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

J Gerard Wolff*,†

Abstract

This paper, a companion to [37], describes problems in AI research and how the *SP System* (described in sources detailed in the paper) may help to solve them. Most of these problems are described by leading researchers in AI in interviews with science writer Martin Ford, and reported by him in his book *Architects of Intelligence* [3]. Problems and their potential solutions that are described in this paper are: the need to rebalance research towards top-down strategies; how to minimise the risk of accidents with self-driving vehicles; the need for strong compositionality in the structure of knowledge; the challenges of commonsense reasoning and commonsense knowledge; establishing the key importance of information compression in AI research; establishing the importance of biological validity in AI research; whether knowledge in the brain is represented in 'distributed' or 'localist' form; the limited scope for adaptation of deep neural networks; and reasons are given for why the important subjects of motivations and emotions have not so far been considered. The evidence in this paper and [37] suggests that *the SP System provides a firmer foundation for the development of artificial general intelligence than any alternative.*

1 Introduction

This paper describes problems in research in artificial intelligence (AI) and how the *SP System* may to help solve them. It is a companion to the paper [37] which describes other problems in AI research, with potential solutions.

The *SP System*, meaning the *SP Theory of Intelligence* and its realisation in the *SP Computer Model*, is described in outline in Appendix A, in more detail in [23], and even more fully in [21].[1]

*CognitionResearch.org, UK

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

[1]There are also outline descriptions of the SP System in [35, Section 3] and [34, Section 2.2].

As with [37]:

- Most of the problems 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* [3].

- This paper is not a review of Ford's book. Rather, it is mainly an extended response to the very interesting material that Ford has gathered in his book. Each problem in AI research is described in its own section, with a description of how SP System may help solve it.

- Although the form of each paper is unusual, it performs an important task in any field of science: the evaluation of problems in the field, and how they may be solved.

On the strength of evidence presented in this paper, and in [37], there is clear potential for the SP System to help solve all the problems which are the subjects of the two papers. Since there is no other system with anything close to that potential, it is good evidence that *the SP System provides a firmer foundation for the development of artificial general intelligence (AGI) than any alternative.*

In most of this paper, and [37], deep neural networks (DNNs) are the implicit or explicit alternative to the SP System. This is because of their dominance in AI research today, which is itself due to their undoubted successes in several areas of application. Nevertheless, most of the problems in AI research that are the subject of this paper and of [37] arise from shortcomings of deep neural networks.

To forestall any misunderstanding, the SP System is *not* some version of a DNN. It is entirely different, both in its organisation and in its workings, as summarised in Appendix A.5.

It is assumed in this paper that what we know about human learning, perception, and cognition (HLPC), and neuroscience, can be helpful in the development of AI, and *vice versa*.

The aforementioned problems in AI research and how they may be solved are described in the main sections that follow.

1.1 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 intelligence* (Section 1): 'AI'.
- *Artificial General AI* (Section 1): 'AGI'.
- *Deep Neural Network* (Section 1): 'DNN'.
- *Human learning, perception, and cognition* (Section 1): 'HLPC'.
- *SP-multiple-alignment* (Section 2): 'SPMA'.
- *Commonsense reasoning and commonsense knowledge* (Section 5): 'CSRK'.
- *Information compression* (Section 6): 'IC'.
- *Information compression via the matching and unification of patterns* (Appendix A.3): 'ICMUP'.

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

2 The need to rebalance research towards top-down strategies

> "... one of [the] stepping stones [towards progress in AI] 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 [3, p. 502].

This quote is, in effect, a call for a top-down strategy in AI research, developing a theory or theories that can be applied to a range of phenomena, not just one or two things in a narrow area. The potential advantages of that kind of strategy in terms of the generality of theories, and their value in terms of Ockham's razor, are described in Section 2.1, below.

In this connection, the SP System scores well. Appendix A.1 describes how *the SP System has adopted a unique top-down strategy, attempting to simplify and integrate observations and concepts across AI, mainstream computing, mathematics, and HLPC.* And it seems fair to say that development of the SP System, with the powerful concept of SP-multiple-alignment (SPMA) at its core, has achieved a favourable combination of conceptual Simplicity with descriptive and explanatory Power (*ibid.*).

2.1 Some advantages of a top-down strategy

Key features of a top-down strategy in research, and its potential benefits, are mainly these:

1. *Broad scope and Ockham's razor.* To achieve generality, the data from which a theory is derived should have a broad reach, like the overarching goal of the SP programme of research, described in Appendix A.1, repeated at the beginning of this main section: to simplify and integrate observations and concepts across AI, mainstream computing, mathematics, and HLPC.

   That broad scope is important for two reasons:

   - In accordance with Ockham's razor, a theory should be as *Simple* as possible but, at the same time, it should retain as much as possible of the descriptive and explanatory *Power* of the data from which it is derived (Appendix A.1).
   - But those two measures are far more significant when they apply to a wide range of phenomena than when they apply only to a small piece of data.

2. *If you can't solve a problem, enlarge it.* A broad scope, as above, can be challenging, but it can also make things easier. Thus President Eisenhower is reputed to have said: "If you can't solve a problem, enlarge it", meaning that putting a problem in a broader context may make it easier to solve. Good solutions to a problem may be hard to see when the problem is viewed through a keyhole, but become visible when the door is opened.

3. *Micro-theories rarely generalise well.* Apart from the potential value of 'enlarging' a problem (point 2 above), and broad scope (point 1), a danger of adopting a narrow scope is that any micro-theory or theories that one may develop for that narrow area are unlikely to generalise well to a wider context—with correspondingly poor results in terms of Simplicity and Power.

   For reasons of that kind, Allen Newell, in his famous essay "You can't play 20 questions with nature and win" [13], urges researchers in psychology to develop theories with wide scope (pp. 284–289), and, accordingly, to work with "a genuine slab of human behaviour" (p. 303). This kind of thinking is the basis of his book *Unified Theories of Cognition* [14].

4. *Bottom-up strategies and the fragmentation of research.* The prevailing view about how to reach AGI seems to be "... that we'll get to general intelligence step by step by solving one problem at a time." expressed by Ray Kurzweil

[3, p. 234]. And much research in AI has been, and to a large extent still is, working within this kind of bottom-up strategy: developing ideas in one area, and then perhaps trying to generalise them to another area, and so on.

But it seems that in practice the research rarely gets beyond two areas, and, as a consequence, there is much fragmentation of research.

2.2 The fragmentation of research

In connection with the fragmentation of research (point 4, above), John Kelly and Steve Hamm (both of IBM) write:

> "Today, as scientists labor to create machine technologies to augment our senses, there's a strong tendency to view each sensory field in isolation as specialists focus only on a single sensory capability. Experts in each sense don't read journals devoted to the others senses, and they don't attend one another's conferences. Even within IBM, our specialists in different sensing technologies don't interact much." [6, location 1004].

And Pamela McCorduck writes:

> "The goals once articulated with debonair intellectual verve by AI pioneers appeared unreachable ... Subfields broke off—vision, robotics, natural language processing, machine learning, decision theory—to pursue singular goals in solitary splendor, without reference to other kinds of intelligent behaviour." [11, p. 417]. Later, she writes of "the rough shattering of AI into subfields ... and these with their own sub-subfields—that would hardly have anything to say to each other for years to come." [11, p. 424]. She adds: "Worse, for a variety of reasons, not all of them scientific, each subfield soon began settling for smaller, more modest, and measurable advances, while the grand vision held by AI's founding fathers, a general machine intelligence, seemed to contract into a negligible, probably impossible dream." (*ibid.*).

Honorable attempts to overcome these problems are two main strands of work: 1) research inspired by Newell's *Unified Theories of Cognition* [14]; and 2) research aiming to develop AGI (see, for example, [5]). But, while both strands of research are welcome, it seems that neither of them have yet managed to escape properly from the problems of bottom-up research.

That a top-down approach to the development of a fully-integrated AGI is proving difficult is suggested by the following observations 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." [5, Preface, Location 51].

In other words, attempts to develop unified theories of cognition, or AGI, have, so far, not overcome the fragmentation of AI so well described by Pamela McCorduck [11, p. 417], quoted above.

2.3 Publish or perish

It appears that the main reason for this tendency for researchers to work in small fields is because it is much easier to produce publishable results in a small field than it is with something more ambitious. And the motive for working in an area where it is easy to produce publishable results appears to be largely because of the relentless pressure to "publish or perish".

And it appears that that pressure to "publish or perish" is due largely to what is probably the false belief amongst research managers and politicians that high pressure is needed to ensure that researchers do something useful. In accordance with the Yerkes–Dodson laws,[2] too much pressure can be counter-productive. There is a need for a radical revision of the goals and incentives for researchers.

3 How to minimise the risk of accidents with self-driving vehicles

> "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

[2]The Yerkes–Dodson curves [38] relate 'arousal', on the x axis, to 'performance' on some task, on the y axis. If 'arousal' is seen to be the result of pressure to achieve results, performance on simple tasks increases up to a plateau with increased pressure, but performance on complex tasks improves up to a peak with increasing pressure but then declines back to 'weak' performance. Since academic or applied research is clearly a complex task, the lesson is clear: be careful not to apply too much pressure.

> have a rule for that circumstance. So, then they have to add a new one, and so on, and so on." Stuart J. Russell [3, 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 [3, p. 321].

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

It seems that, for any kind of driver, either human or artificial, some kind of generalisation from experience is essential [37, Section 5]. In that connection, people will have the benefit of all their visual experience prior to their driving lessons, but the same principles apply.

If a person or a driverless car has learned to apply the brakes when a child runs out in front, that learning should be indifferent to the multitude of images that may be seen: the child may be fat or thin; tall or short; running, skipping, or jumping; in a skirt or wearing trousers; facing towards the car or away from it; seen through rain or not; lit by street lights or by the sun; and so on.

There may be some assistance from 'generalisation via perception' [37, Section 5.2] but that in itself is unlikely to be sufficient. It seems that something like 'generalisation via unsupervised learning' [37, Section 5.1] is also needed.

With those two kinds of generalisation, it seems possible that, with reasonable amounts of driving experience across a range of driving conditions, the risk of accidents may be minimised.

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

4 The need for strong compositionality in the structure of knowledge

> "By the end of the '90s and through the early 2000s, neural networks were not trendy, and very few groups were involved with them. I had a strong intuition that by throwing out neural networks, we were throwing out something really important.
>
> "Part of that was because of something that we now call compositionality: The ability of these systems to represent very rich information about the data in a compositional way, where you compose many building blocks that correspond to the neurons and the layers." Yoshua Bengio [3, p. 25].

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

It appears that in this connection, the SP System has a striking advantage compared with DNNs. Any SP-pattern may contain SP-symbols that serve as references to other SP-patterns, a mechanism which allows hierarchical structures to be built up through as many levels as are required ((([23, Section 9,1], [21, Section 6.4])).

This can be seen in Figure 3, where the SP-pattern '`D Dp 4 t w o #D`' (which represents a word), and this connects with the SP-pattern '`NP NPp D Dp #D N Np #N #NP`' which which represents a noun phrase, and that connects with the SP-pattern '`VP VPp Vr #Vr #VP`' which represents a verb phrase, and that connects with the SP-pattern '`S Num ; NP #NP VP #VP #S`' which represents a sentence.

Apart from part-whole hierarchies, the SP System also lends itself to the representation and processing of class-inclusion hierarchies, as can be seen in [23, Figure 16, Section 9.1].

5 The challenges of commonsense reasoning and commonsense knowledge

> "We don't know how to build machines that have human-like 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 [3, 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 [3, p. 472]. Here, 'AGI' is an abbreviation for 'Artificial General Intelligence'.

Although 'commonsense reasoning' (CSR) is a kind of reasoning, it is discussed here, with 'commonsense knowledge' (CSK), in a section that is separate from [37, Section 11] (about the strengths and potential of the SP Computer Model for several kinds of probabilistic reasoning) 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, [2]).

Judging by the nature of DNNs and the paucity of research on how they might be applied in the CSRK area [17], it seems that DNNs are not well suited to this aspect of AI. But preliminary studies suggest that the SP System has potential in this area.

- It may prove useful with CSRK [33, Section 3], and more so when 'unfinished business' in the development of the SP Computer Model has been completed (Appendix A.13).

- Aspects of CSRK may be modelled with the SP Computer Model [33, Sections 4 to 6]: how to interpret a noun phrase like "water bird"; how, under various scenarios, to assess the strength of evidence that a given person committed a murder; how to interpret the horse's head scene in *The Godfather* film.

  A fourth problem—how to model the process of cracking an egg into a bowl—is beyond what can be done with the SP System as it is now [33, 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 [32], 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" [7].

6 Establishing the key importance of information compression in AI research

There is little about information compression in *Architects of Intelligence*, except for some brief remarks about autoencoders by Yoshua Bengio in [3, p. 26]:

> "Autoencoders have changed quite a bit since [the] 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.

Now it seems that interest in autoencoders has waned: "... a few years later, we discovered that we didn't need these approaches to train deep networks, we could just change the nonlinearity." Yoshua Bengio [3, p. 26].

This fairly relaxed view of IC in AI research, as described by a leading researcher in AI, contrasts with the view in the SP programme of research that IC is fundamental in HLPC [34], in mathematics [35], and in the design of the SP System [23, 21].

It seems that there is a need for communication and discussion about this important issue.

7 Establishing the importance of a biological perspective in AI research

With regard to the importance of biology in AI research:

> "Deep learning will do some things, but biological systems rely on hundreds of algorithms, not just one algorithm. [AI researchers] will need hundreds more algorithms before we can make that progress, and we cannot predict when they will pop." Rodney Brooks [3, p. 427].

What Rodney Brooks describes here is much like Marvin Minsky's concept of diverse agents [12] as the basis for AI. It seems that both of them are unfalsifiable theories because, for every attempt to prove the theory wrong, a new algorithm may be added to plug the gap. And that is likely to mean a theory with ever-decreasing merit in terms of Ockham's razor.

Here is some evidence relating to the importance of biology in AI research:

- The paper [34, Section 4] describes powerful reasons in terms of natural selection why we should expect to find IC in brains and nervous systems of all animals including people.

- The rest of the paper [34] describes quite a lot of other evidence from people and other animals for the importance of IC in natural intelligence and other aspects of neural functioning.

- In his book *Kluge* [9], Marcus provides compelling arguments, with evidence, why the haphazard nature of natural selection should have produced kluges, meaning clumsy makeshift solutions that nevertheless work.

In this paper, and other writings in the SP programme of research, the main emphasis has been on the importance of IC in HLPC. But this is mainly for the sake of clarity, and because little attention has so far been given to the possibility of kluges from the haphazard nature of natural selection.

At some stage in the SP programme of research, the latter influence should receive attention.

8 Whether knowledge in the brain is represented in 'distributed' or 'localist' form

> "In a hologram, information about the scene is distributed across the whole hologram, which is very different from what we're used to. It's very different from a photograph, where if you cut out a piece of a photograph you lose the information about what was in that piece of the photograph, it doesn't just make the whole photograph go fuzzier." Geoffrey Hinton [3, p. 79].

A persistent issue in AI and theories of HLPC is whether knowledge in the brain is represented in a 'distributed' or 'localist' form.

In DNNs, knowledge is distributed in the sense that it is encoded in the strengths of connections between many neurons across several layers of the network. Since DNNs provide the most fully developed examples of distributed knowledge, it is assumed in present discussion that they are representative for that kind of system for representing knowledge.

In SP-Neural, a 'neural' version of the SP System (Appendix A.9), knowledge is localised in the sense that there may be a single neuron, or more likely a small cluster of neurons, representing any one concept such as 'my house', but such a concept is likely to have links to many other concepts in other places such as 'roof', 'window', 'doors', and so on.

This issue is essentially the much-debated issue of whether the concept of 'my grandmother' is represented in one place or whether the concept may be represented via a diffuse collection of neurons throughout the brain.

Although the following conclusion will not be free from controversy, it seems that the weight of evidence now favours a localist view, and the SP System is in keeping with that conclusion:

- The SP System, in both its abstract form (Appendix A) and as SP-Neural (Appendix A.9), is unambiguously localist. To the extent that the SP System provides a plausible framework for the development of AGI, it provides evidence in support of localist forms for knowledge.

- Since DNNs are vulnerable to the problem of catastrophic forgetting [37, Section 12], this seems to be a problem more generally for the distributed representation of knowledge. With knowledge stored in the strengths of connections between neurons, there seems, at the least, to be a risk that concepts will interfere with each other.

- It is true that if knowledge of one's grandmother is contained within a single neuron, death of that neuron would destroy one's knowledge of one's grandmother. But:

  - As Barlow points out [1, pp. 389–390], a small amount of replication will give considerable protection against this kind of catastrophe.

  - Any person who has suffered a stroke, or is suffering from dementia, may indeed lose the ability to recognise close relatives.

- Is it conceivable that, with a localist representation, there are enough neurons in the human brain to store the knowledge that a typical person, or, more to the point, an exceptionally knowledgeable person, may have? Arguments and calculations relating to this issue suggest that it is indeed possible for us to store what we know in localist form, and with substantial room to spare for multiple copies [21, Section 11.4.9]. A summary of the arguments and calculations may be found in [29, Section 4.4].

- In support of the conclusion in the preceding point, the key importance of IC in the organisation and workings of the SP System (Appendix A.8) provides a good reason for supposing that, in a mature SP Machine [16], knowledge acquired via unsupervised learning will be stored in a highly compressed form.

- Mike Page [15, pp. 461–463] discusses several studies that provide direct or indirect evidence in support of localist encoding of knowledge in the brain.

9 The limited scope for adaptation in deep neural networks

> "What's missing from AI today—and likely to stay missing, until and unless the field takes a fresh approach—is broad (or 'general') intelligence. AI needs to be able to deal not only with specific situations for which there is an enormous amount of cheaply obtained relevant data, but also problems that are novel, and variations that have not been seen before.
>
> "Broad intelligence, where progress has been much slower, is about being able to adapt flexibly to a world that is fundamentally open-ended—which is the one thing humans have, in spades, that machines haven't yet touched. But that's where the field needs to go, if we are to take AI to the next level." Gary Marcus and Ernest Davis [10, p. 15].

A problem which is closely related to catastrophic forgetting [37, Section 12] and also to the distinction between 'broad' and 'narrow' AI [37, Section 13] is that any one DNN is really only designed to learn a single concept. It is true that one could provide multiple DNNs for the learning of multiple concepts but, since a DNN has multiple layers and multiple connections between layers (which is what makes it 'deep'), the provision of a DNN for each of the many concepts that people can learn would be expensive.

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

By contrast, the layers in a DNN, and the potential connections amongst them, are finite and pre-defined [37, Section 12]. It is true that the connections can vary in strength but only within pre-defined limits.

10 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 [3, 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 [3, 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 [3, 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 [3, 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 [3, p. 457].

In developing AGI, motivations and emotions are clearly important, not least because of the possibility that super-intelligent AIs might come to regard people as dispensable. But, in the SP programme of research, there has, so far, been no attempt to give the SP Computer Model any kind of motivation (except the motivation 'compress information'), or any kind of emotion. This is because of the belief that, in relation to the SP concepts and their development, it would be trying to run before we can walk. When the SP System is more mature, there will likely be a case for exploring how it may be applied to the complexities of motivations and emotions.

11 Conclusion

This paper, with its sister paper [37], describes problems in AI research and how the *SP System* may help solve them.

Here, *SP System* means the *SP Theory of Intelligence* and its realisation in the *SP Computer Model*. There is an outline description of the SP System in Appendix A, with pointers to where fuller information may be found.

Many of the problems in AI research considered in this paper and [37] are described by leading researchers in AI in interviews with science writer Martin Ford, and reported by him in his book *Architects of Intelligence* [3].

On the strength of evidence presented in this paper, and in [37], there is clear potential for the SP System to help solve all the problems which are the subjects of the two papers. Since there is no other system with anything close to that potential, it is good evidence that *the SP System provides a firmer foundation for the development of AGI than any alternative.*

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 rebalance research towards top-down strategies* (Section 2). Unlike most research in AI today, the SP System has been developed via a top-down strategy, aiming for simplification and integration across AI, mainstream computing, mathematics, and HLPC.

- *Minimising the risk of accidents with driverless cars* (Section 3). The complexities of driving a car seem to require human-like abilities in generalising its knowledge [37], with the correction of over- and under-generalisations. This may be done both via unsupervised learning ([37, Section 5.1], Appendix A.6), and also via processes of perception [37, Section 5.2].

- *The need for strong compositionality* (Section 4). Although DNNs appear to provide some approximations of the idea of one concept being composed of smaller elements, this idea is much more strongly developed in the SP System where it is entirely feasible to create class-inclusion hierarchies and part-whole hierarchies of any depth, and without the vagueness and ambiguity of DNN approximations of compositionality.

- *The challenges of commonsense reasoning and commonsense knowledge* (Section 5). 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 are described.

- *Establishing the key importance of IC in AI research* (Section 6). The SP System appears to be unique in employing IC as the basis for all aspects of intelligence and the representation of knowledge. There is much evidence for the importance of IC in HLPC [34].

- *Biological validity* (Section 7). 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.

- *Is knowledge in the brain represented in 'distributed' or 'localist' form?* (Section 8). In contrast to DNNs, the SP System stores knowledge in an unambiguously 'localist' form, meaning that each concept is represented by a small cluster of neurons in one location. In this respect it conforms to what appears to be the balance of evidence in favour of localist kinds of representation.

- *Scope for adaptation* (Section 9). The representation of knowledge with SP-patterns, with the SPMA construct, provides for much greater scope for adaptation than the layers of a DNN.

- *Motivations and emotions*. Despite the importance for people of motivations and emotions (Section 10), no attempt has yet been made to study them in the SP programme of research.

Other problems in AI research and how the SP System may help solve them are described in [37].

Acknowledgements

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

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

The *SP System*—meaning the *SP Theory of Intelligence* and its realisation in the *SP Computer Model* is the product of a lengthy programme of research, from about 1987 to now with a break between early 2006 and late 2012. This programme of research has included the creation and testing of many versions of the SP Computer Model. A major discovery has been the concept of *SP-multiple-alignment* and its versatility in many aspects of intelligence (Appendix A.4).

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

The overarching goal of the SP programme of research is to simplify and integrate observations and concepts across AI, mainstream computing, mathematics, and HLPC. In effect, this means developing concepts that combine conceptual *Simplicity* with high levels of descriptive or explanatory *Power*. This in turn means the same as IC by increasing the *simplicity* of a body of information **I**, by the removal of *redundancy* from **I**, whilst retaining as much as possible of its non-redundant, expressive *power*.

As readers may guess, the Simplicity and Power concepts, which apply to both the aims of the research and the workings of the SP Computer Model, are the reasons for the name 'SP'.

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 SPMA (Appendix A.4), with relatively simple processes for unsupervised learning (Appendix A.6), has strengths and potential across diverse aspects of intelligence and the representation of diverse kinds of knowledge (Appendix A.11).

A.2 High level view of the SP System

The SP System is described in outline here, in more detail in [23], and much more fully in [21]. Distinctive features and advantages of the SP System are described in [30]. Other papers in this programme of research are detailed, with download links, on www.cognitionresearch.org/sp.htm.

In broad terms, the SP System is a brain-like system that takes in *New* information through its senses and stores some or all of it as *Old* information that is compressed, as shown schematically in Figure 1.

In the SP System, all kinds of knowledge are represented with *SP-patterns*, where each such SP-pattern is an array of atomic *SP-symbols* in one or two di-

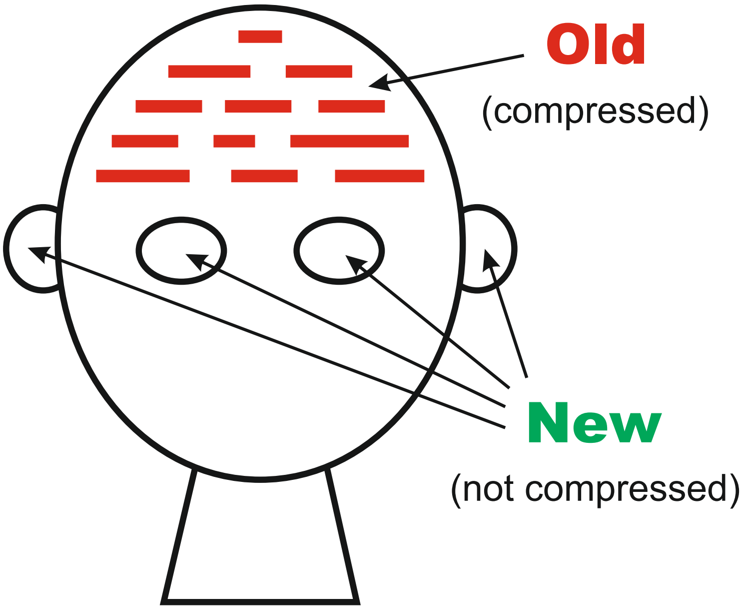

Figure 1: Schematic representation of the SP System from an 'input' perspective. Reproduced, with permission, from Figure 1 in [23].

mensions. An SP-symbol is simply a 'mark' that can be matched with any other symbol to determine whether it is the same or different.

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.3 Information compression via the matching and unification of patterns

A central idea in the SP System is that all kinds of processing would be achieved via IC. Evidence for the importance of IC in HLPC is described in [34].

In the development of the SP System, it has proved useful to understand IC as a process of searching for patterns that match each other and the merging or 'unifying' patterns that are the same. The expression 'information compression via the matching and unification of patterns' may be abbreviated as 'ICMUP'.

More specifically IC in the SP System is achieved largely via the creation of SPMAs (Appendix A.4) and, via unsupervised learning (with SPMAs playing an important role), the creation of SP-grammars (Appendix A.6).

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, [8]), 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 itself a theory of computing [30, Section II-C] founded on concepts of ICMUP and SPMA.

A.4 SP-multiple-alignment

A central idea in the SP System, is the simple but powerful concept of SPMA, borrowed and adapted from the concept of 'multiple sequence alignment' in bioinformatics.

SPMA is the last of seven variants of ICMUP described in [35, Section 5]. It may be seen to be a generalisation of the other six variants [35, Section 5.7].

Within the SP System, the SPMA concept is largely responsible for the strengths and potential of the SP System as outlined in Appendix A.11. The versatility of the SP System may also be seen in its several potential areas of application described in [27] and several other papers that are detailed with download links on www.cognitionresearch.org/sp.htm.

Bearing in mind that 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.*

Probably the best way to explain the idea is by way of examples. Figure 2 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 A G G A     T G     G   G G A
  | | |     |     | | | | | | | | |     | |     |   | | |
  G G | G   G C C C A G G G A G G A     | G G C G   G G A
  | | |     | | | | | | | | | | | |     | |     |   | | |
A | G A C T G C C C A G G G | G G | G C T G     G A | G A
  | | |           | | | | | | | | |   |   |     |   | | |
  G G A A         | A G G G A G G A   | A G     G   G G A
  | |   |         | | | | | | | |     |   |     |   | | |
  G G C A         C A G G G A G G     C   G     G   G G A
```

Figure 2: A 'good' multiple sequence 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'[3]—which produce results that are good enough for practical purposes.

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

Figure 3 shows an example of an SPMA, superficially similar to the one in Figure 2, except that the sequences are called *SP-patterns*, the SP-pattern in row 0 is New information and the remaining SP-patterns, one per row, are Old SP-pattern, selected from a relatively large pool of such SP-patterns. A 'good' SPMA is one which allows the New SP-pattern to be encoded economically in terms of the Old SP-patterns.

```
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 NPp D Dp         #D N Np                        #N #NP             | | | |            4
             |   |                                                  |              | | | |
5            |   |                                                  |         Vr 1 p l a y #Vr        5
             |   |                                                  |         |            |
6            |   |                                                  |  VP VPp Vr           #Vr #VP    6
             |   |                                                  |  |   |                    |
7 S Num    ; NP  |                                                 #NP VP  |                   #VP #S 7
     |     |     |                                                         |
8   Num PL ;    NPp                                                       VPp                         8
```

Figure 3: The best SPMA created by the SP Computer Model with a store of Old SP-patterns like those in rows 1 to 8 (representing grammatical structures, including words) and a New SP-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). Adapted from Figure 1 in [22], with permission.

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

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 the SPMA in this example is the parsing a sentence ('`f o r t u n e f a v o u r s t h e b r a v e`') into its grammatical parts and sub-parts.

Contrary to the impression that may be given by Figure 3, the SPMA concept is very versatile and is largely responsible for the strengths and potential of the SP System, as described in Appendix A.11.

As with multiple sequence alignments, it is almost always necessary to use heuristic methods to achieve useful results without undue computational demands. The use of heuristic methods helps to ensure that computational complexities in the SP System are within reasonable bounds [21, 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 The SP System is *quite different* from a deep neural network

The several levels in an SPMA may give the impression that the SP System in its structure and workings simply a variant of the structure and workings of a DNN. This is entirely false.

In DNNs, the layers are provided at the beginning of processing and do not change except in the strengthening of links between neurons. By contrast, the SP System stores its knowledge in the form of SP-patterns, and those SP-patterns become the rows in each of a multitude of different SPMAs, each of which contains its own distinctive array of SP-patterns, normally a unique set of SP-patterns but sometimes two or more sets the same but with different alignments. Also, IC is of central importance in the SP System by contrast with most DNNs in which IC has little or no role.

A.6 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.* [4]).

Unsupervised learning in the SP System is quite unlike 'Hebbian' learning via the gradual strengthening or weakening of neural connections ([37, Section 6]), variants of which are the mainstay of learning in DNNs. In the SP System, unsupervised learning incorporates the building of SPMAs 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. All learning in the SP System starts with the taking in of information from the environment:

- If that information is the same as one or more Old SP-patterns, then the frequency of the one SP-pattern is increase, or frequencies of the two or more SP-patterns are increased.

- If that information is entirely new, 'ID' SP-symbols[4] are added at the beginning and end of the pattern so that it becomes an SP-pattern. Then it is added directly to the store of Old SP-patterns.

- If partial matches can be made between the newly-received information and one or more of the stored Old SP-patterns, then each of the parts that match, and each of the parts that don't match, are made into SP-patterns by the addition of ID SP-symbols at the beginning and end, and the newly-created SP-patterns are added to the store of Old SP-patterns.

With a given body of New SP-patterns, the system processes them as just sketched, and then searches for one or two 'good' *SP-grammars*, where an SP-grammar is a collection of Old SP-patterns, and it is 'good' if it is effective in the economical encoding of the original set of New SP-patterns, where that economical encoding is achieved via SPMA.

As with the building of SPMAs, 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 SPMAs, the use of heuristic methods helps to ensure that computational complexities in the SP System are within reasonable bounds [21, 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 ([23, Section 5], [21, Chapter 9]).

But there are (at least) two shortcomings in the system [23, Section 3.3]: it cannot learn intermediate levels of structure or discontinuous dependencies in grammar, although the SPMA framework can accommodate structures of those kinds.

[4]These are SP-symbols for identification and classification such as '`Nr`', '`5`', and '`#Nr`', in the SP-pattern '`Nr 5 k i t t e n #Nr`' in Figure 3 (a).

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.7 The probabilistic nature of the SP System

Owing to the intimate relation that is known to exist between IC and concepts of probability [18, 19], and owing to the fundamental role of IC in the workings of the SP System, the system is inherently probabilistic ([23, Section 4.4], [21, Section 3.7]).

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.11.2).

The very close connection that exists between IC and concepts of probability may suggest that there is nothing to choose between them. But [35, Section 8.2] argues that, in research on aspects of AI and HLPC, there are reasons to regard IC as more fundamental than probability and a better starting point for theorising.

A.8 Two main mechanisms for information compression in the SP System, and their functions

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

1. *The building of SP-multiple-alignments.* The process of building SPMAs achieves compression of New information. At the same time it may achieve any or all of the following functions described in [21, Chapters 5 to 8] and [23, Sections 7 to 12], with potential for more:

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

(b) Pattern recognition which is robust in the face of errors of omission, commission, or substitution; and pattern recognition at multiple levels of abstraction.
(c) Information retrieval which is robust in the face of errors of omission, commission, or substitution.
(d) Several kinds of probabilistic reasoning, as summarised in Section A.11.2.
(e) Planning such as, for example, finding a flying route between London and Beijing.
(f) Problem solving such as solving the kinds of puzzle that are popular in IQ tests.

The building of SPMAs is also part of the process of unsupervised learning, next.

2. *Unsupervised learning.* Unsupervised learning, outlined in Appendix A.6, means the creation of one or two *SP-grammars* which are collections of SP-patterns which are effective in the economical encoding of a given set of New SP-patterns.

A.9 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 [29].

The concept of an SP-symbol may realised as a *neural symbol* comprising a single neuron or, more likely, a small cluster of neurons. An SP-pattern maps quite well on to the concept of a *pattern assembly* comprising a group of inter-connected SP-symbols. And an SPMA may be realised in terms of pattern assemblies and their interconnections, as illustrated in Figure 4.

In this connection, it is relevant to mention that the SP System, in both its abstract and neural forms, is quite different from DNNs [17] and has substantial advantages compared with such systems, as described in several sections in [37] and [36], and in [30, Section V].

A.10 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,

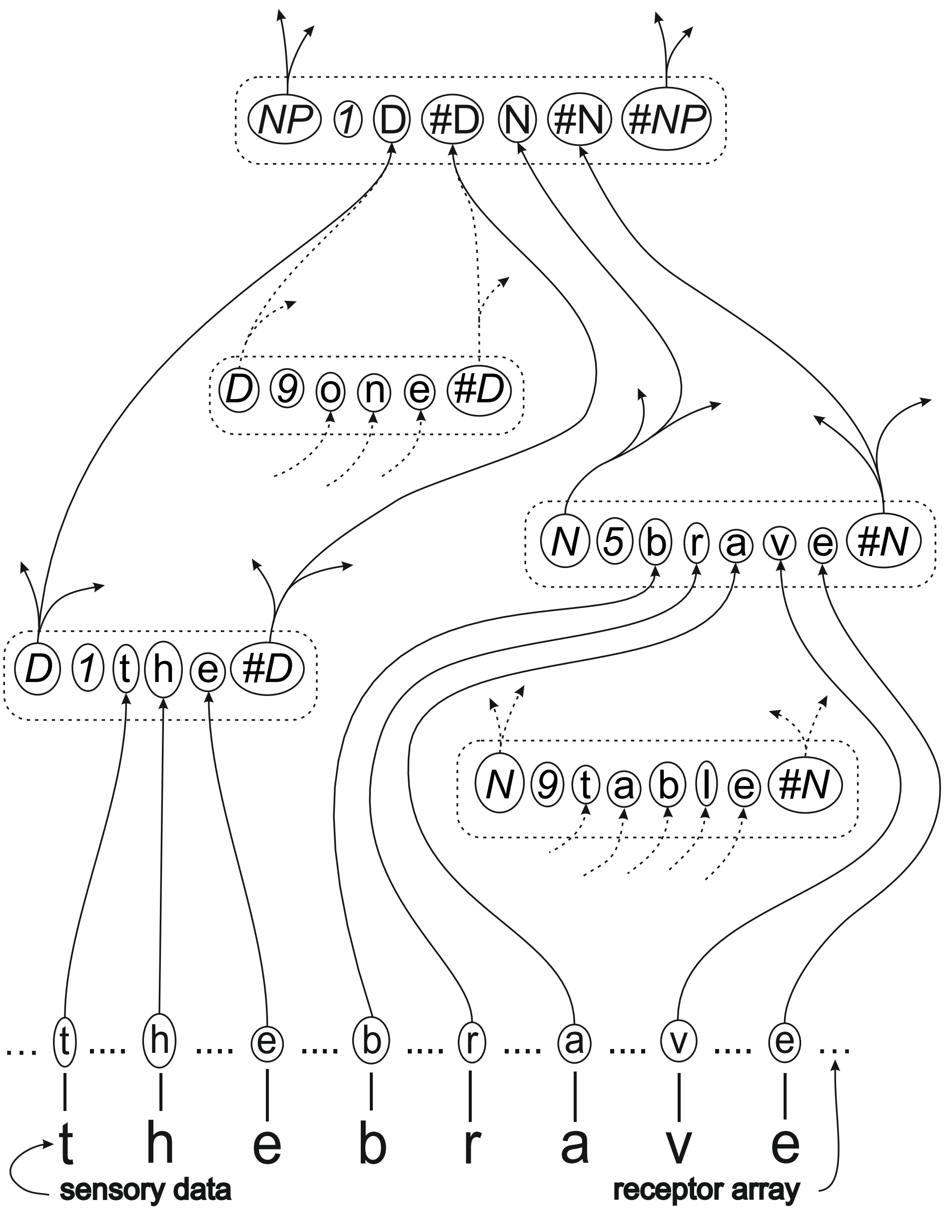

Figure 4: A schematic representation of a partial SPMA in SP-Neural, as discussed in [29, Section 4]. Each broken-line rectangle with rounded corners represents a *pattern assembly*—corresponding to an SP-pattern in the main SP Theory of Intelligence; each character or group of characters enclosed in 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.

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 [24, Section 6.1 and 6,2], and procedures that work in parallel as described in [25, Sections V-G, V-H, and V-I, and C].

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

A.11 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 [23, Sections 5 to 12], [21, Chapters 5 to 9], [30], and in other sources referenced in the subsections that follow.

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

A.11.1 Versatility in aspects of intelligence

The SP System has strengths and potential in the 'unsupervised' learning of new knowledge. As noted in Appendix A.6, 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 SPMAs.

Other aspects of intelligence where the SP System has strengths or potential are modelled via the building of SPMAs. 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 [24]; best-match and semantic kinds of information retrieval; several kinds of reasoning (next subsection); planning; and problem solving.

A.11.2 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 ([21, Section 3.7], [23, Section 4.4]).

There is also potential in the system for spatial reasoning [25, Section IV-F.1], and for what-if reasoning [25, Section IV-F.2].

It seems unlikely that the features of intelligence mentioned above are the full extent of the SP System's potential to imitate what people can do. The close connection that is known to exist between IC and concepts of probability (Appendix A.7), the central role of IC in the SPMA framework, and the versatility of the SPMA framework in aspects of intelligence suggest that there are more insights to come.

As noted in Appendix A.7, 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.11.3 Versatility in the representation of knowledge

Although SP-patterns are not very expressive in themselves, they come to life in the SPMA 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 [22, Sections 3 and 4]; relational tuples (*ibid.*, Section 3), and concepts in mathematics, logic, and computing, such as 'function', 'variable', 'value', 'set', and 'type definition' ([21, Chapter 10], [27, Section 6.6.1], [31, 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 SPMA 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.11.4 The seamless integration of diverse aspects of intelligence, and diverse kinds of knowledge, in any combination

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

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

Figure 5 shows schematically how the SP System, with SPMA 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.

A.12 Potential benefits and applications of the SP System

Apart from its strengths and potential in modelling AI-related functions (Appendix A.11), it appears that, in more humdrum terms, the SP System has several potential benefits and applications, several of them described in peer-reviewed papers. These include:

- *Big data.* Somewhat unexpectedly, it has been discovered that the SP System has potential to help solve nine significant problems associated with big data [26]. These are: overcoming the problem of variety in big data; the unsupervised learning of structures and relationships in big data; interpretation of big data via pattern recognition, natural language processing; the analysis of streaming data; compression of big data; model-based coding for the efficient transmission of big data; potential gains in computational and energy efficiency in the analysis of big data; managing errors and uncertainties in data; and visualisation of structure in big data and providing an audit trail in the processing of big data.

- *Autonomous robots.* The SP System opens up a radically new approach to the development of intelligence in autonomous robots [25];

- *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 [22]. 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

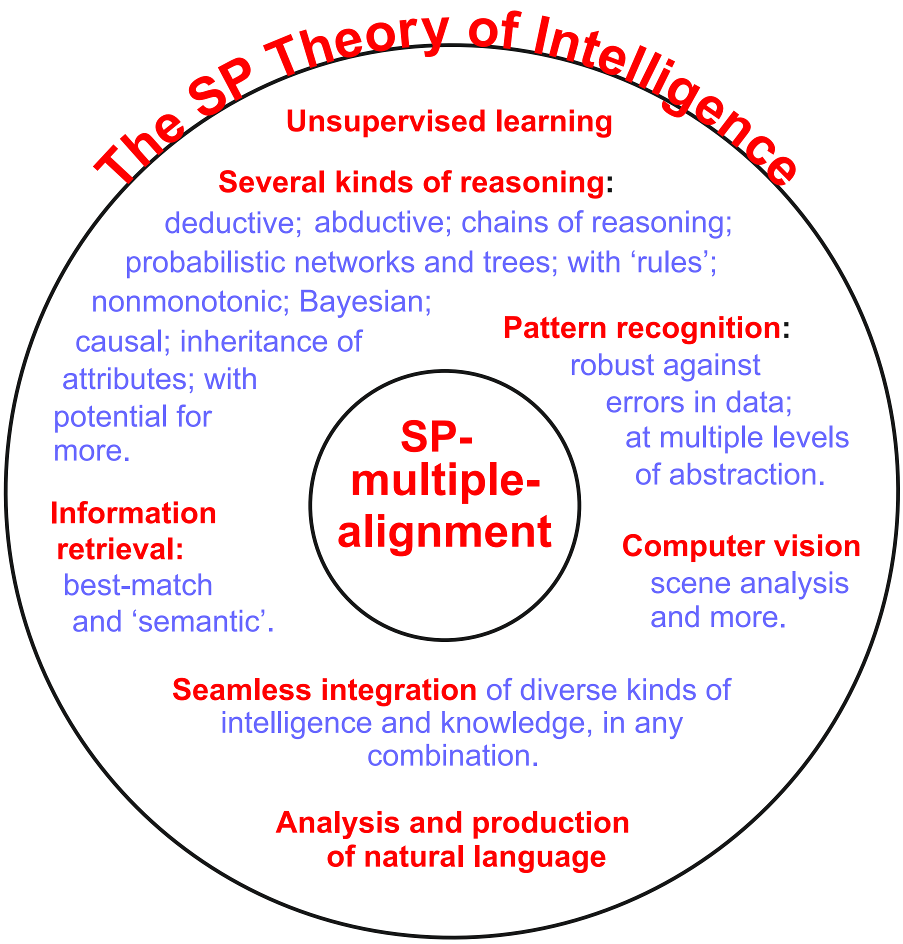

Figure 5: A schematic representation of versatility and integration in the SP System, with SPMA centre stage.

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 [20];

- *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 [24];

- *Neuroscience.* As outlined in Appendix A.9, 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* ([29], [21, 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 [2]. How the SP System may meet the several challenges in this area is described in [28].

- *Other areas of application.* The SP System has potential in several other areas of application including [27]: the simplification and integration of computing systems; best-match and semantic forms of information retrieval; software engineering [31]; the representation of knowledge, reasoning, and the semantic web; information compression; bioinformatics; the detection of computer viruses; and data fusion.

- *Mathematics.* The concept of ICMUP provides an entirely novel interpretation of mathematics [35]. 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.13 Unfinished business and the SP Machine

Like most theories, the SP Theory is not complete. Four pieces of 'unfinished business' are described in [23, Section 3.3]: 1) The SP Computer Model needs to be generalised to include SP-patterns in two dimensions, with associated processing; 2) 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; 3) More work is needed on the development of unsupervised learning in the SP Computer Model; 4) And although the SP Theory has led to the proposal that much of mathematics, perhaps all of it, may be understood as IC [35], research is needed to discover whether or how the SP concepts may be applied in the representation of numbers. A better understanding is also needed of how quantitative concepts such as time, speed, distance, and so on, may be represented in the SP System.

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

More generally, a programme of research is envisaged, with one or more teams of researchers, or individual researchers, to create a more mature *SP Machine*, based on the SP Computer Model, and shown schematically in Figure 6. A roadmap for the development of the SP Machine is described in [16].

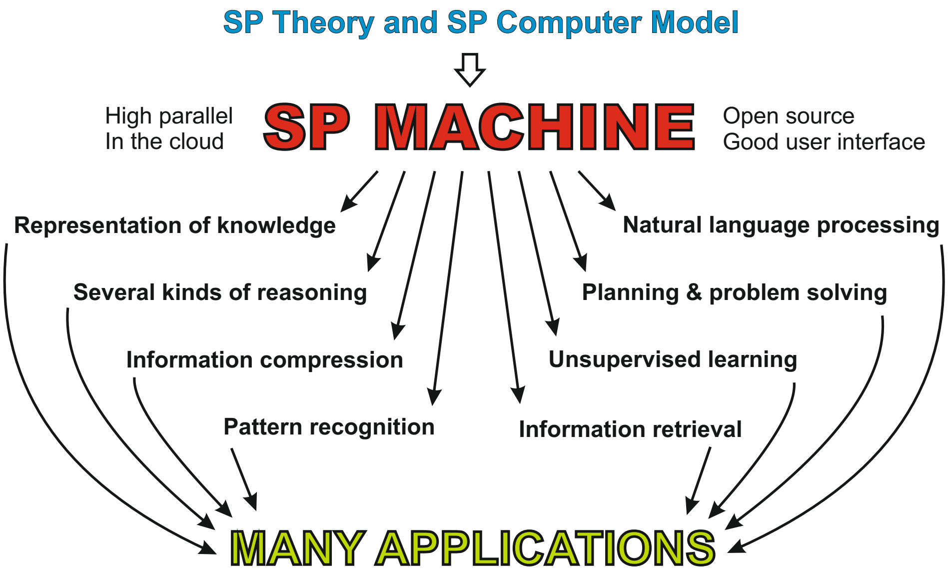

Figure 6: Schematic representation of the development and application of the SP Machine. Reproduced from Figure 2 in [23], with permission.

References

[1] H. B. Barlow. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1:371–394, 1972.

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

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

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

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

[6] J. E. Kelly and S. Hamm. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Columbia University Press, New York, Kindle edition, 2013.

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

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

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

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

[11] P. McCorduck. *Machines Who Think: a Personal Inquiry Into the History and Prospects of Artificial Intelligence*. A. K. Peters Ltd, Natick, MA, second edition, 2004.

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

[13] A. Newell. You can't play 20 questions with nature and win: Projective comments on the papers in this symposium. In W. G. Chase, editor, *Visual Information Processing*, pages 283–308. Academic Press, New York, 1973.

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

[15] M. Page. Connectionist modelling in psychology: a localist manifesto. *Behavioral and Brain Sciences*, 23:443–512, 2000.

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

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

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

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

[21] 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.

[22] 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.

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

[24] 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.

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

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

[27] 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.

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

[29] 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.

[30] 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.

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

[32] 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.

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

[34] 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.

[35] 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.

[36] J. G. Wolff. More problems in AI research and how the SP System may help to solve them. Technical report, CognitionResearch.org, 2020. tinyurl.com/y3l3w5a6.

[37] J. G. Wolff. Problems in AI research and how the SP System may help to solve them. 2020. tinyurl.com/yxb8h822, Submitted for publication.

[38] R. M. Yerkes and J. D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5):459–482, 1908.