify the generality of IC in the workings of the SP System.

13.4 A fresh approach: generalisation and the correction of over- and under-generalisations

A potentially valuable bonus from a theory of HLPC and AI which is founded on IC, for both the representation and the processing of knowledge, is that it provides a robust, coherent theory of generalisation and the correction of both over- and under-generalisations, which has some empirical support (Section 5).

As described in [32, Section 3], such a theory of generalisation may help to overcome problems of adaptability in driverless cars.

The variety of situations that a driver may encounter is far too large for any driver (human or computer) to achieve competence by learning specific responses to specific situations. In much the same way that we learn our native language or languages, learner drivers must generalise beyond their experience to date, but they must not overgeneralise. And it seems that IC provides the key.

13.5 A fresh approach: potential benefits and applications of the SP System

Another means of assessing the generality of the SP System is via potential benefits and applications of the system which are credible in the sense that they are not mere speculations but have supporting evidence via the workings and performance of the system.

In this respect, the SP System scores well, as can be seen from the range of potential benefits and applications described in [26] and in several other papers that may be downloaded via links from cognitionresearch.org/sp.htm, including peer-reviewed published papers on the application of the SP System to problems in: big data, the development of human-like intelligence in autonomous robots, computer vision and natural vision, intelligent databases, medical diagnosis, and mathematics.

14 Conclusion

This paper describes several problems in AI research and how the *SP System* (Section 1.2) may help to solve them.

This paper with two others—[30] and [31]—demonstrate some of the potential of information compression (IC) as a unifying principle in different aspects of intelligence and the representation and processing of knowledge. More generally, they may be seen to demonstrate, via three different sources of evidence, the potential of IC as a means of making sense of complexity in some of its manifestations.

Most of the problems in AI research considered in this paper are described by leading researchers in AI, in interviews with science writer Martin Ford, and reported by him in his book *Architects of Intelligence* [7].

On the strength of evidence presented in this paper, there is clear potential for the SP System to help solve all the problems described in the paper. It appears that *the SP System provides a relatively promising foundation for the development of artificial general intelligence.*

The problems discussed in this paper are as follows, each one with a brief summary of how the SP System may help to solve it:

- *The need to bridge the divide between symbolic and non-symbolic kinds of knowledge representation and processing* (Section 2). The fact that people can deal effectively with symbolic systems such as natural languages and mathematics, and they can also do what appear to be non-symbolic things like recognising objects, playing tennis, and so on, points to the need for a system that can bridge those two styles of knowledge representation and processing. The SP System has clear potential in that area.

- *Errors that can be made by deep neural networks* (Section 3). The SP System appears to be entirely free from the tendency of DNNs to make large and unexpected errors in recognition.

- *The challenges of unsupervised learning* (Section 4). Unlike most DNNs, learning in the SP System is entirely unsupervised. Although there are shortcomings as the system is now, it appears that the problems are soluble. This form of learning is prominent in the way people learn, and it may provide a foundation for other kinds of learning such as learning by being told, learning by imitation, learning via rewards and punishments (reinforcement learning), and so on.

- *How to account for generalisation, under-generalisation, and over-generalisation* (Section 5). The central role of IC in the SP system provides the basis for generalisation in accordance with coherent principles.

  In brief, the theory of generalisation embodied in the SP System may be expressed quite simply: for a given body of information, **I**, compress **I** as much as possible, separate 'encoding' from 'grammar' and discard the encoding. Section 5.1 describes how this works with the correction of over-generalisations, under-generalisations, and 'dirty data', and Section 5.2 describes how the principle would work with perception or related processes such as parsing.

Informal tests with the SP Computer Model and earlier IC-based models of learning, suggest that these principles are sound.

- *How to achieve unsupervised learning of useful knowledge from a single exposure or experience* (Section 6). Like people, and unlike DNNs, the SP System can learn useful knowledge from a single occurrence or experience. This is because: 1) the first stage in learning in the SP System is to take in information from the system's 'environment' and to interpret it in terms of existing knowledge; and 2) without any further processing, that new knowledge can serve immediately in other AI functions such as reasoning, problem-solving, and so on. This contrasts with DNNs where much repetition is required before any new knowledge is usable.

- *How to achieve transfer learning* (Section 7). 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. As noted in Section 7, this capability is related to but distinct from 'catastrophic forgetting'.

- *How, in learning, to increase speeds, and to reduce large demands for volumes of data, and computational resources* (Section 8). Like people, and unlike DNNs, the SP System can demonstrate useful learning fast with relatively small demands for data and computational resources. This probably mainly because of the SP System's strengths in one-shot learning and transfer learning. With the learning of complex knowledge or skills such as a new natural language, the skills needed for gymnastics to a high level, and so on, the SP System, like people, is likely to need more time, because of the volume of knowledge and the complexity of the task of organising it efficiently.

- *How to achieve transparency in the representation and processing of knowledge* (Section 9). Unlike DNNs, the SP System is entirely transparent in how it represents knowledge, and it provides a full and comprehensible audit trail for all its processing.

- *The need to strengthen the representation and processing of natural languages* (Section 10). Unlike DNNs, the SP System has strengths in the representation and processing of natural language with the kinds of syntactic structures that are recognised by linguists, with potential for the representation of semantic structures and their integration with syntax. It appears that such structures have psychological validity, and will probably be necessary for the achievement of human-like capabilities with natural languages.

- *How to create human-like capabilities in probabilistic reasoning* (Section 11). As a by-product of its design, the SP System exhibits several forms of probabilistic reasoning, such as abductive reasoning, reasoning with discrimination networks and trees, nonmonotonic reasoning and more. It seems likely that capabilities like these will be needed in any system that aspires to AGI. As with the processing of natural language, there is clear potential for smooth integration of one aspect of AI—the SP System's strengths in probabilistic reasoning—with any or all of the SP System's other strengths in AI and the representation of knowledge.

- *How to eliminate the problem of catastrophic forgetting* (Section 12). The problem of catastrophic forgetting, where new knowledge wipes out old knowledge, is a significant problem with most DNNs. By contrast, the SP System is entirely free of the problem.

- *The need to develop broad AI* (Section 13). Owing to the overarching strategy adopted in the SP programme of research—seeking to simplify and integrate across a broad convass—and owing to the SP System's favourable combination of conceptual Simplicity with descriptive and explanatory Power, it appears that *the SP System provides a relatively promising foundation for the development of artificial general intelligence.*

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Funding of research and publication

The author declares that he pays for the research and for its publication.

Data availability

The author declares that the research is concerned exclusively with the development of theory, and there are no data of the kind that would provide the basis of an empirical study.

Source code for the SP Computer Model, and Windows executable code, may be downloaded via links under the heading 'SOURCE CODE' on:

www.cognitionresearch.org/sp.htm#SOURCE-CODE.

Acknowledgements

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

References

[1] C. Brown. *My Left Foot.* Vintage Digital, London, Kindle edition, 2014. First published in 1954.

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

[3] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, Ramesh A, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners. 2020. arXiv:2005.14165v3 [cs.CL].

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

[5] J. W. Dauben. *Georg Cantor: His Mathematics and Philosophy of the Infinite.* Princeton University Press, Princeton, New Ed edition, 1990.

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

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

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

[9] D. O. Hebb. *The Organization of Behaviour.* John Wiley & Sons, New York, Kindle edition, 1949.

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

[11] F. Jabr. Does thinking really hard burn more calories? *Scientific American*, 2012, 2012.

[12] E. H. Lenneberg. Understanding language without the ability to speak: a case report. *Journal of Abnormal and Social Psychology*, 65:419–425, 1962.

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

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

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

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

[17] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Francisco, revised second printing edition, 1997.

[18] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and INSTITUTION = OpenAI NOTE = tinyurl.com/tna3aud YEAR = 2020 I. Sutskever, TITLE = Language models are unsupervised multitask learners. Technical report.

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

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

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

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

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

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

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

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

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

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

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

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

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

[32] J. G. Wolff. More problems in AI research and how the SP System may help to solve them. Technical report, CognitionResearch.org, 2020. Download: tinyurl.com/y29xngn7.