A RADICAL ALTERNATIVE TO CURRENT APPROACHES TO AI

J Gerard Wolff*

Tuesday 21st January, 2025

Most of the AI which is in the news today is based on 'deep neural networks' (DNNs). Although DNNs are the basis for some notable successes, they have several weaknesses, most of which may be overcome via the powerful concept of SP-Multiple-Alignment*, a* **a major discovery***, which models diverse aspects of intelligence via compression of information. Somewhat unexpectedly,* **a second major discovery** *in this research is that* **much of mathematics, perhaps all of it, may be seen as a set of techniques for compression of information, and their application**.

The concept of 'deep neural network' (DNN), which is the basis of most AI systems today, started out as the concept of an 'artificial neural network' (ANN), a computational model based loosely on our imperfect understanding of real neural networks in brains.

DNNs could be derived from ANNs by adding extra layers of artificial neurons (which makes DNNs 'deep'), and by adding such things as 'backpropogation'. With developments like that, DNNs performed much better at tasks such as learning new concepts and recognising things.

In later research, DNNs have provided the basis for AI systems with impressive capabilities: such as beating the best human players at the game of Go, a game which is very challenging for people; and greatly speeding up the process of discovering how sequences of amino acids may be folded into three-dimensional structures, a task which is needed in medical research but which is difficult and slow to do without AI assistance.

*Dr Gerry Wolff, BA (Cantab) PhD (Wales) CEng MIEEE, ORCID ID 0000-0002-4624-8904. CognitionResearch.org, Menai Bridge, UK; jgw@cognitionresearch.org; +44 (0) 7746 290775; *X:* @gerrywolff65; *Web:* www.cognitionresearch.org.

Understandably, successes with DNNs have led to much research with them, with a widespread assumption that, despite their shortcomings, DNNs would provide a good foundation for further research towards the development of human-level AI.

So what is this "radical alternative to current approaches to AI", mentioned in the title of this article? It was inspired originally by the ground-breaking discovery by Claude Shannon, published in 1949, that the size of a body of information could be defined precisely with numbers.

This discovery led to pioneering research in the 1950s and later, most notably by Fred Attneave [1, 2], Horace Barlow [3, 4], and Satosi Watanabe [12, 13], around the idea that the processing of information by brains and nervous systems, could be expressed numerically. More specifically, it soon became clear that much of the workings of brains and nervous systems could be understood in terms of information compression (IC) by reducing the amount of repetition in it. Here, below, is a couple of examples, and there are more in [19].

Barlow pointed out that, in mammals at least, the optic nerve (between each eye and the brain), is too small by a wide margin to carry the enormous quantity of visual information coming in to each eye. He suggested that things would work better if information coming in to each eye was made smaller by compressing it. In support of that idea, the neural process of 'lateral inhibition' in the retina of each eye does indeed compress visual information.

Here's another example. If, when we are looking at something, we close our eyes for a moment and open them again, what do we see? Normally, it is the same as what we saw before. But creating a single view out of the before and after views means merging or unifying the two views to make one and thus compressing the information, as shown schematically in Figure 1.

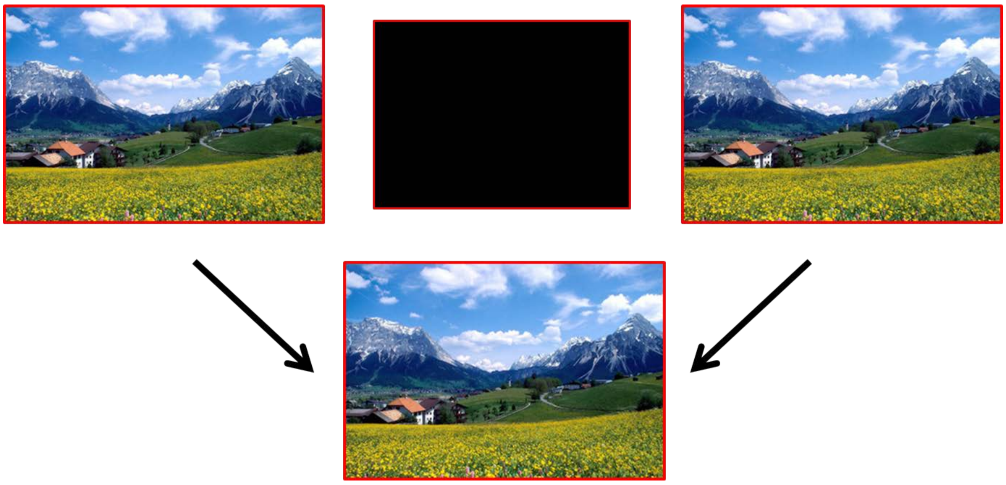

Figure 1: A schematic view of how, if we close our eyes for a moment and open them again, we normally merge the before and after views to make one. The landscape here is from Wallpapers Buzz (www.wallpapersbuzz.com), reproduced with permission.

It is interesting that, as early as 1969, Barlow recognised that IC might be relevant to intelligence: "... the operations needed [for IC] have a rather fascinating similarity to the task of answering an intelligence test, finding an appropriate scientific concept, or other exercises in the use of inductive reasoning. Thus, [IC] may lead one towards understanding something about the organization of memory and intelligence, as well as pattern recognition and discrimination."

One would think that, in the light of the two examples outlined above, with more in [19], most modern attempts to create human-like AI would put IC at centre-stage. But apart from some lip-service for the importance of IC within the DNN literature, and some occasional weak claims for IC in AI, there have been few serious attempts to frame IC as the basis for all or most aspects of intelligence.[1]

In my research, the start was postulating a general theory of AI called the *SP Theory of Intelligence* (SPTI), to be realised eventually in the *SP Computer Model* (SPCM). Here, 'SP' is short for 'Simplicity' and 'Power', two ideas which mean the same as IC because IC may be seen as a search for Simplicity in a body of information, while at the same time seeking to retain as much as possible of its descriptive or explanatory Power.

The SPTI is conceived as a brain-like system as shown schematically in Figure 2, with *New* information (green) coming in via the senses (eyes and ears in the figure), and with some or all of that information compressed and stored as *Old*

[1] Honourable exceptions include research by Ray Solomonoff [10, 11], Nick Chater and Paul Vitányi [5], Marcus Hutter [7], and Jürgen Schmidhuber [8].

information (red), in the brain.

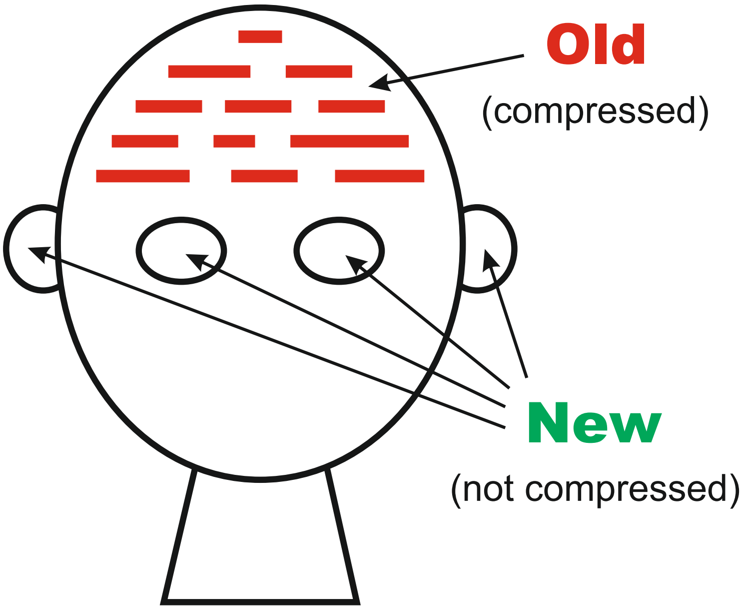

Figure 2: Schematic representation of the SPTI. Reproduced from Figure 1 in [17].

In the SPTI, all kinds of knowledge are represented by *SP-Patterns*, where an SP-Pattern is an array of *SP-Symbols* in one or two dimensions. An SP-Symbol is simply a mark from an alphabet of alternatives where each SP-Symbol can be matched in a yes/no manner with any other SP-Symbol. An SP-Symbol does not have any hidden meaning, such as 'add' for the SP-Symbol '+' in arithmetic, or 'multiply' for the SP-Symbol '×', and so on. Any meaning attaching to an SP-Symbol is provided by one or more other SP-Symbols with which it is associated.

After a fairly lengthy process of R&D, inspired in part by the bioinformatics concept of a 'multiple sequence alignment', I arrived at the concept of an *SP-Multiple-Alignment* (SPMA), expressed in the SPCM. An example of an SPMA created by the SPCM is shown in Figure 3.

With this example, the general idea about how it works is: that row 0 shows a sentence, '`t h e p l u m s a r e r i p e`' which is a New SP-Pattern read in from the system's environment; and each of rows 1 to 9 shows an Old SP-Pattern, which may be a grammatical structure or a word, drawn from a store of many Old SP-Patterns.

The program finds one or more 'good' alignments amongst these SP-Patterns, where 'good' means that a compressed version of the sentence may be created using short codes from the stored SP-Patterns, and the IC achieved via an SPMA is relatively large.

The overall result in the figure shows how the sentence may be analysed into its grammatical parts and subparts, in much the same way that a human specialist in grammatical analysis would analyse the sentence.

```
0                       t h e                p l u m      s                      a r e         r i p e           0
                        | | |                | | | |      |                      | | |         | | | |
1                       | | |          Nrt 6 p l u m #Nrt |                      | | |         | | | |           1
                        | | |           |             |   |                      | | |         | | | |
2                       | | |    N Npl Nrt           #Nrt s #N                   | | |         | | | |           2
                        | | |    |  |                       |                    | | |         | | | |
3                  D 17 t h e #D |  |                       |                    | | |         | | | |           3
                   |          |  |  |                       |                    | | |         | | | |
4            NP 0a D          #D N  |                       #N #NP               | | |         | | | |           4
             |                      |                           |                | | |         | | | |
5            |                      |                           |       V Vpl 11 a r e #V      | | | |           5
             |                      |                           |       |  |           |       | | | |
6            |                      |                           |  VP 1 V  |           #V A    | | | | #A #VP    6
             |                      |                           |  |       |              |    | | | | |   |
7 S Num    ; NP                     |                          #NP VP      |              |    | | | | |  #VP #S 7
     |     |                        |                                      |              |    | | | | |
8   Num PL ;                       Npl                                    Vpl             |    | | | | |         8
                                                                                          |    | | | | |
9                                                                                         A 21 r i p e #A        9
```

Figure 3: The best SPMA created by the SPCM that achieves the effect of parsing a sentence (‘`t h e p l u m s a r e r i p e`’) into its parts and sub-parts, as described in the text. The sentence in row 0 ia a New SP-Pattern, while each of the rows 1 to 9 contains a single Old SP-Pattern, drawn from a relatively-large repository of Old SP-Patterns. Reproduced from Figure A2 in [15].

It turns out that the SPMA concept is remarkably powerful and largely responsible for the versatility of the SPTI across several different aspects of intelligence. That versatility includes strengths in several different aspects of how natural languages may be processed, strengths in several different aspects of the recognition and retrieval of patterns, strengths in several different kinds of reasoning, and more.

In keeping with the versatility of the SPMA across diverse aspects of intelligence, it has been shown to have versatility across several different techniques for IC.

I believe it is fair to say that *the SP-Multiple-Alignment concept is a major discovery with the potential to be as significant for an understanding of intelligence as is the concept of DNA for an understanding of biology. It may prove to be the ‘double helix’ of intelligence!*

The SPMA is also largely responsible for the strengths of the SPTI in overcoming several weaknesses of DNNs. Because DNNs are so dominant in AI research today, their weaknesses are in effect described in a book by Martin Ford which is about problems in AI research described by recognised experts in AI [6]. The following summary describes several of those problems briefly, plus a few others not discussed in the book, each problem with a summary of how the SPTI may solve it.

1. *The symbolic versus sub-symbolic divide*: The need to bridge the divide between symbolic and sub-symbolic kinds of knowledge and processing [22, Section 2]. The concept of an SP-Symbol can represent a relatively large symbolic kind of thing such as a word or a relatively fine-grained kind of thing such as a pixel.

2. *Errors in recognition*: The tendency of DNNs to make large and unexpected errors in recognition [22, Section 3]. The overall workings of the SPTI and its ability, in the recognition of patterns, to correct errors in data suggests that it is unlikely to suffer from these kinds of error.

3. *Natural languages*: The need to strengthen the representation and processing of natural languages, including the understanding of natural languages and the production of natural language from meanings [22, Section 4]. The SPTI has clear potential in the representation and processing of several aspects of natural language.

4. *Unsupervised learning*: Overcoming the challenges of unsupervised learning. Although DNNs can be used in unsupervised mode, they seem to lend themselves best to the supervised learning of tagged examples [22, Section 5]. In contrast, learning in the SPTI is entirely unsupervised.

   It is clear that most human learning, including the learning of our first language or languages [14], is achieved largely via unsupervised learning.

   Incidentally, a working hypothesis in the SP programme of research is that unsupervised learning can be the foundation for all other forms of learning, including learning by imitation, learning by being told, and so on.

5. *Generalisation, over-generalisation, and under-generalisation*: The need for a coherent account of generalisation, over-generalisation (under-fitting) and under-generalisation (over-fitting) [22, Section 6] in unsupervised learning. In the SPTI, those three things are achieved entirely via IC.

6. *Reduce or eliminate the corrupting effect of 'dirty data' in unsupervised learning*: Although this is not mentioned in Ford's book [6], there is the problem of reducing or eliminating the corrupting effect of errors in the data which is the basis for unsupervised learning, a problem for which the SPTI provides a solution.

7. *One-Shot Learning*: Unlike people, DNNs are ill-suited to the learning of usable knowledge from one exposure or experience. The ability to learn usable knowledge from a single exposure or experience is an integral and important part of the SPTI ([22, Section 7]).

8. *Transfer learning*: Although transfer learning—incorporating old learning in newer learning—can be done to some extent with DNNs [9, Section 2.1], DNNs fail to capture the fundamental importance of transfer learning for people, or transfer learning's integral and important part of how the SPTI works [22, Section 8].

9. *Reduced demands for data and for computational resources compared with DNNs*: The ability of the SPTI to learn from a single exposure or experience (above), and the fact that transfer learning is an integral part of how it works (above), is likely to mean that, compared with DNNs, the SPTI will make greatly reduced computational demands and greatly reduced demands for data [22, Section 9]:

10. *Transparency*: By contrast with DNNs, which are opaque in how they represent knowledge, and how they process it, the SPTI is entirely transparent in both the representation and processing of knowledge [22, Section 10].

11. *Probabilistic reasoning*: The SPTI is entirely probabilistic in all its inferences, including the forms of probabilistic reasoning described in [16, Chapter 7], [17, Section 1].

12. *Commonsense reasoning and commonsense knowledge*: Unlike probabilistic reasoning, the area of commonsense reasoning and commonsense knowledge is surprisingly challenging. With qualifications, the SPTI shows some promise in this area [15, 18], [22, Section 12].

13. *How to minimise the risk of accidents with self-driving vehicles*: Notwithstanding the hype about self-driving vehicles, there are still significant problems in minimising the risk of accidents with such vehicles. The SPTI has potential in this area [21], [22, Section 13].

14. *Compositionality in the representation of knowledge*: DNNs are not well suited to the representation of Part-Whole Hierarchies or Class-Inclusion Hierarchies. By contrast, the SPTI has robust capabilities in this area [22, Section 14].

15. *Establishing the importance of IC in AI research*: There is a need to raise awareness of the significance of IC in AI. The importance of IC in the workings of brains and nervous systems is described in [19] and its importance in

the SPTI is described in this article and most other publications about the SPTI [22, Section 15].

16. *Establishing the importance of IC across diverse aspects of AI and human cognition*: A point which deserves emphasis which was not mentioned in [22] is that, while there is some recognition amongst other researchers of the importance of IC in machine learning, there appears to be less recognition of the importance of IC in other aspects of intelligence.

    The importance of IC in the SPTI across several aspects of intelligence is a major strength of the SPTI.

17. *Establishing the importance of a biological perspective in AI research*: There is a need to raise awareness of the importance of a biological perspective in AI research. This is very much part of the SPTI research [22, Section 16].

18. *Distributed versus localist representations for knowledge*: A persistent issue in studies of human learning, perception, and cognition, and in AI, is whether knowledge in brains is represented in distributed or localist form, and which of those two forms works best in AI systems. DNNs employ a distributed form for knowledge, but the SPTI, which is firmly in the localist camp, has distinct advantages compared with DNNs. This is in keeping with other evidence for localist representations in brains [22, Section 17].

19. *The learning of structures from raw data*: DNNs are weak in the learning of structures from raw data, either linuistic or non-linguistic. By contrast, this is a clear advantage in the workings of the SPTI [22, Section 18].

20. *Overcoming the limited scope for adaptation in deep neural networks*: An apparent problem with DNNs is that, unless many DNNs are joined together, each one is designed to learn only one concept, and the learning is restricted to what can be done with a fixed set of layers. By contrast, the SPTI, like people, can learn multiple concepts, and these multiple concepts are often in hierarchies of classes or in part-whole hierarchies. This adaptability is largely because, via the SP-Multiple-Alignment concept, many different SP-Multiple-Alignments may be created in response to one body of data [22, Section 20].

21. *The problem of catastrophic forgetting*: Although there are somewhat clumsy workarounds for this problem, an ordinary DNN is prone to the problem of catastrophic forgetting, meaning that new learning wipes out old learning. There is no such problem with the SPTI which may store new learning independently of old learning, or form composite structures which preserve

both old and new learning, in the manner of transfer learning (above) [22, Section 21].

22. *A weakness of DNNs not mentioned in [22]*: A matter which has become increasingly clear with further thought is that, despite the impressive things that have been done with DNNs,[2] DNNs are relatively restricted in the aspects of intelligence that, without augmentation, they can model. They show little of the versatility of the SPTI in modelling diverse aspects of intelligence.

Somewhat unexpectedly, a second major discovery in this research is that much of mathematics, perhaps all of it, may be seen as a set of techniques for IC, and their application [20].

In summary, the concept of SP-Multiple-Alignment is largely responsible for the versatility of the SP Theory of Intelligence, together with the SP Computer Model, in modelling diverse aspects of intelligence, with information compression as a unifying theme.

As a vehicle for further research, it is envisaged that an industrial-strength *SP Machine* will be developed, based on the SP Computer Model with high levels of parallel processing. It is shown schematically in Figure 4.

[2]Forming part of a system that has beaten the best human players at the game of Go, and forming part of a system that has automated the difficult task of working out likely 3D structures for sequences of amino-acid residues.

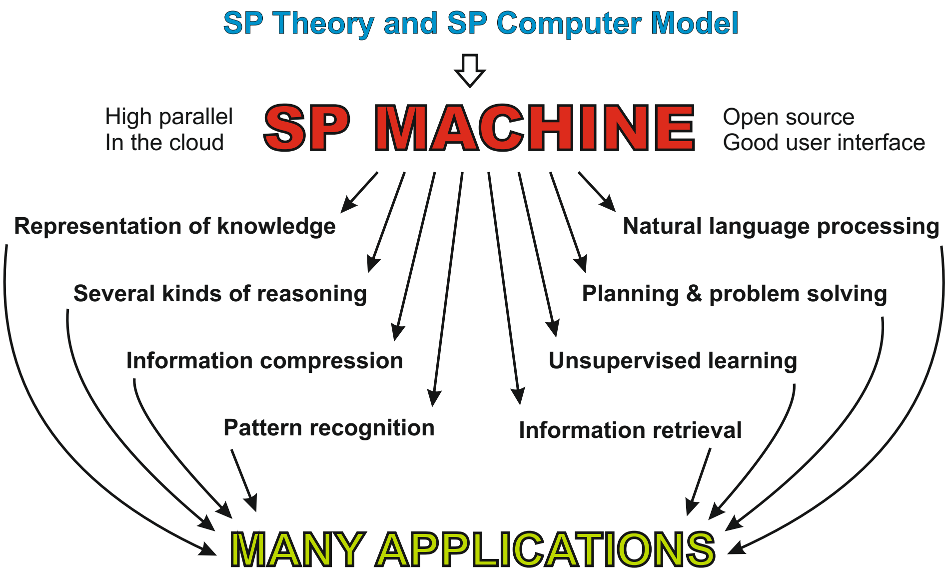

Figure 4: Schematic representation of the development and application of the proposed SP machine. Reproduced from Figure 2 in [17].

References

[1] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954.

[2] F. Attneave. *Applications of Information Theory to Psychology*. Holt, Rinehart and Winston, New York, 1959.

[3] H. B. Barlow. Sensory mechanisms, the reduction of redundancy, and intelligence. In HMSO, editor, *The Mechanisation of Thought Processes*, pages 535–559. Her Majesty's Stationery Office, London, 1959.

[4] H. B. Barlow. Trigger features, adaptation and economy of impulses. In K. N. Leibovic, editor, *Information Processes in the Nervous System*, pages 209–230. Springer, New York, 1969.

[5] N. Chater and P. Vitányi. Simplicity: a unifying principle in cognitive science? *TRENDS in Cognitive Sciences*, 7(1):19–22, 2003.

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

[7] M. Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability.* Springer, Berlin, 2005. www.hutter1.net/ai/uaibook.htm.

[8] J. Schmidhuber. Driven by compression progress: a simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In G. Pezzulo, M. V. Butz, O. Sigaud, and G. Baldassarre, editors, *Anticipatory Behavior in Adaptive Learning Systems, from Sensorimotor to Higher-level Cognitive Capabilities*, Lecture Notes in Artificial Intelligence. Springer, Berlin, 2009.

[9] J. Schmidhuber. One big net for everything. Technical report, The Swiss AI Lab, IDSIA, 2018.

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

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

[12] Satosi Watanabe. Information-theoretical aspects of inductive and deductive inference. *IBM Journal of Research and Development*, 4:208–231, 1960.

[13] Satosi Watanabe. Pattern recognition as information compression. In Satosi Watanabe, editor, *Frontiers of Pattern Recognition.* Academic Press, New York, 1972.

[14] 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.

[15] J. G. Wolff. The SP Theory of Intelligence, and its realisation in the SP Computer Model as a foundation for the development of artificial general intelligence. *Analytics*, 2(1):163–192, 20023. Web: www.mdpi.com/2813-2203/2/1/10.

[16] J. G. Wolff. *Unifying Computing and Cognition: the SP Theory and Its Applications.* CognitionResearch.org, Menai Bridge, 2006. Distributors include Amazon.com and Amazon.co.uk. The print version is produced via print-on-demand from "INGRAM Lightning Source" and, via that technology, is unlikely to go out of print.

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

[18] 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. bit.ly/2ME8DOA.

[19] J. G. Wolff. Information compression as a unifying principle in human learning, perception, and cognition. *Complexity*, 2019:38, February 2019.

[20] 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. This paper is reproduced in *New Ideas Concerning Science and Technology*, Vol. 13, 12 June 2021, Pages 132–169, doi.org/10.9734/bpi/nicst/v13/8449D.

[21] J. G. Wolff. A proposed solution to problems in learning the knowledge needed by self-driving vehicles. Technical report, CognitionResearch.org, 2021. tinyurl.com/n6vpxcuf.

[22] J. G. Wolff. Twenty significant problems in AI research, with potential solutions via the SP Theory of Intelligence and its realisation in the SP Computer Model. *Foundations*, 2:1045–1079, 2022.