analytics

MDPI

Article

The *SP Theory of Intelligence*, and Its Realisation in the *SP Computer Model*, as a Foundation for the Development of Artificial General Intelligence

J. Gerard Wolff [1]

CognitionResearch.org, 18 Penlon, Menai Bridge, Anglesey LL59 5LR, UK; jgw@cognitionresearch.org

Abstract: The theme of this paper is that the *SP Theory of Intelligence* (SPTI), and its realisation in the *SP Computer Model*, is a promising foundation for the development of artificial intelligence at the level of people or higher, also known as 'artificial general intelligence' (AGI). The SPTI, and alternatives to the SPTI chosen to be representative of potential foundations for the development of AGI, are considered and compared. A key principle in the SPTI and its development is the importance of information compression (IC) in human learning, perception, and cognition. More specifically, IC in the SPTI is achieved via the powerful concept of *SP-multiple-alignment*, the key to the versatility of the SPTI in diverse aspects of intelligence, and thus a favourable combination of Simplicity with descriptive and explanatory Power. Since there are many uncertainties between where we are now and, far into the future, anything that might qualify as an AGI, a multi-pronged attack on the problem is needed. The SPTI qualifies as the basis for one of those prongs. Although it will take time to achieve AGI, there is potential along the road for many useful benefits and applications of the research.

Keywords: artificial general intelligence; information compression; SP Computer Model; SP-multiple-alignment; SP Theory of Intelligence

1. Introduction

The theme of this paper is that the *SP Theory of Intelligence* (SPTI), and its realisation in the *SP Computer Model* (SPCM), is a promising foundation for the development of artificial intelligence at the level of people, or higher, also known as 'artificial general intelligence' (AGI).

Readers who are not already familiar with the SPTI are urged to read the introduction to it in Appendix A, and perhaps other sources referenced there.

Because AGI is far from being achieved by any system, the focus of this paper is on *foundations* for the development of AGI, abbreviated as 'FDAGIs'.

In the paper, six alternatives to the SPTI, abbreviated as 'ALTs', are described and compared with the SPTI in terms of their potential as FDAGIs.

Although the SPTI scores well in comparison with the ALTs, it would be wrong, and quite unrealistic, to assume that all research effort should be switched into its development. Since there are many uncertainties between where we are now and, far into the future, anything that may come close to human intelligence, a multi-pronged attack on the problem is needed. And the SPTI qualifies as the basis for one of those prongs.

Although the quest for AGI, if it succeeds, is likely to take some time, any programme of research like this is likely to produce many potential benefits and applications at points along the road.

Presentation

The main sections of the paper are these:

- In Section 2, six ALTs are described.

- Section 3 describes how the ALTs and the SPTI may be evaluated as potential FDAGIs.
- Section 4 evaluates the ALTs and the SPTI as FDAGIs in terms of the criteria in Section 3.
- Section 5 summarises how the SPTI compares with the ALTs, and how the ALTs compare with each other, in terms of the criteria in Section 3.
- The paper concludes that the SPTI is indeed relatively promising as an FDAGI (Section 6).

Abbreviations are listed with their meanings after the Conclusion (Section 6), and the meaning of each one is also made clear where it is first used.

The appendices that follow the Conclusion (Section 6) should not be regarded as part of the main substance of the paper in Sections 1–6. They present background information in support of the main presentation. Although they describe some elements of previous publications in this programme of research, they are merely assisting the main presentation, and should not be regarded as self-plagiarism.

- Appendix A introduces the SPTI, with pointers to where fuller information may be found.
- Appendix B describes strengths of the SPTI in both intelligence-related and non-intelligence-related domains.
- Appendix C describes topics related to the role of information compression (IC) in biology, especially HLPC, and in the SPTI.
- Appendix D provides an entirely novel perspective on the foundations of mathematics.
- Appendix E describes the benefits of a top-down, breadth-first research strategy with wide scope.

2. Six Systems That May Serve as FDAGIs

This section describes six systems, each of which is a potential FDAGI.

2.1. 'The Society of Mind' by Marvin Minsky

In the book *The Society of Mind* [1], Marvin Minsky writes:

"I'll call 'Society of Mind' this scheme in which each mind is made of many smaller processes. These we'll call *agents*. Each mental agent by itself can only do some simple thing that needs no mind or thought at all. Yet when we join these agents in societies—in certain very special ways—this leads to true intelligence." [1] (Prolog, p. 17), emphasis in the original.

Later, he writes:

"Since most of the statements in this book are speculations, it would have been too tedious to mention this on every page. ... Each idea should be seen not as a firm hypothesis about the mind, but as another implement to keep inside one's toolbox for making theories about the mind." [1] (Postscript, p, 323).

In short, the society-of-mind idea (SOM) is largely a council of despair: the human mind is too complicated for there to be any coherent theory of its structure and workings. Nevertheless, the SOM represents a distinct approach to the development of AI which is relevant to issues considered in this paper.

2.2. 'Gato' from DeepMind

A team of researchers from the DeepMind company has created "A generalist agent" called 'Gato', described in [2]. They say:

"In this paper, we describe the current iteration of a general-purpose agent which we call Gato, instantiated as a single, large, transformer sequence model. With a single set of weights, Gato can engage in dialogue, caption images, stack blocks

with a real robot arm, outperform humans at playing Atari games, navigate in simulated 3D environments, follow instructions, and more.

> "While no agent can be expected to excel in all imaginable control tasks, especially those far outside of its training distribution, we here test the hypothesis that training an agent which is generally capable on a large number of tasks is possible; and that this general agent can be adapted with little extra data to succeed at an even larger number of tasks." [2] (p. 2).

A "transformer sequence model" mentioned in the quote is "a transformer [deep] neural network akin to a large language model" [2] (Caption to Figure 2, p. 1), and 'transformer' means that different parts of the input data are given different weights, somewhat like human attention.

The paper [2] does not say explicitly that the authors are aiming for AGI, but it is clear that they see Gato as a stepping stone towards AGI: in [2] (p. 18) they reference the book by Nick Bostrom [3] with its main focus on possible dangers from the development of AGI.

2.3. 'DALL·E 2' from OpenAI

The OpenAI research organisation says on its website openai.com (accessed on 1 December 2022):

> "OpenAI's mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity.

> "We will attempt to directly build safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome."

The latest system in this quest is 'DALL·E 2', which in 2022-06-09 was described as follows on the OpenAI website openai.com/dall-e-2/ (accessed on 1 December 2022):

> "DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.

> "DALL·E 2 can make realistic edits to existing images from a natural language caption. It can add and remove elements while taking shadows, reflections, and textures into account.

> "DALL·E 2 can take an image and create different variations of it inspired by the original.

> "DALL·E 2 has learned the relationship between images and the text used to describe them. It uses a process called "diffusion," which starts with a pattern of random dots and gradually alters that pattern towards an image when it recognizes specific aspects of that image.

> "In January 2021, OpenAI introduced DALL·E. One year later, our newest system, DALL·E 2, generates more realistic and accurate images with 4x greater resolution.

> "DALL·E 2 is preferred over DALL·E 1 for its caption matching and photorealism when evaluators were asked to compare 1,000 image generations from each model.

> "DALL·E 2 is a research project which we currently do not make available in our API. As part of our effort to develop and deploy AI responsibly, we are studying DALL·E's limitations and capabilities with a select group of users. Safety mitigations we have already developed include: Preventing Harmful Generations ... Curbing Misuse ... Phased Deployment Based on Learning ...

> "Our hope is that DALL·E 2 will empower people to express themselves creatively. DALL·E 2 also helps us understand how advanced AI systems see and understand our world, which is critical to our mission of creating AI that benefits humanity.".

Overall, it is clear that OpenAI is aiming for AGI, and that DALL·E 2 is seen as a stepping stone in that direction. Hence, in the terms of this paper, it has potential as an FDAGI.

2.4. 'Soar' from John Laird, Paul Rosenbloom, and Allen Newell

The 'Soar' cognitive architecture was first described in [4] and it is described at various points throughout Allen Newell's book *Unified Theories of Cognition* [5]. Now, the most comprehensive description of Soar is in *The Soar Cognitive Architecture* by John Laird [6]. And there is a more recent introduction to Soar in [7].

Some idea of how Soar is organised may be gained from Figure 1 which shows its memories, processing modules, learning modules and their connections.
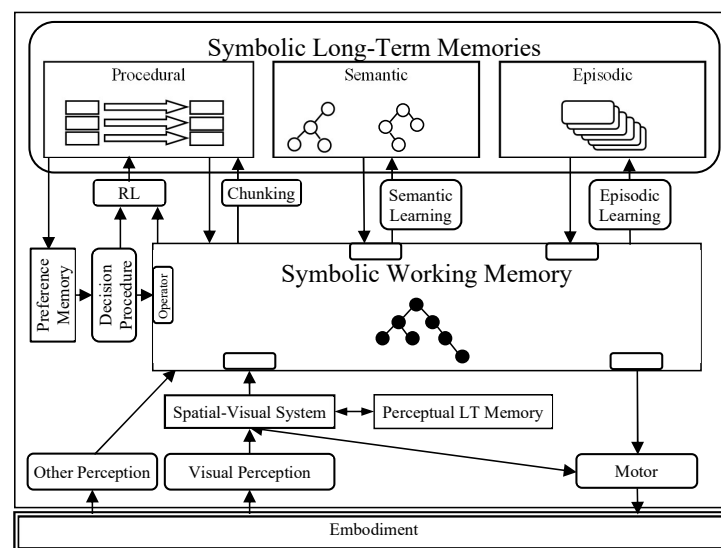

Figure 1. The structure of Soar memories, processing modules, learning modules and their connections. Reproduced from [7] (Figure 1), with permission from John Laird.

In Laird's introduction to Soar [6], it is described like this:

"Soar is meant to be a general cognitive architecture [8] that provides the fixed computational building blocks for creating AI agents whose cognitive characteristics and capabilities approach those found in humans [5,6]. A cognitive architecture is not a single algorithm or method for solving a specific problem; rather, it is the task-independent infrastructure that learns, encodes, and applies an agent's knowledge to produce behavior, making a cognitive architecture a software implementation of a general theory of intelligence. One of the most difficult challenges in cognitive architecture design is to create sufficient structure to support coherent and purposeful behavior, while at the same time providing sufficient flexibility so that an agent can adapt (via learning) to the specifics of its tasks and environment. The structure of Soar is inspired by the human mind and as Allen Newell (Newell, 1990) suggested over 30 years ago, it attempts to embody a unified theory of cognition." [7] (p. 1).

Apart from being a cognitive architecture, Soar may of course also be a potential FDAGI.

2.5. 'ACT-R' from John Anderson, Christian Lebiere, and Others

'ACT-R' (which stands for "Adaptive Control of Thought—Rational") is a cognitive architecture developed by John Anderson and Christian Lebiere, and others, which, like Soar, is inspired by Newell's writings on the need for unified theories of cognition. A relatively full account is in Anderson and Lebiere's book *The Atomic Components of Thought* [9].

On the ACT-R website act-r.psy.cmu.edu (accessed 15 December 2022), ACT-R is described as "A theory for understanding and simulating human cognition". In an overview of the system, the authors say:

> "Adaptive control of thought–rational (ACT–R ...) has evolved into a theory that consists of multiple modules but also explains how these modules are integrated to produce coherent cognition. The perceptual-motor modules, the goal module, and the declarative memory module are presented as examples of specialized systems in ACT–R. These modules are associated with distinct cortical regions. These modules place chunks in buffers where they can be detected by a production system that responds to patterns of information in the buffers. At any point in time, a single production rule is selected to respond to the current pattern. Subsymbolic processes serve to guide the selection of rules to fire as well as the internal operations of some modules. Much of learning involves tuning of these subsymbolic processes." [10] (Abstract).

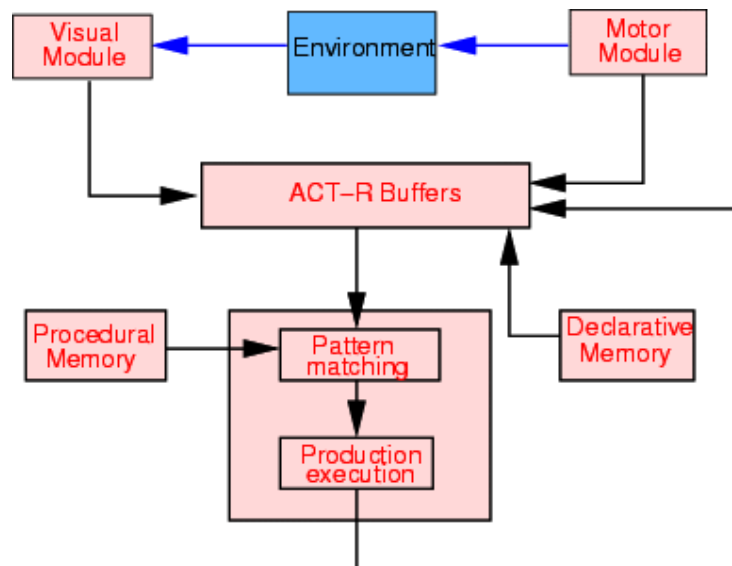A schematic representation of the ACT-R cognitive architecture is shown in Figure 2.

Figure 2. Schematic representation of the ACT-R cognitive architecture. Reproduced, with permission, from the ACT-R website, http://act-r.psy.cmu.edu/about/act-r.psy.cmu.edu/about/ (accessed on 1 July 2022).

As with Soar (Section 2.4), ACT-R is a cognitive architecture, but it is also a potential FDAGI.

2.6. 'NARS' from Pei Wang

"Artificial General Intelligence" is a group of researchers, aiming for AGI, who hold an annual conference about their research. Details of all the conferences held to date may be seen via agi-conference.org (accessed on 25 November 2022). Although AGI is the goal of this research, it is acknowledged that the problem is difficult, and no real breakthrough has yet been achieved.

One of the systems that is associated with that AGI research is Pei Wang's *Non-Axiomatic Reasoning System* (NARS) (see, for example, Refs. [11–13], and sources referenced there).

NARS is an environment for *Non-Axiomatic Logic* (NAL): "This [NAL] logic is designed for the creation of general-purpose Artificial Intelligence (AI) systems, by formulating the fundamental regularity of human thinking at a general level." [14] (Location 94). Wang recognises that, to achieve AGI, there must be integration of different aspects of intelligence.

In [11] he discusses how reasoning and learning may be integrated within the NARS environment.

So NARS is a potential FDAGI, and is evaluated in this paper alongside other ALTs and the SPTI.

3. How the ALTs and the SPTI May Be Evaluated, Viewed as FDAGIs

This section describes how the ALTs and the SPTI may be evaluated, viewed as FDAGIs.

Although the ALTs are not normally seen as scientific theories, they represent ideas about the nature of AGI which seem to need the same kind of evaluation as may be applied to scientific theories. Hence, this section applies some widely-accepted conclusions from the philosophy of science about desirable or mandatory features in any scientific theory:

- *Ockham's Razor*. It has long been recognised that any good scientific theory should conform to the rule that, in words attributed to the English Franciscan friar William of Ockham: "Entities should not be multiplied beyond necessity". As described in Appendix C.1, this means that any good theory should combine conceptual *Simplicity* with descriptive or explanatory *Power*.
- *Falsifiability*. Karl Popper proposed that any scientific theory should be *falsifiable* [15,16], meaning that, for the given theory, it should be possible to conceive of evidence that would show that the theory was wrong.

3.1. Evaluation Headings

In Section 4, which follows this main section, the ALTs and the SPTI are evaluated under the four headings shown here:

- *Simplicity*. It may not be possible to measure the Simplicity of a system precisely in terms of bits or bytes of information, but a more informal assessment may nevertheless be useful. A large system should have a low score (0) for Simplicity, while a small system may have a high score (2), and something in between may be given a score of 1.
- *Power in Modelling Aspects of Human Intelligence*. As with Simplicity, it may not be possible to measure Power precisely in terms of bits or bytes of information, but a more informal measure may nevertheless be useful. A system with high Power should have a high score (2), while a system with low Power should have a low score (0), and something in between may be given a score of 1.
- *Other Strengths or Weaknesses*. In addition to Simplicity and Power, there may be other strengths or weaknesses of a given system that are relevant to the development of AGI. An assessment that emphasises weaknesses is scored −1, an assessment which is mainly about strengths is scored 1, and when strengths and weaknesses are at least roughly equal, or when there are none, the score is 0.
  Under this heading it is appropriate to consider anything that suggests that the given ALT or the SPTI, viewed as a theory of intelligence, might be unfalsifiable, as described above.
- *Combined Score*. To simplify comparisons amongst two or more systems, a 'combined score' is calculated as the sum of the scores from the three headings above.

3.2. A Definition of Intelligence

Evaluating the Power of the ALTs and the SPTI in Section 4, below, means assessing their strengths or weaknesses in modelling human intelligence.

Arriving at a meaning for 'human intelligence' that will be widely accepted is difficult, witness the variety of definitions that have been advanced, many of which have been documented by Shane Legg and Marcus Hutter in [17].

As mentioned in Section 3, this section describes a definition of intelligence that may serve in the evaluation of any system, viewed as a potential FDAGI. This definition reflects the state of the SPTI as it has been developed to date. No doubt, other features will be

added with further development of the SPTI. And, given the complexity of the concept of human intelligence, it should not be surprising to find that other sets of features, with at least equal claims for their validity, have been adopted by other researchers in other studies.

Some readers may object that the adoption of a definition of intelligence that reflects what the SPTI can do means the creation of an unreasonable bias in favour of the SPTI. That might be true if that definition was the only one to be adopted in the assessment of potential FDAGIs. However, as noted in the previous paragraph, alternative definitions of intelligence are recognised in this paper and given the same weight as the definition in Section 3.4, below.

3.3. Taking Account of the Distinction Between the 'Core' of Each System and How It May Be When It Is Enhanced Via Learning or Programming

In Soar, ACT-R, and the SPTI, there is a 'core' to the system which may be enhanced via learning or via programming.

In these cases, it is probably best for the core to be the primary focus of the evaluation. The main reasons are:

- The core is relatively well defined but the way in which the core may be enhanced via learning or programming is less well defined.
- The core comprises aspects of human intelligence, such as those described in Section 3.4, which are probably inborn and not learned.
- It seems likely that features like those described in Section 3.4 are desirable features in any FDAGI. Likewise for other intelligence-related features with comparable validity mentioned near the beginning of Section 3.2.

3.4. The SPTI Core as a Definition of Intelligence

It seems likely that, although unsupervised learning is an important part of the SPTI, the features described here are, in people, inborn and not learned. These features, which may be seen as the central 'core' of human intelligence (Section 3.3), are described quite fully in [18] and more briefly in [19].

What follows is an outline of the SPTI definition of intelligence. There is more detail in Appendix B.1.

- *Information Compression*. In view of substantial evidence for the importance of IC in HLPC [20], IC should be seen as an important feature of human intelligence.
- *Natural Language Processing*. Under the general heading of "Natural Language Processing" are capabilities that facilitate the learning and use of natural languages. These include:
  - The ability to structure syntactic and semantic knowledge into hierarchies of classes and sub-classes, and into parts and sub-parts.
  - The ability to integrate syntactic and semantic knowledge.
  - The ability to encode discontinuous dependencies in syntax such as the number dependency (singular or plural) between the subject of a sentence and its main verb, or gender dependencies (masculine or feminine) in French—where 'discontinuous' means that the dependencies can jump over arbitrarily large intervening structures.
  - Also important in this connection is that different kinds of dependency (e.g., number and gender) can co-exist without interfering with each other.
  - The ability to accommodate recursive structures in syntax, and perhaps also in semantics.
- *Recognition and Retrieval*. Capabilities that facilitate recognition of entities or retrieval of information include:
  - The ability to recognise something or retrieve information on the strength of a good partial match between features as well as an exact match.

- Recognition or retrieval within a class-inclusion hierarchy with 'inheritance' of attributes, and recognition or retrieval within a hierarchy of parts and sub-parts.

- *Probabilistic Reasoning*. Capabilities here include: one-step 'deductive' reasoning; abductive reasoning; probabilistic networks and trees; reasoning with 'rules'; non-monotonic reasoning; 'explaining away' meaning 'If A implies B, C implies B, and B is true, then finding that C is true makes A less credible.' In other words, finding a second explanation for an item of data makes the first explanation less credible; probabilistic causal diagnosis; and reasoning which is not supported by evidence.

- *Planning and Problem Solving*. Capabilities here include:

  - The ability to plan a route, such as for example a flying route between cities A and B, given information about direct flights between pairs of cities including those that may be assembled into a route between A and B.
  - The ability to solve analogues of GAPs in textual form, where a 'GAP' is a Geometric Analogy Problem.

- *Unsupervised Learning*. Chapter 9 of [18] describes how the SPCM may achieve unsupervised learning from a body of 'raw' data, **I**, to create an *SP-grammar*, **G**, and an *Encoding* of **I** in terms of **G**, where the encoding may be referred to as **E**. At present the learning process has shortcomings summarised in [19] (Section 3.3) but it appears that these problems are soluble.

  In its essentials, unsupervised learning in the SPCM means searching for one or more 'good' SP-grammars, where a good SP-grammar is a set of SP-patterns which is relatively effective in the economical encoding of **I** via SP-multiple-alignment (SPMA, Appendix C.1),

4. Evaluation of the ALTs and the SPTI as FDAGIs

This section evaluates the ALTs and the SPTI in terms of the criteria described in Section 3. These evaluations have been made carefully, trying to avoid biases or distortions arising from the author's association with the SPTI.

4.1. The SOM

Since no computational core is specified for the SOM, the evaluation necessarily focuses on the nature of the system when it is populated by many small agents:

- *Simplicity, Score = 0*. At first sight, the SOM looks simple: intelligence is merely lots of little agents. However, a little reflection suggests that in the SOM as Minsky describes it:

  - There would be large numbers of agents, which collectively would be far from simple.
  - Since IC is not mentioned, we may assume that there would be no corresponding benefit via the simplification of those many agents.

  In short, the SOM is weak in terms of Simplicity and it has been assigned a corresponding score of 0.

- *Power, Score = 0*. Regarding the (intelligence-related) descriptive/explanatory Power of the SOM:

  - In the second of the quotes in Section 2.1, Minsky says: "... most of the statements in this book are speculations" [1] (p. 17). In other words, there is little evidence in support of the proposals in the book.
  - The SOM concept of human intelligence is little better than a theory that merely redescribes what it is meant to explain (Appendix C.1).

    An example may be seen in [1] (p. 20) where Minsky suggests that the picking up of a cup of tea by a person may be analysed into such agents as one for *grasping* the cup, an agent for *balancing* the cup to avoid spills, an agent for the *thirst* of the person picking up the cup, and an agent for *moving* the cup to the person's lips.

In short, the SOM is weak in terms of its descriptive/explanatory Power and it has been assigned a corresponding score of 0.

- *Other Strengths or Weaknesses, Score = −1.* Apart from the weakness of the SOM in terms of Simplicity and Power, it is weak because, in terms of Popper's ideas: the theory is unfalsifiable. This is because any proposed falsification of the theory could be met by the addition, omission, or substitution of agents within the theory. The SOM is too malleable to be plausible as a theory of human intelligence. Because of this additional weakness in the SOM, it has been assigned a score of −1 for 'other strengths or weaknesses'.
- *Combined Score = 0 + 0 − 1 = −1.*

4.2. Gato

Since the case for Gato depends critically on the idea that intelligence may be built up via the learning of many more-or-less simple skills, it should be evaluated in terms of how it will be when many such skills have been learned.

- *Simplicity, Score = 1.* For Gato to do many things, as indicated in the quote at the beginning of Section 2.2, it must be trained in all of them (apart from the ability to learn required to learn those many skills). And training across many such tasks means adding a substantial amount of information to the core system. From the perspective of Simplicity, things do not look so good for Gato. However, since Gato can achieve quite a lot with a small computational core, it seems reasonable to give it a middling score of 1.
- *Power, Score = 2.* As noted in [2] (p. 2), Gato, with appropriate training, "... can... engage in dialogue, caption images, stack blocks with a real robot arm, outperform humans at playing Atari games, navigate in simulated 3D environments, follow instructions, and more."

    Although the researchers' strategy is not entirely clear, it seems that the Gato research is based on the assumption in [2] that any artificial system that can learn the kinds of things that people can learn may be seen to have achieved AGI, or it will have done after it has been scaled up. This assumption has been adopted by at least one of the authors of the [2] paper:

    > "According to Doctor Nando de Freitas, a lead researcher at Google's DeepMind, humanity is apparently on the verge of solving artificial general intelligence (AGI) within our lifetimes.

    > "In response to an opinion piece penned by [the author of the article in which this quote appears], the scientist posted a thread on Twitter that began with what's perhaps the boldest statement [seen by writers at The Guardian] from anyone at DeepMind concerning its current progress toward AGI:

    > > '... My opinion: It's all about scale now! The Game is Over!' (Tristan Green, *The Guardian*, "DeepMind researcher claims new 'Gato' AI could lead to AGI, says 'the game is over!'", 16 May 2022).

    There are (at least) two difficulties with the assumption that AGI is merely the ability to learn the variety of skills that can be learned by people:

    - There is an implicit assumption that the learning of those skills is achieved via supervised learning or reinforcement learning, ("For simplicity Gato was trained offline in a purely supervised manner; however, in principle, there is no reason it could not also be trained with either offline or online reinforcement learning (RL)" [2] (p. 3] but it appears that: most human learning is unsupervised, meaning that it is learning without predefined associations between, for example, words and pictures, and without rewards or punishments [18] (Chapter 9).
    - Gato suffers from a weakness that is similar to that in the SOM and described in Section 4.1 (item 'Power'): the belief that human intelligence is merely a collection

of skills, somewhat like the discredited theory that human cognition may be understood as a collection of instincts including such implausible instincts as a putting-on-of-shoes instinct, a planting-of-seeds-instinct, a car-driving instinct, and so on.

With regard to the second point, an alternative view, adopted with varying degrees of confidence by many people working in AI, is that some capabilities are more central to the concept of intelligence than others (Section 3.2). It seems, for example, that skills like those outlined in Section 3.2 are likely to be inborn and fundamental in human intelligence, whereas others such as how to drive a car, how to play cricket, how to play the piano, and so on, are learned and not inborn, and only tangentially relevant to our concept of intelligence.

Thus, in brief, with regard to the Power of Gato: the model of learning that has been adopted in Gato is unlikely to conform to the fundamentals of learning in people; and the emphasis in Gato on the variety of skills that the system may learn says little or nothing about the fundamentals of intelligence. It is possible that, like the SOM, Gato is too malleable to be plausible as a theory of human intelligence.

However, since it is not impossible that human intelligence is merely a knowledge of many skills, and since Gato clearly has the potential to model many of them, the Power of Gato is hereby assigned a score of 2 (in the range 0 to 2).

- *Other strengths or weaknesses, Score* $= -1$. Although Gato, with appropriate training, can demonstrate a variety of capabilities, we may say in a similar way that, with the installation of appropriate apps, a smartphone or any ordinary computer can demonstrate a variety of capabilities, and these may include AI capabilities. Does this make the smartphone or ordinary computer a good FDAGI? Certainly not. This is an additional reason to doubt the potential of Gato as an FDAGI.

  Apart from that, the reliance of Gato on a variant of the DNN model is a weakness that stems from the well-known shortcomings of DNNs, most of which are summarised in Appendix B.2.2. Hence Gato has been assigned a score of -1 under the heading 'other strengths or weaknesses'.

- *Combined Score* $= 1 + 2 - 1 = 2$.

4.3. DALL·E 2

From the description of DALL·E 2 in Section 2.3, it is clear that in attempting to integrate two aspects of intelligence—the processing of images and the processing of natural language—it is being developed within a bottom-up strategy with its associated problems, described in Appendix E. This has a bearing on judgements both of its Simplicity and of its descriptive/explanatory Power:

- *Simplicity, Score* $= 1$. Because DALL·E 2 is being developed via a bottom-up strategy, and because it is at a relatively early stage in that process, it may be seen to be relatively strong in terms of Simplicity.

  Although a mature version of DALL·E 2 that embraces several aspects of intelligence is likely to be relatively large, it seems best to stick with the assessment of the Simplicity of DALL·E 2 as it is now. Hence, the Simplicity of DALL·E 2 is hereby assigned a middling Score of 1.

- *Power, Score* $= 2$. In some respects, DALL·E 2 has shortcomings in its Power to model aspects of human intelligence:

  – Although the bottom-up strategy for the development of DALL·E 2 seems plausible and is popular, it has apparently never proved successful (Appendix E.2). For this reason, and because DALL·E 2 is still at an early stage of that bottom-up process, its Power with respect to the development of AGI may be judged to be weak.

  – While some parts of the project are relevant to the development of AGI (e.g., The creation of "a new AI system that can create realistic images and art from a

description in natural language", see Section 2.3.), other parts seem to be more focussed on meeting the needs of potential users of the system (e.g., "Safety mitigations we have already developed include: Preventing Harmful Generations ... Curbing Misuse ... Phased Deployment Based on Learning ..."), see Section 2.3. More generally, the project is not tightly focussed on modelling aspects of human intelligence.

- – One study concludes that "DALL·E 2 is unable reliably to infer meanings that are consistent with the syntax. These results challenge recent claims concerning the capacity of such systems to understand of human language." [21] (Abstract).
- – DALL·E is a version of GPT-3 trained on both images and text [22] (p. 4, footnote). However, despite the impressive capabilities of GPT-3 with natural language, it has a tendency to produce 'tortured phrases' such as 'colossal information' instead of 'big data', 'counterfeit consciousness' instead of 'artificial intelligence', and more [23].

However, it would be perverse to give DALL·E 2 a score of 0 for Power because the kinds of things it can do are undoubtedly impressive. Hence it has seemed most appropriate to give it a Power score of 2, on the scale 0 to 2.

- *Other strengths or weaknesses, Score* $= -1$. DALL·E 2 is a 'transformer' model [24], and a transformer model is a kind of DNN (Section 2.2), and DNNs have well known shortcomings compared with people and the SPTI, most of which are summarised in Appendix B.2.2. There is more detail about the shortcomings of DNNs in [25]. As with Gato, these shortcomings are why DALL·E 2 has been assigned a score of -1 for 'other strengths or weaknesses'.
- Combined Score $= 1 + 2 - 1 = 2$.

4.4. Soar

Soar is one of the systems mentioned in Section 3.3 which is designed as an intelligence-related 'core' which needs to be programmed to create one or more specific systems. Hence, it is evaluated in this and following subsections in terms of its intelligence-related core features rather than the features of any of the systems that may be derived from the Soar architecture via programming.

- *Simplicity, Score* $= 1$. In keeping with what is said about ACT-R, and the SPTI, in Section 3.3, it seems best to evaluate the Simplicity of Soar in terms of the size of its important computational 'core'. Since that core expresses several aspects of intelligence via mechanisms for achieving those aspects of intelligence (next bullet point), and since, overall, there is little simplification via integration (as suggested by the structures shown in Figure 1), it seems reasonable to say that Soar has a middling score of 1 for Simplicity, that it is neither very complex nor very succinct.
- *Power, Score* $= 2$. In John Laird's book, *The Soar Cognitive Architecture* [6], the way in which Soar expresses aspects of human intelligence is described largely via descriptions of the mechanisms within Soar which relate to each aspect. Nevertheless, the descriptive/explanatory Power of Soar with respect to each aspect of intelligence comes over reasonably clearly. For example:

  - – Chapter 6 describes how Soar achieves 'chunking', a widely-recognised aspect of human learning which became prominent largely because of George Miller's paper about "The magical number seven, plus or minus two: ..." [26].
  - – Chapter 7 describes how reinforcement learning may be achieved via the modification or tuning of existing rules. Although unsupervised learning is probably more fundamental, there is no doubt that reinforcement can play a part in learning.
  - – Chapter 8 (co-authored with Yongjia Wang and Nate Derbinsky) describes Soar's semantic memory, "... a repository for long-term declarative knowledge that supplements what is contained in short-term working memory (and production memory)." [6] (p. 203).

- And so on.

Clearly, Soar embodies a definition of intelligence that is different from that described in Section 3.2. However, as noted in that section, there may be other definitions of intelligence with equal validity. Overall, it seems reasonable to say that the Power of Soar in modelling human intelligence is good, so it has been assigned a score of 2.

- *Other strengths or weaknesses, Score* = 0. There seem to be no other notable strengths or weaknesses of Soar.
- *Combined Score* = 1 + 2 + 0 = 3.

4.5. ACT-R

The background to ACT-R is similar to that of Soar, and the evaluation here is similar.

- *Simplicity, Score* = 1. In keeping with what is said about Soar, ACT-R, and the SPTI, in Section 3.3, it seems best to evaluate the Simplicity of ACT-R in terms of the size of its important computational 'core'. Since that core expresses several aspects of intelligence via mechanisms for achieving those aspects of intelligence (next bullet point), and since, overall, there is little simplification via integration (as suggested by the structures shown in Figure 2), it seems reasonable to say that the Simplicity of ACT-R is moderate, and to assign it a score of 1.
- *Power, Score* = 2. ACT-R embodies a fairly detailed model of intelligence. It is different from the definition that is specified in Section 3.2 and is implicit in the structure of ACT-R (Section 2.5, second bullet point). However, as with Soar, there may be definitions of intelligence that are different from that in Section 3.2 but with equal validity. Overall, it seems reasonable to say that the Power of ACT-R in modelling human intelligence is good, and to assign it a score of 2.
- *Other strengths or weaknesses, Score* = 0. There seem to be no other notable strengths or weaknesses of ACT-R.
- *Combined Score* = 1 + 2 + 0 = 3.

4.6. NARS

Like DALL·E 2, NARS is being developed via a bottom-up strategy (Appendix E), and, so far, only two aspects of intelligence have been considered: Non-Axiomatic Logic and learning (Section 2.6). As will be described, this has a bearing on estimates of the Simplicity and Power of NARS.

- *Simplicity, Score* = 1. Although a mature version of NARS that embraces several aspects of intelligence is likely to be relatively large, it seems best to stick with the assessment of the Simplicity of NARS as it is now. Hence, although the Simplicity of NARS is merely because it is at an early stage of a bottom-up process, the Simplicity of NARS is assessed as moderate, and it has been assigned a score of 1.
- *Power, Score* = 1. Since NARS is still at an early stage in the integration of different aspects of intelligence, it seems reasonable to judge its descriptive/explanatory Power to be moderate, and to assign it a score of 1.
- *Other strengths or weaknesses, Score* = 0. There seem to be no other notable strengths or weaknesses of NARS.
- *Combined Score* = 1 + 1 + 0 = 2.

4.7. SPTI

- *Simplicity, Score* = 2. An important feature of the SPTI (including the SPCM) is that *all* the intelligence-related Power of the system, summarised in the next bullet point, flows from the computational 'core' of the SPCM, without the need for any kind of additional learning or programming.

  Viewed as a model of natural intelligence, all of the SPCM's intelligence-related 'core' may be seen as inborn capabilities, present at the system's 'birth'.

That computational core is remarkably simple: it is largely the SPMA concept (Appendix A.2) and the SPTI's procedures for unsupervised learning (Appendix A.4), much of which is the repeated application of the SPMA concept. In other words, the SPCM is largely the SPMA concept.

In short, the SPTI with the SPCM is remarkably small, meaning that its Simplicity is strong. Accordingly, it has been given the highest available score of 2.

A point that deserves emphasis here is that, in the SPTI, Simplicity may be combined with repetition of information, as described in Appendix C.3. In case this sounds like nonsense, the point is simply that the SPMA concept may be applied in several different aspects of intelligence: in hearing, in vision, in touch, and so on. That a powerful technique for compression of information may be applied in several different areas of brain function is a point that appears to have been missed in the ALTs described in this paper, and probably in other systems as well.

- *Power, Score* = 2. The intelligence-related capabilities of the SPTI (including the SPCM), which are substantial, are described in: Appendices B.1 and B.2.1, with indications of a few exceptions not yet demonstrated in the SPCM; and Appendix B.2.2.

  In short, the SPTI is strong in its Power to model aspects of human intelligence, so it has accordingly been assigned the highest available score of 2.

- *Other Strengths or Weaknesses, Score* = 1. Largely because of substantial evidence for the importance of IC in HLPC [20], IC is central in the structure and workings of the SPCM (Appendix C.5).

  In addition, a major discovery in the SP programme of research is the powerful concept of *SP-multiple-alignment* (Appendix A.2). This is largely responsible for the versatility of the SPCM across several aspects of intelligence, summarised above, and for the potential of the SPCM in other areas (Appendices B.2.3 and D).

  This versatility is itself due to the way in which the SPMA concept is a generalisation of six other methods for compression of information via ICMUP [27].

  Thus for two related reasons—the versatility of the SPMA in modelling human intelligence and beyond; and more generally the central role for IC in the SPTI—a score of 1 has been assigned to the SPTI under the heading 'other strengths or weaknesses'.

- *Combined Score* = 2 + 2 + 1 = 5.

5. Comparison of the SPTI with the ALTs

Table 1 summarises the evaluation scores assigned to the ALTs and the SPTI in Section 4.

Table 1. This table summarises the evaluation scores assigned to the ALTs and the SPTI in Section 4. "Other S/W" is short for "Other strengths or weaknesses".

System	Simplicity	Power	Other S/W	Combined Score
SOM	0	0	−1	0 + 0 − 1 = −1
Gato	1	2	−1	1 + 2 − 1 = 2
DALL·E 2	1	2	−1	1 + 2 − 1 = 2
Soar	1	2	0	1 + 2 + 0 = 3
ACT-R	1	2	0	1 + 2 + 0 = 3
NARS	1	1	0	1 + 1 + 0 = 2
SPTI	2	2	1	2 + 2 + 1 = 5

Assuming that the evaluations are fair, and care has been taken to ensure that they are, the SPTI, with a combined score of 5, stands well above the ALTs, the best of which—Soar and ACT-R—each have a combined score of 3.

The main reason for the relative strength of the SPTI is the concept of SP-multiple-alignment, which is largely responsible for the versatility of the SPCM, both within AI and beyond, and for its small size.

6. Conclusion

This paper argues that the SP Theory of Intelligence and its realisation as the SP Computer Model is a promising foundation for the development of human-like broad AI, at the level of humans or beyond, also known as Artificial General Intelligence (AGI).

In that connection, the main intelligence-related strengths of the SPTI, and other strengths, are summarised in Appendix B. It appears that the SPTI has significant advantages over six other systems, chosen to be representative of potential foundations for the development of AGI.

The Simplicity of the SPTI, and its Power in modelling aspects of human intelligence, is almost entirely due to the powerful concept of *SP-multiple-alignment* (Appendix A.2), itself part of the ICMUP approach to the achievement of information compression (Appendix C.2).

As noted in the Introduction, it would be a mistake for all available research eggs to be put into one basket. Given the uncertainties in any vision of how AGI might be achieved, it would make more sense to hedge our bets with parallel streams of research. The SPTI qualifies as the basis for one of those streams of research.

Again, as noted in the Introduction, achieving AGI is likely to be far into the future, but research that is aiming for AGI is likely to produce many potential benefits and applications at points along the road to AGI. Some of those potential benefits and applications from the SPTI are described in Appendix B.2.

Abbreviations

The following abbreviations are used in this manuscript:

- **AI** : *Artificial Intelligence*. Any aspect of intelligence, at any level, that is not natural.
- **AGI**: *Artificial General Intelligence*. AI which aims to encompass all aspects of intelligence at human levels or higher.
- **ALT**: *Alternative to the SPTI as a potential FDAGI*.
- **DNN**: *Deep Neural Network*.
- **GAP**: *Geometric Analogy Problem*.
- **HLPC**: *Human Learning, Perception, and Cognition*.
- **IC**: *Information Compression*.
- **ICMUP**: *Information Compression via the Matching and Unification of Patterns*.
- **FDAGI**: *Foundation for the Development of AGI*. A set of ideas intended to provide a firm basis for the development of AGI.
- **SOM**: *Society of Mind*. Section 2.1.

- **SPCM**: *SP Computer Model*. Appendix A.
- **SPTI**: *SP Theory of Intelligence*. Appendix A.

As noted in Appendix C.1, it is intended that 'SP' should be treated as a name, like 'IBM' or 'BBC', not an abbreviation.

Appendix A. The SP Theory of Intelligence and the SP Computer Model in Brief

This appendix introduces the *SP Theory of Intelligence* and its realisation in the *SP Computer Model* with sufficient detail to ensure that the rest of the paper is intelligible. More detail may be found in the paper [19], and there is a much fuller account of the system in the book *Unifying Computing and Cognition* [18].

The SPTI is conceived as a brain-like system as shown in Figure A1, 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* information (red), in the brain.

As described in more detail below, the processing of New information to create Old information is central in how the SPCM works and lies at the heart of the strengths of the SPTI, outlined in Appendix B.
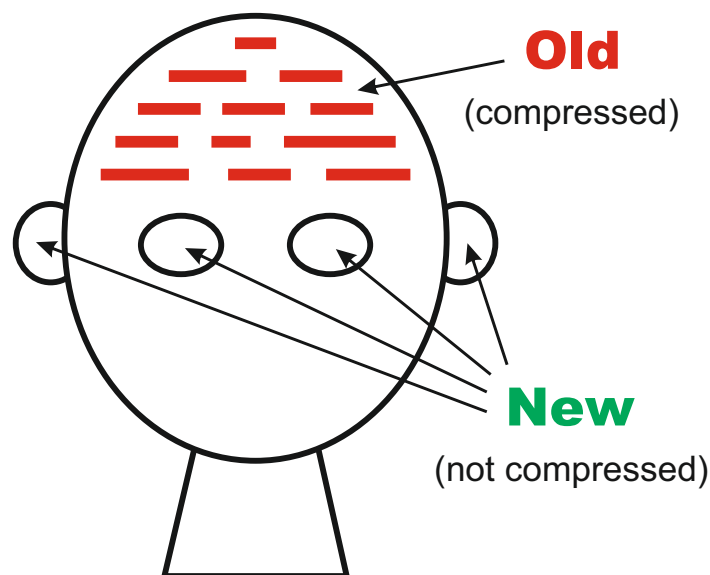
Figure A1. Schematic representation of the SPTI from an 'input' perspective. Reproduced, with permission, from Figure 1 in [19].

Appendix A.1. SP-Patterns and SP-Symbols

In the SPTI, all information is represented by *SP-patterns*, where an SP-pattern is array of *SP-symbols* in one or two dimensions. An SP-symbol is simply a mark from an alphabet of alternatives that can be matched in a yes/no manner with any other SP-symbol.

Examples of SP-patterns may be seen in Figure A2, as described in the caption to the figure.

At present, the SPCM works only with one-dimensional SP-patterns but it is envisaged that, at some stage, the SPCM will be generalised to work with two-dimensional SP-patterns as well as one-dimensional SP-patterns. This should open up the system for the representation and processing of diagrams and pictures, and, as described in [28] (Sections 6.1 and 6.2), structures in three dimensions.

Appendix A.2. The SP-Multiple-Alignment Concept

The concept of *SP-multiple-alignment* (SPMA) is described in outline here.

The SPMA concept is largely responsible for the strengths of the SPTI as summarised in Appendix B. Apart from some additional programming in the procedures for unsupervised learning (Appendix A.4), the SPMA concept is the means by which the SPCM achieves IC.

The SPMA concept in the SPCM has been borrowed and adapted from the concept of 'multiple sequence alignment' in bioinformatics [19] (Section 4). An example of an SPMA is shown in Figure A2.
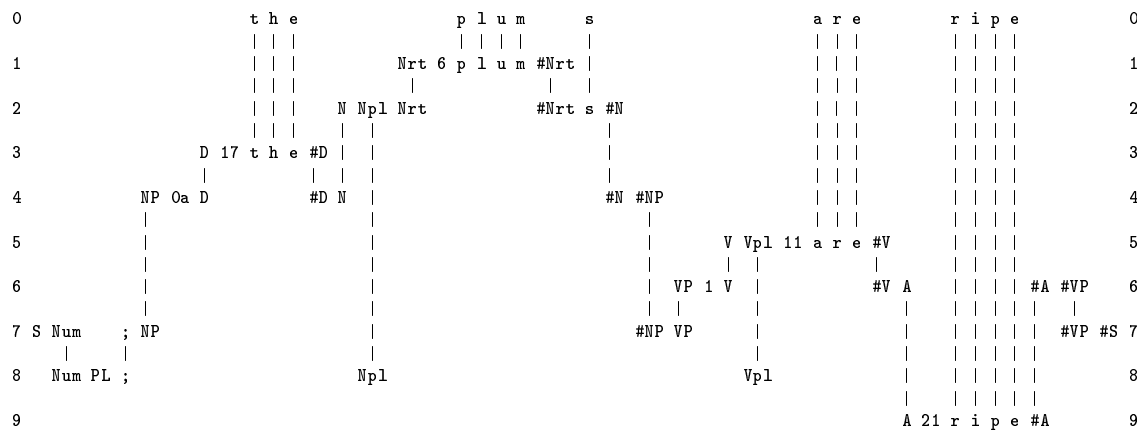
```
0                          t h e              p l u m      s                    a r e       r i p e          0
                           | | |              | | | |      |                    | | |       | | | |
1                          | | |       Nrt 6 p l u m #Nrt |                     | | |       | | | |          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  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 A2. 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 repository of Old SP-patterns.

Here is a summary of how SP-multiple-alignments like the one shown in Figure A2 are formed:

1. At the beginning of processing, the SPCM has a store of Old SP-patterns including those shown in rows 1 to 9 (one SP-pattern per row), and many others. When the SPCM is more fully developed, those Old SP-patterns would have been learned from raw data as outlined in Appendix A.4, but for now they are supplied to the program by the user.

2. The next step is to read in the New SP-pattern, 't h e p l u m s a r e r i p e'.

3. Then the program searches for 'good' matches between SP-patterns, where 'good' matches are ones that yield relatively high levels of compression of the New SP-pattern in terms of Old SP-patterns with which it has been unified.

4. As can be seen in the figure, matches are identified at early stages between (parts of) the New SP-pattern and (parts of) the Old SP-patterns 'D 17 t h e #D', 'Nrt 6 p l u m #Nrt', 'V Vpl 11 a r e #V', and 'A 21 r i p e #A'.

5. In SPMAs, IC is achieved by the merging or *unification* of SP-patterns, or parts of SP-patterns, that are the same, like the match between 't h e' in the New SP-pattern and the same three letters in the Old SP-pattern 'D 17 t h e #D'.

6. The unification of 't h e' with 'D 17 t h e #D' yields the unified SP-pattern 'D 17 t h e #D', with exactly the same sequence of SP-symbols as the second of the two SP-patterns from which it was derived.

7. The details of how IC for any one SPMA is calculated are given in [19] (Section 4.1) and [18] (Section 3.5).

8. As processing proceeds, similar pair-wise matches and unifications eventually lead to the creation of SP-multiple-alignments like that shown in Figure A2. At every stage, all the SP-multiple-alignments that have been created are evaluated in terms of IC, and then the best SP-multiple-alignments are retained and the remainder are discarded. In this case, the overall 'winner' is the SPMA shown in Figure A2.

9. This process of searching for good SP-multiple-alignments in stages, with selection of good partial solutions at each stage, is an example of heuristic search. This kind

of search is necessary because there are too many possibilities for anything useful to be achieved by exhaustive search. By contrast, heuristic search can normally deliver results that are reasonably good within a reasonable time, but it cannot guarantee that the best possible solution has been found.

Appendix A.3. Versatility of the SP-Multiple-Alignment Concept

As noted in the caption to Figure A2, the SPMA in the figure achieves the effect of parsing the sentence into its parts and sub-parts. However, *the beauty of the SPMA concept is that it is largely responsible for the strengths of the SPTI in several different areas*, summarised in Appendix B.

As noted in Appendix C.2, the SPMA concept is the last of seven variants of ICMUP described there, and it has been shown to be a generalisation of the other six variants [27]. This generalisation is probably the main reason for the strengths and potential of the SPTI mentioned above.

For readers not already familiar with the SPTI, it is appropriate to repeat what has been said elsewhere that, bearing in mind that it would be just as bad to downplay any feature of the SPTI as would over-selling any aspect of the system, *the SPMA concept promises to be as significant for our understanding of intelligence as is DNA for many aspects of biology. The SPMA concept may prove to be the 'double helix' of intelligence.*

Appendix A.4. Unsupervised Learning

"Unsupervised learning represents one of the most promising avenues for progress in AI. ... However, it is also one of the most difficult challenges facing the field. *A breakthrough that allowed machines to efficiently learn in a truly unsupervised way would likely be considered one of the biggest events in AI so far, and an important waypoint on the road to AGI.*" Martin Ford [29] (pp. 11–12), emphasis added.

As indicated near the beginning of Appendix A.2, above, unsupervised learning in the SPTI, described in the following two subsections, is intimately combined with processes of interpretation.

Appendix A.4.1. Learning with a *Tabula Rasa*

When the SPCM is a *tabula rasa*, with no stored Old SP-patterns, the system learns by taking in New SP-patterns via its 'senses' and storing them directly as received, except that 'ID' SP-symbols are added at the beginning and end, like the SP-symbols 'A', '21', and '#A', in the SP-pattern 'A 21 r i p e #A' in row 9 of Figure A2. Those added SP-symbols provide the means of identifying and classifying SP-patterns, and they may be modified or added to by later processing.

This kind of direct learning of new information reflects the way that people may learn from a single event or experience. One experience of getting burned may teach a child to take care with hot things, and the lesson may stay with him or her for life. However, we may remember quite incidental things from one experience that have no great significance in terms of pain or pleasure—such as a glimpse we may have had of a red squirrel climbing a tree.

Any or all of this one-shot learning may go into service immediately without the need for repetition, as for example: when we ask for directions in a place that we have not been to before; or how, in a discussion, we normally respond to what other people are saying.

These kinds of one-shot learning contrast sharply with learning in deep neural networks (DNNs) which requires large volumes of data and many repetitions before anything useful is learned.

"We can imagine systems that can learn by themselves without the need for huge volumes of labeled training data." Martin Ford [29] (p. 12).

"... the first time you train a convolutional network you train it with thousands, possibly even millions of images of various categories." Yann LeCun [29] (p. 124).

Appendix A.4.2. Learning with Previously Stored Knowledge

Of course, with people, the closest we come to learning as a *tabular rasa* is when we are babies. At all other times, learning occurs when we already have some knowledge. In people, and in the SPTI, two kinds of things can happen:

- The New information is interpreted via the SPMA in terms of the Old information, as described in Appendix A.2. The example illustrated in Figure A2 is of a purely syntactic analysis, but with the SPCM, semantic analysis is feasible too [18] (Section 5.7).
- Partial matches between New and Old SP-patterns may lead to the creation of additional Old SP-patterns, as outlined next.

As an example, Figure A3 shows an SPMA between New and Old SP-patterns.

```
0     t h e b l a c k c a t w a l k s   0
      | | |             | | | | | | | |
1 < 1 t h e       c a t w a l k s > 1
```

Figure A3. The best SPMA created by the SPCM with a New SP-pattern 't h e b l a c k c a t w a l k s' and an Old SP-pattern '< 1 t h e c a t w a l k s >'.

From a partial matching like this, the SPCM derives SP-patterns that reflect coherent sequences of matched and unmatched SP-symbols, and it stores the newly-created SP-patterns in its repository of Old SP-patterns, each SP-pattern with added 'ID' SP-symbols. The results in this case are the SP-patterns '< 13 t h e >', '< 19 b l a c k >', and '< 20 c a t w a l k s >'.

With this small amount of information, the SP-pattern '< 20 c a t w a l k s >' is a 'word'. However, with more information such as 'c a t r u n s' or 'd o g w a l k s', the SP-patterns 'c a t' and 'w a l k s' would become separate words.

Even with simple examples like these, there is a lot of complexity in the many alternative structures that the program considers. However, with the IC heuristic, the structures that are intuitively 'bad' are normally weeded out, leaving behind the structures that people regard intuitively as 'good'.

Appendix A.4.3. Unsupervised Learning of *SP-Grammars*

In the SPCM, processes like those just described provide the foundation for the unsupervised learning of *SP-grammars*, where an SP-grammar is simply a set of Old SP-patterns that is relatively good at compressing a given set of New SP-patterns.

To create 'good' SP-grammars requires step-wise processes of selection, very much like processes of that kind in the creation of 'good' SPMAs (Appendix A.2).

Appendix A.5. How to Make Generalisations without over- or under-Generalisation; and How to Minimise the Corrupting Effect of 'Dirty Data'

The importance of IC in HLPC (Appendix C.4) and in the SPTI (Appendix C.5) and, in particular, in unsupervised learning within the SPCM (Appendix A.4), provides what appears to be a sound solution to two problems with unsupervised learning: how to generalise beyond a body of data (**I**) without either over-generalisations (under-fitting) or under-generalisations (over-fitting); and how to learn correct forms despite the fact that **I** normally contains errors of various kinds, otherwise called 'dirty data'.

The proposed solution, indebted to Ray Solomonoff [30,31], is described in [19] (Section 5.3) and [18] (Section 2.2.12). In brief: compress **I** as thoroughly as possible via unsupervised learning to yield a grammar (**G**), and an encoding (**E**) of **I** in terms of **G**. Then discard **E** which contains all of the dirty data or most of it, and retain **G** which provides a compact description of **I**, including 'correct' generalisations from **I**.

Informal tests with unsupervised learning in the SPCM, and also in the MK10 and SNPR computer models of language learning [32], suggest that these principles are sound, including the exclusion of over- and under-generalisations, and the learning of 'correct' forms without corruption by 'dirty data'.

Appendix A.6. The SP Computer Model

The SP theory is realised most fully in the SP Computer Model, with capabilities in the building of multiple alignments and in unsupervised learning. The source code for the SPCM (the current version of which is 'SP71'), with a Windows executable file and some other files, may be obtained via sources detailed in 'Statements and Declarations/Software Availability', after the Conclusion (Section 6).

The SPCM and its precursors have played a key part in the development of the SPTI:

- As an antidote to vagueness. As with all computer programs, processes must be defined with sufficient detail to ensure that the program actually works.
- By providing a convenient means of encoding the simple but important mathematics that underpins the SP theory, and performing relevant calculations, including calculations of probability.
- By providing a means of seeing quickly the strengths and weaknesses of proposed mechanisms or processes. Many ideas that looked promising have been dropped as a result of this kind of testing.
- By providing a means of demonstrating what can be achieved with the theory.

The workings of the SPCM is described in some detail in [18] (Sections 3.9, 3.10, and 9.2) and more briefly in Appendices A.2 and A.4, above.

Appendix A.7. SP-Neural: A Preliminary Version of the SP Theory of Intelligence in Terms of Neurons and Their Interconnections and Inter-Communications

The SPTI has been developed primarily in terms of abstract concepts such as the SPMA concept. However, a version of the SPTI called SP-Neural has also been described in outline, expressed in terms of neurons and their inter-connections and inter-communications. Current thinking in that area is described in [18] (Chapter 11) and more recently in [33].

It appears that, within SP-Neural, the SP concepts of SP-pattern and SP-multiple-alignment can be expressed in terms of neurons and their interconnections. The main challenge is how the *processes* of building SP-multiple-alignments, and of unsupervised learning, can be expressed in terms of neural processes.

Appendix A.8. Neural Inhibition

In view of evidence for the importance of neural *inhibition* in the workings of brains and nervous systems and its role in IC in human learning, perception, and cognition [20], it seems likely that inhibition will be an important part of SP-Neural [33] (Section 9).

What appears to be a promising line of attack is *the idea that inhibition plays the part of unification in the ICMUP concept of IC (Appendix C.2)*:

- Unification in the ICMUP concept is when (within a body of information **I**) two or more patterns that match each other are reduced to a single instance.
- Providing that the patterns to be unified are more frequent within **I** than one would expect by chance, the merging of multiple instances to make one instance has the effect of removing redundancy from **I**.
- In a similar way, inhibition in the nervous system kicks in when two signals are the same. In lateral inhibition in the eye, for example, neighbouring fibres carrying incoming signals inhibit each other when they are both active.

As with the important role of the SPCM in the development of the abstract version of the SPTI (Section A.6), it seems likely that creating a computer model will be an effective way of developing SP-Neural, reducing vagueness in ideas, providing a means of testing ideas, and providing a means of demonstrating what can be done with the system when it is more mature.

Appendix A.9. Unfinished Business

Like most theories, the SP theory has shortcomings, but it appears that they may be overcome. At present, the most immediate problems are:

- *Processing of Information in Two or More Dimensions*. No attempt has yet been made to generalise the SP model to work with patterns in two dimensions, although that appears to be feasible to do, as outlined in [18] (Section 13.2.1). As noted in [18] (Section 13.2.2), it is possible that information with dimensions higher than two may be encoded in terms of patterns in one or two dimensions, somewhat in the manner of architects' drawings. A 3D structure may be stitched together from several partially-overlapping 2D views, in much the same way that, in digital photography, a panoramic view may be created from partially-overlapping pictures [28] (Sections 6.1 and 6.2).

- *Recognition of Perceptual Features in Speech and Visual Images*. For the SP system to be effective in the processing of speech or visual images, it seems likely that some kind of preliminary processing will be required to identify low level perceptual features such as, in the case of speech, phonemes, formant ratios, or formant transitions, or, in the case of visual images, edges, angles, colours, luminances, or textures. In vision, at least, it seems likely that the SP framework itself will prove relevant since edges may be seen as zones of non-redundant information between uniform areas containing more redundancy and, likewise, angles may be seen to provide significant information where straight edges, with more redundancy, come together [28] (Section 3). As a stop-gap solution, the preliminary processing may be done using existing techniques for the identification of low-level perceptual features [34] (Chapter 13).

- *Unsupervised Learning*. A limitation of the SP computer model as it is now is that it cannot learn intermediate levels of abstraction in grammars (e.g., phrases and clauses), and it cannot learn the kinds of discontinuous dependencies in natural language syntax that are described in [19] (Sections 8.1 and 8.2). I believe these problems are soluble and that solving them will greatly enhance the capabilities of the system for the unsupervised learning of structure in data [19] (Section 5.1).

- *Processing of Numbers*. The SP model works with atomic symbols such as ASCII characters or strings of characters with no intrinsic meaning. In itself, the SP system does not recognise the arithmetic meaning of numbers such as '37' or '652' and will not process them correctly. However, the system has the potential to handle mathematical concepts if it is supplied with patterns representing Peano's axioms or similar information [18] (Chapter 10). As a stop-gap solution, existing technologies may provide whatever arithmetic processing may be required.

In the process of solving these and other problems in the development of the SPTI, it seems likely that the proposed *SP Machine* (Appendix A.10, next) will be a useful vehicle for the representation and testing of ideas.

Appendix A.10. Future Developments and the SP Machine

In view of the potential of the SPTI in diverse areas (Appendix B), the SPCM appears to hold promise as the foundation for the development of an *SP Machine*, described in [35], and illustrated schematically in Figure A4.

It is envisaged that the SP Machine will feature high levels of parallel processing and a good user interface. It may serve as a vehicle for further development of the SPTI by researchers anywhere. Eventually, it should become a system with industrial strength that may be applied to the solution of many problems in science, government, commerce, industry, and in non-profit endeavours.

It is envisaged that the best way forward is to develop the *SP Machine* by porting the SPCM onto a platform which will provide for the application of high levels of parallel processing, and to adapt the CPCM to exploit those high levels of parallel processing. Additionally, there is a need to give the system a good 'friendly' user interface.

Although it is likely that a mature version of the SP Machine will be very much more efficient than the extraordinarily power-hungry and data-hungry DNNs [25] (Section 9), high levels of parallel processing are likely to be needed for relatively demanding operations such as unsupervised learning, especially with 'big data' and the like.
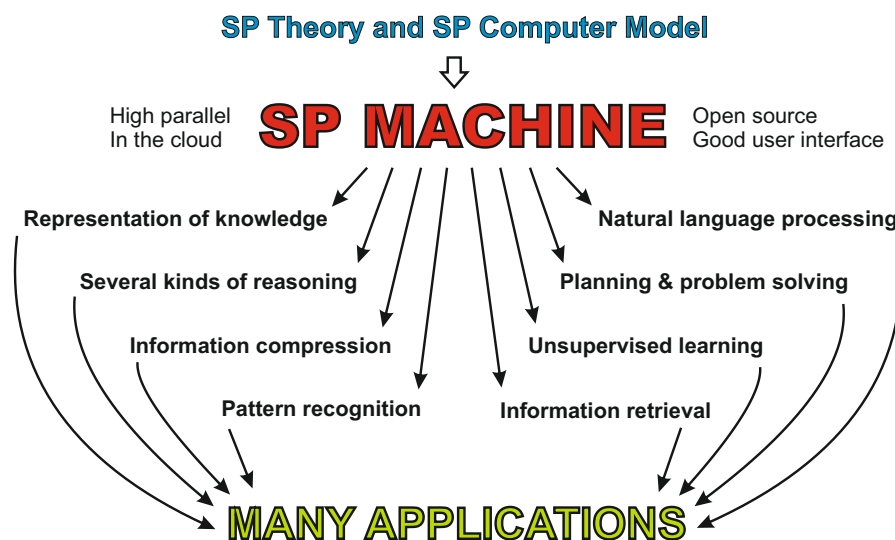

Figure A4. Schematic representation of the development and application of the SP Machine. Reproduced from Figure 2 in [19], with permission.

It is envisaged that the SP Machine will be entirely open so that researchers anywhere may test the system and develop it, perhaps following the suggestions in [35]. To make things easy for other researchers, the SP Machine may be hosted on one or more of the following platforms:

- A workstation with GPUs providing high levels of parallel processing. Other researchers would need to buy one or more such workstations, and then, on each machine, they may install the open-source software of the SPCM, ready for further development.
- Facilities in the cloud that provide for high levels of parallel processing.
- Since pattern-matching processes in the foundations of the SPCM are similar to the kinds of pattern matching that are fundamental in any good search engine, an interesting possibility is to create the SP Machine as an adjunct to one or more search engines. This would mean that, with search engines that are not open access, permission would be needed to access functions in relevant parts of the search engine, so that those functions may be used within the SP Machine.

Appendix B. Strengths of the SPTI

The strengths of the SPTI in intelligence-related functions and other attributes are summarised in this appendix. Further information may be found in [19] (Sections 5–12), in [18] (Chapters 5–9), on the CognitionResearch.org website (accessed on 22 November 2022), and in other sources referenced below.

Appendix B.1. Intelligence-Related Strengths of the SPCM

Most of the intelligence-related capabilities described in this appendix are demonstrable with the SPCM. Exceptions to that rule are noted at appropriate points below.

Apart from recording the intelligence-related strengths of the SPTI, this appendix also serves as the definition of intelligence expressed by the SPTI and outlined in Section 3.4.

Appendix B.1.1. Intelligence-Related Strengths Excluding Reasoning

Most of the aspects of intelligence described here have been demonstrated with the SPCM. In cases where there is merely potential and not actual demonstrations, this is indicated in the text.

- *Compression and Decompression of Information*. In view of substantial evidence for the importance of IC in HLPC [20], IC should be seen as an important feature of human intelligence. Paradoxical as this may seem, the SPCM provides for decompression of information via the compression of information (Appendix C.7).
- *Natural Language Processing*. Under the general heading of "Natural Language Processing" are capabilities that facilitate the learning and use of natural languages. These include:
  - The ability to structure syntactic and semantic knowledge into hierarchies of classes and sub-classes, and into parts and sub-parts.
  - The ability to integrate syntactic and semantic knowledge.
  - The ability to encode discontinuous dependencies in syntax such as the number dependency (singular or plural) between the subject of a sentence and its main verb, or gender dependencies (masculine or feminine) in French—where 'discontinuous' means that the dependencies can jump over arbitrarily large intervening structures. Also important in this connection is that different kinds of dependency (e.g., number and gender) can co-exist without interfering with each other.
  - The ability to accommodate recursive structures in syntax.
  - The production of natural language. A point of interest here is that the SPCM provides for the production of language as well as the analysis of language, and it uses exactly the same processes for IC in the two cases—in the same way that the SPCM uses exactly the same processes for both the compression and decompression of information (Appendix C.7).
- *Recognition and Retrieval*. Capabilities that facilitate recognition of entities or retrieval of information include:
  - The ability to recognise something or retrieve information on the strength of a good partial match between features as well as an exact match.
  - Recognition or retrieval within a class-inclusion hierarchy with 'inheritance' of attributes, and recognition or retrieval within an hierarchy of parts and sub-parts.
  - 'Semantic' kinds of information retrieval—retrieving information via 'meanings'.
  - Computer vision [28], including visual learning of 3D structures ([28] (Sections 6.1 and 6.2)
- *Several Kinds of Probabilistic Reasoning*. See Section B.1.2.
- *Planning and Problem Solving*. Capabilities here include:
  - The ability to plan a route, such as for example a flying route between cities A and B, given information about direct flights between pairs of cities including those that may be assembled into a route between A and B.
  - The ability to solve geometric analogy problems, or analogues in textual form.
- *Unsupervised Learning*. Chapter 9 of [18] describes how the SPCM may achieve unsupervised learning from a body of 'raw' data, **I**, to create an *SP-grammar*, **G**, and an **Encoding** of **I** in terms of **G**, where the encoding may be referred to as **E**. At present the learning process has shortcomings summarised in [19] (Section 3.3) but it appears that these problems may be overcome.

  In its essentials, unsupervised learning in the SPCM means searching for one or more 'good' SP-grammars, where a good SP-grammar is a set of SP-patterns which is relatively effective in the economical encoding of **I** via SP-multiple-alignment (Appendix C.1).

  This kind of learning includes the discovery of segmental structures in data (including hierarchies of segments and subsegments) and the learning classes (including hierarchies of classes and subclasses).

Appendix B.1.2. Probabilistic Reasoning

Capabilities here include:

- *One-Step 'Deductive' Reasoning*. A simple example of *modus ponens* syllogistic reasoning goes like this:

    – If something is a bird then it can fly.
    – Tweety is a bird.
    – Therefore, Tweety can fly.

- *Abductive Reasoning*. Abductive reasoning is more obviously probabilistic than deductive reasoning:

    "One morning you enter the kitchen to find a plate and cup on the table, with breadcrumbs and a pat of butter on it, and surrounded by a jar of jam, a pack of sugar, and an empty carton of milk. You conclude that one of your house-mates got up at night to make him- or herself a midnight snack and was too tired to clear the table. This, you think, best explains the scene you are facing. To be sure, it might be that someone burgled the house and took the time to have a bite while on the job, or a house-mate might have arranged the things on the table without having a midnight snack but just to make you believe that someone had a midnight snack. But these hypotheses strike you as providing much more contrived explanations of the data than the one you infer to." [36].

- *Probabilistic Networks and Trees*. One of the simplest kinds of system that supports reasoning in more than one step (as well as single step reasoning) is a 'decision network' or a 'decision tree'. In such a system, a path is traced through the network or tree from a start node to two or more alternative destination nodes depending on the answers to multiple-choice questions at intermediate nodes. Any such network or tree may be given a probabilistic dimension by attaching a value for probability or frequency to each of the alternative answers to questions at the intermediate nodes.

- *Reasoning With 'Rules'*. SP-patterns may serve very well within the SPCM for the expression of such probabilistic regularities as 'sunshine with broken glass may create fire', 'matches create fire', and the like. Alongside other information, rules like those may help determine one or more of the more likely scenarios leading to the burning down of a building, or a forest fire.

- *Nonmonotonic Reasoning*. The conclusion that "Socrates is mortal", deduced from "All humans are mortal" and "Socrates is human" remains true for all time, regardless of anything we learn later. By contrast, the inference that "Tweety can probably fly" from the propositions that "Most birds fly" and "Tweety is a bird" is nonmonotonic because it may be changed if, for example, we learn that Tweety is a penguin.

- *'Explaining Away'*. This means "If A implies B, C implies B, and B is true, then finding that C is true makes A less credible." In other words, finding a second explanation for an item of data makes the first explanation less credible.

There is also potential in the system for:

- *Spatial Reasoning*. The potential is described in [37] (Section IV-F.1).
- *What-If Reasoning*. The potential is described in [37] (Section IV-F.2).

Appendix B.1.3. The Representation and Processing of Several Kinds of
Intelligence-Related Knowledge

Although SP-patterns are not very expressive in themselves, they come to life in the SPMA framework within the SPCM. Within the SPMA framework, they provide relevant knowledge for each aspect of intelligence mentioned in Appendix B.1.

More specifically, they may serve in the representation and processing of such things as: the syntax of natural languages; class-inclusion hierarchies (with or without cross

classification); part-whole hierarchies; discrimination networks and trees; if-then rules; entity-relationship structures [38] (Sections 3 and 4); relational tuples ([38, Sections 3]) and concepts in mathematics, logic, and computing, such as 'function', 'variable', 'value', 'set', and 'type definition' ([18] (Chapter 10), [39] (Section 6.6.1), [40] (Section 2)).

As previously noted (Appendix A), the addition of two-dimensional SP patterns to the SPCM is likely to expand the capabilities of the SPTI to the representation and processing of structures in two-dimensions and three-dimensions, and the representation of procedural knowledge with parallel processing.

Appendix B.1.4. The Seamless Integration of Diverse Aspects of Intelligence, and Diverse Kinds of Knowledge, in Any Combination

An important additional feature of the SPCM, alongside its versatility in aspects of intelligence and diverse forms of reasoning, and its versatility in the representation and processing of diverse kinds of knowledge, is that *there is clear potential for the SPCM to provide for the seamless integration of diverse aspects of intelligence and diverse forms of knowledge, in any combination.* This is because those several aspects of intelligence and several kinds of knowledge all flow from a single coherent and relatively simple source: the SPMA framework.

It appears that this kind of seamless integration is *essential* in any artificial system that aspires to AGI.

Figure A5 shows schematically how the SPTI, with SPMA at centre stage, exhibits versatility and seamless integration.
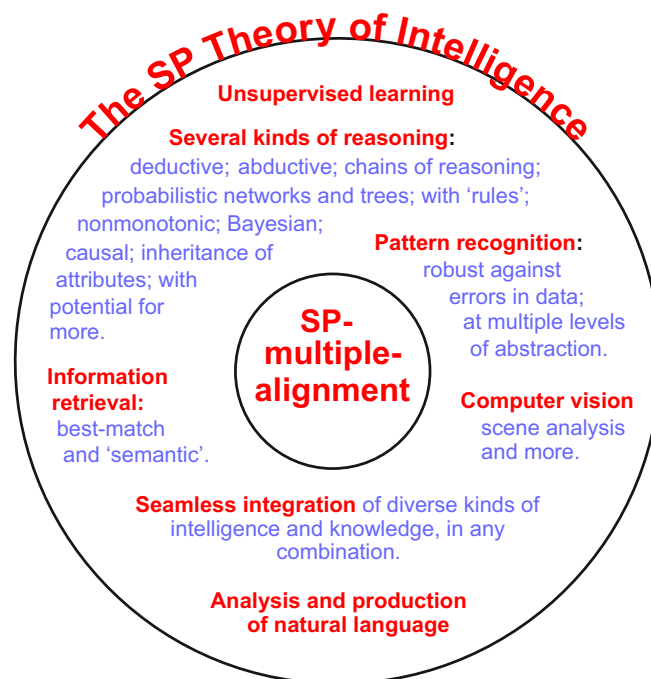

Figure A5. A schematic representation of versatility and seamless integration in the SPTI, with the SPMA concept centre stage.

Appendix B.2. Potential Benefits and Applications of the SPTI

Appendix B.2.1. Intelligence-related Benefits and Applications

- *Overview of Potential Benefits and Applications.* The paper [39] describes several potential benefits and applications of the SPTI, including some which are fairly directly related to AGI: best-match and semantic forms of information retrieval; the representation of knowledge, reasoning, and the semantic web.

- *The Development of Intelligence in Autonomous Robots*. The SPTI opens up a radically new approach to the development of intelligence in autonomous robots [37].
- *Commonsense Reasoning and Commonsense Knowledge*. Largely because of research by Ernest Davis and Gary Marcus (see, for example, [41]), the challenges in this area of AI research are now better known. Preliminary work shows that the SPTI has promise in this area [42].
- *An Intelligent Database System*. The SPTI has potential in the development of an intelligent database system with several advantages compared with traditional database systems [38].
- *Medical Diagnosis*. The SPTI 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 [43].
- *Natural Language Processing*. The SPTI has strengths in the processing of natural language ([19] (Section 8), [18] (Chapter 5)).
- *Vision, Both Artificial and Natural*. The SPTI opens up a new approach to the development of computer vision and its integration with other aspects of intelligence, and it throws light on several aspects of natural vision: [28,44].

Appendix B.2.2. The Clear Potential of the SPTI to Solve 20 Significant Problems in AI Research

Strong support for the SPTI has arisen, indirectly, from the book *Architects of Intelligence* by science writer Martin Ford [29]. To prepare for the book, he interviewed several influential experts in AI to hear their views about AI research, including opportunities and problems in the field:

> "The purpose of this book is to illuminate the field of artificial intelligence—as well as the opportunities and risks associated with it—by having a series of deep, wide-ranging conversations with some of the world's most prominent AI research scientists and entrepreneurs." Martin Ford [29] (p. 2).

In the book, Ford reports what the AI experts say, giving them the opportunity to correct errors he may have made so that the text is a reliable description of their thinking.

This source of information has proved to be very useful in defining problems in AI research that influential experts in AI deem to be significant. This has been important from the SP perspective because, with 17 of those problems and three others—20 in all—there is clear potential for the SPTI to provide a solution.

Since these are problems with broad significance, not micro-problems of little consequence, the clear potential of the SPTI to solve them is a major result from the SP programme of research, demonstrating some of the power of the SPTI.

The paper [25] describes those 20 significant and how the SPTI may solve them. The following summary describes each of the problems briefly. Readers are invited to read [25] to see how the SPTI may solve them:

1. *The Symbolic Versus Sub-Symbolic Divide*. The need to bridge the divide between symbolic and sub-symbolic kinds of knowledge and processing [25] (Section 3).
2. *Errors in Recognition*. The tendency of DNNs to make large and unexpected errors in recognition [25] (Section 3).
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 [25] (Section 5).
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 [25] (Section 6).
   It is clear that most human learning, including the learning of our first language or languages [32], is achieved via unsupervised learning, without needing tagged

examples, or reinforcement learning, or a 'teacher', or other form of assistance in learning (cf. [45]).

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, learning with rewards and punishments, and so on.

5. *Generalisation*. The need for a coherent account of generalisation, under-generalisation (over-fitting), and over-generalisation (under-fitting). Although this is not mentioned in Ford's book [29], there is the related problem of reducing or eliminating the corrupting effect of errors in the data which is the basis of learning [25] (Section 7).

6. *One-Shot Learning*. Unlike people, DNNs are ill-suited to the learning of usable knowledge from one exposure or experience [25] (Section 8).

7. *Transfer Learning*. Although transfer learning—incorporating old learning in newer learning—can be done to some extent with DNNs [46] (Section 2.1), DNNs fail to capture the fundamental importance of transfer learning for people, or the central importance of transfer learning in the SPCM [25] (Section 9).

8. *Reducing Computational Demands*. How to increase the speed of learning in AI systems, and how to reduce the demands of AI learning for large volumes of data, and for large computational resources [25] (Section 10).

9. *Transparency*. Although transfer learning—incorporating old learning in newer learning—can be done to some extent with DNNs [46] (Section 2.1), DNNs fail to capture the fundamental importance of transfer learning for people, or the central importance of transfer learning in the SPCM [25] (Section 9).

10. *Probabilistic Reasoning*. How to achieve probabilistic reasoning that integrates with other aspects of intelligence [25] (Section 12).

11. *Commonsense*. The challenges of commonsense reasoning and commonsense knowledge [25] (Section 13).

12. *Top-Down Strategies*. The need to re-balance research towards top-down strategies [25] (Section 14).

13. *Self-Driving Vehicles*. How to minimise the risk of accidents with self-driving vehicles [25] (Section 15).

14. *Compositionality*. By contrast with people, and the SPTI, DNNs are not well suited to the learning and representation of such compositional structures as part-whole hierarchies and class-inclusion hierarchies [25] (Section 16).

15. *Commonsense Reasoning and Commonsense Knowledge*. The challenges of commonsense reasoning and commonsense knowledge [25] (Section 17).

16. *Information Compression*. Establishing the key importance of IC in AI research [25] (Section 18). There is good evidence that much of HLPC may be understood as IC, and for that reason, IC is fundamental in the SPTI, including the SPCM (Appendix A, Appendix C.4). By contrast, IC receives no mention in [2], and does not receive much emphasis in Schmidhuber's review of DNNs (see, for example, [47] (e.g., Sections 4.4, 5.6.3 and 6.7)).

17. *A Biological Perspective*. Establishing the importance of a biological perspective in AI research [25] (Section 19).

18. *Distributed Versus Localist Knowledge*. Establishing whether or not knowledge in the brain is represented in 'distributed' or 'localist' form [25] (Section 20).

19. *Adaptation*. How to bypass the limited scope for adaptation in DNNs [25] (Section 21).

20. *Catastrophic Forgetting*. Catastrophic forgetting is the way in which, when a given DNN has learned one thing and then it learns something else, the new learning wipes out the earlier learning. This problem is quite different from human learning, where new learning normally builds on earlier learning, although of course we all have a tendency to forget some things.

    However, one may make a copy of a DNN that has already learned something, and then train it on some new concept that is related to what has already been learned.

The prior knowledge may help in the learning of the new concept. Additionally, one may provide a very large DNN, divided into sections, and train each section on a different concept [46].

Appendix B.2.3. Other Potential Benefits and Applications of the SPTI, with Less Relevance to Intelligence

This section describes other potential benefits and applications of the SPTI that are less closely related to AI. They include:

- *Overview of Potential Benefits and Applications*. As mentioned above, several potential areas of application of the SPTI are described in [39]. The ones that are less directly relevant to AI include: the simplification and integration of computing systems; software engineering; the representation of knowledge IC; bioinformatics; the detection of computer viruses; and data fusion.
- *Big Data*. The SPTI has potential in helping to solve several problem with big data [48]. These include: overcoming the problem of variety in big data; the unsupervised learning or discovery of 'natural' structures in data; the interpretation of data; the analysis of streaming data; making big data smaller; economies in the transmission of data; managing errors and uncertainties in data; visualisation of knowledge structures.
- *Sustainability*. The SPTI has potential for substantial reductions in the very large demands for energy of standard DNNs, and applications that need to manage huge quantities of data such as those produced by the Square Kilometre Array [49]. Where those demands are met by the burning of fossil fuels, there would be corresponding reductions in the emissions of CO_2.
- *Transparency in Computing*. By contrast with applications with DNNs, the SPTI provides a very full and detailed audit trail of all its processing, and all its knowledge is transparent and open to inspection. Additionally, there are reasons to believe that, when the system is more fully developed, its knowledge will normally be structured in forms that are familiar such as class-inclusion hierarchies, part-whole hierarchies, run-length coding, and more. Strengths of the SPTI in these area are described in [50].

Appendix C. Information Compression in Biology and the SPTI

As its title suggests, this appendix considers the role of IC in biology, especially in HLPC, and in the SPTI.

Appendix C.1. Information compression, Simplicity and Power

In words attributed to the English Franciscan friar William of Ockham: "Entities should not be multiplied beyond necessity". This principle, known as 'Ockham's razor', is commonly understood to mean that a good theory should be simple but not so simple that it says little or nothing that is useful.

- Any good theory may be seen as the product of a process that aims to simplify and integrate observations and concepts across a broad canvass (Appendix E), and this means applying IC to those observations and concepts.
- In all cases, IC may be seen as a process that increases the *Simplicity* of a body of information, **I**, by reducing or eliminating *redundancy* in **I**, whilst retaining as much as possible of the non-redundant descriptive and explanatory *Power* of **I**.
- For any one theory, it may be difficult or impossible to obtain precise values for Simplicity and Power. In cases like that, it may be necessary to use informal estimates.
- Since, for any one theory, the range of observations and concepts in **I** is likely to vary amongst alternative theories in the given area of interest, two or more theories of that area may be compared via some kind of combination of Simplicity, Power, and other strengths or weaknesses. In the example described in Section 3.1, simple measures of those attributes are simply added together.

- Care should be taken to ensure that the estimates of Simplicity, Power, and other strengths or weaknesses, are derived from a broad base of evidence (what Allen Newell called "a genuine slab of human behaviour," Appendix E), not some trivial corner of the given area of interest.
- Within this framework, two particularly weak kinds of theory may be recognised:
  - Any theory that is so general that, superficially, it can describe or explain anything (e.g., 'Because God wills it') should be rejected. In terms of Simplicity and Power, any such theory is weak because it is too simple and correspondingly lacking in Power.
  - Any theory that merely redescribes observations without any compression is a weak theory that should be rejected. In terms of Simplicity and Power, such a theory is weak because, without compression, the Simplicity of the theory is poor.

In this paper, a favourable combination of Simplicity and Power, or the potential for such a favourable combination, is what is mainly required in a system for it to qualify as an FDAGI. Of course, when the aim is to achieve AGI, Power must be the power of the system to described aspects of human intelligence.

Simplicity and Power are the reason for the name 'SP'. *However, as with such names as 'IBM' or 'BBC', it is intended that 'SP' should be used as a name and not as an abbreviation for Simplicity and Power*.

Appendix C.2. The Working Hypothesis That IC May Always Be Achieved Via the Matching and Unification of Patterns

A working hypothesis in the SP research is that *all kinds of IC may be achieved via ICMUP*. Although this is a 'working' hypothesis, there is much supporting evidence: the powerful concept of SPMA may be understood as an example of ICMUP [27]; the SPMA concept seems to underpin several aspects of intelligence (Appendix B.1), including several kinds of probabilistic reasoning; and much of mathematics, perhaps all of it, may be understood in terms of ICMUP (Appendix D and [51]).

In this research, seven main variants of ICMUP are recognised [51] (Sections 5.1–5.7):

- *Basic ICMUP*. Two or more instances of any pattern may be merged or 'unified' to make one instance [51] (Section 5.1).
- *Chunking-With-Codes*. Any pattern produced by the unification of two or more instances is termed a 'chunk'. A 'code' is a relatively short identifier for a unified chunk which may be used to represent the unified pattern in each of the locations of the original patterns [51] (Section 5.2).
- *Schema-Plus-Correction*. A 'schema' is a chunk that contains one or more 'corrections' to the schema. For example, a menu in a restaurant may be seen as a schema that may be 'corrected' by a choice of starter, a choice of main course, and a choice of pudding [51] (Section 5.3).
- *Run-Length Coding*. In run-length coding, a pattern that repeats two or more times in a sequence may be reduced to a single instance with some indication that it repeats, or perhaps with some indication of when it stops, or even more precisely, with the number of times that it repeats [51] (Section 5.4).
- *Class-Inclusion Hierarchies*. Each class in a hierarchy of classes represents a group of entities that have the same attributes. Each level in the hierarchy *inherits* all the attributes from all the classes, if any, that are above it [51] (Section 5.5).
- *Part-Whole Hierarchies*. A part-whole hierarchy is similar to a class-inclusion hierarchy but it is a hierarchy of part-whole groupings [51] (Section 5.6).
- *SP-multiple-alignment*. The SPMA concept is described in Appendix A.2 and in [51] (Section 5.7). *The SPMA concept may be seen as a generalisation of the other six variants of ICMUP*, as demonstrated via the SPCM in [27].

This list probably does not exhaust the possible variants of ICMUP, but they are the ones that have received most attention so far in the SP programme of research.

Appendix C.3. Clarification: Information Compression, Simplicity, and the Extraction of Redundancy From a Body of Information, **I**

A possible source of confusion is that, while the SPTI and SPCM are dedicated to the promotion of Simplicity in a body of information, **I**, and this means the extraction of *redundancy* from **I** via ICMUP, there may be good reasons why a simplified version of **I** may retain relatively large amounts of redundancy.

Those reasons do not include the fact that, with many instances of **I** that are reasonably large, it is difficult or impossible to find and extract all instances of redundancy in **I**.

Excluding that reason for residual redundancies in **I**, here are reasons why **I** may be seen to exhibit Simplicity and complexity at the same time:

- *Error-Reducing Redundancy*. Redundancy may be retained in **I**, or added to **I**, to help reduce errors in the processing or transmission of **I**. An obvious example is keeping two or more copies of a database to guard against catastrophic loss or corruption of the stored data. Normally, any one copy would be relatively free of redundancy but there may be two or more such copies.
- *Speeding Up Processing*. Databases that provide the basis for search engines would normally be replicated in different parts of the world. Each individual database would normally be relatively free of redundancy but the existence of multiple copies would normally reduce the computational load on each copy.
- *Multiplying the Advantages of Simplicity*. Although different parts of the brain may be seen to perform different functions, such as the processing of visual information in one part, the processing of auditory information in a second part, and the processing of tactile information in a third part, there may be advantages, as suggested by the SPTI, in exploiting the benefits of IC in all three parts, and elsewhere in the brain and nervous system.

In cases like that, the existence of redundancy in **I** should not disguise the simultaneous existence of Simplicity in **I**.

Appendix C.4. Evidence For The Importance of IC in HLPC, and Implications for AGI

A potent idea, pioneered by Fred Attneave [52,53], Horace Barlow [54,55], and others, is that much of the workings of brains and nervous systems may be understood as IC. This idea has been investigated by various researchers up to the present (see, for example, [56–58]). The importance of IC in HLPC became central in a programme of research developing computer models of the learning of a first language by children [32]. Evidence for the importance of IC in HLPC is reviewed in [20].

In connection with this research and the quest for AGI, it is of interest that, as far back as 1969, Barlow wrote:

> "... the operations needed to find a less redundant code 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, redundancy reduction may lead one towards understanding something about the organization of memory and intelligence, as well as pattern recognition and discrimination." [55] (p. 210).

where "find[ing] a less redundant code" leads to "redundancy reduction" which means IC.

Appendix C.5. IC and Its Role in the SPTI

With regard to goal of developing AGI:

- Evidence for the importance of IC in HLPC [20] has provided the motivation for making IC central in the structure and workings of the SPCM;
- In view of the same evidence, it seems clear that IC should be central in the workings of any system that aspires to AGI;

- The central role for IC in the SPCM—mediated by the concept of SPMA (Appendix A.2)—is largely responsible for:
    - The intelligence-related strengths of the SPTI (Appendix B.1);
    - The intelligence-related potential benefits and applications of the SPTI (Appendices B.1 and B.2.2).
    - Other potential benefits and applications of the SPTI, with less relevance to intelligence, as described in Appendix B.2.3.
    - In the formation of generalisations without over-fitting or under-fitting, and in the weeding out of 'dirty data', meaning data containing errors (Appendix A.5);
    - In a resolution of the apparent paradox that IC may achieve decompression as well as compression of data (Appendix C.7).

- In both natural and artificial systems:
    - For a given body of information, **I**, to be stored, IC means that a smaller store is needed. Or for a store of a given capacity, IC facilitates the storage of a larger **I** [20] (Section 4);
    - For a given body of information, **I**, to be transmitted along a given channel, IC means an increase in the speed of transmission. Or for the transmission of **I** at a given speed, IC means a reduction in the bandwidth which is needed [20] (Section 4).

- Because of the intimate relation between IC and concepts of inference and probability (Appendix C.6), and because of the central role of IC in the SPTI, the SPTI is intrinsically probabilistic.
  Correspondingly, it is relatively straightforward to calculate absolute and relative probabilities for all aspects of intelligence exhibited by the SPTI, including several kinds of reasoning, in keeping with the probabilistic nature of human inferences and reasoning.

Appendix C.6. IC and Concepts of Prediction and Probability

It has been known for some time that there is an intimate relation between IC and concepts of prediction and probability [30,31]. That intimate relation makes sense in terms of ICMUP because when two or more patterns are unified, the number of such patterns that are unified yields a measure of frequency, and that may translate into a measure of probability. Related ideas include:

- How absolute and relative probabilities are calculated for each SPMA is described in [19] (Section 4.4) and [18] (Section 3.7).
- The SPTI provides a framework for several kinds of probabilistic reasoning, as described in [18] (Chapter 7).
- A description of how the SPTI may provide an alternative to Bayesian networks to model the phenomenon of 'explaining away' may be found in [18] (Section 7.8).
- In mainstream statistics, it is normally assumed that high frequencies are needed to ensure statistical significance. However, in a search for repeating patterns that may be unified to yield compression of information, *the sizes of repeating patterns are as important as their frequency*. Then frequencies as low as 2 or 3 may yield inferences that are statistically significant [51] (Section 8.2.3).
- The close relation between IC and concepts of prediction and probability may suggest that, in developing any theory of AI or HLPC, it makes no difference whether we work from IC to probability or from probability to IC. However, there are several reasons to start with IC, as described in [51] (Section 8.2).

Appendix C.7. A Resolution of the Apparent Paradox That IC May Achieve Decompression as Well as Compression of Data

It is sometimes said that IC as a central feature of HLPC conflicts with the undoubted fact that people can and do produce information as well as compress it, both in ordinary speech or writing and also in creative areas like creative writing, painting, the composition of music, and so on.

In that connection, an interesting feature of the SPCM is that SPMA processes for the analysis of New information are *exactly* the same as may be used for the production of information. For example, with natural language, processes for the production of a sentence are, without any qualification, the same as may be used for the analysis of the same sentence.

Since the SPCM works by compressing information, this feature of the SPCM looks, paradoxically, like "decompression of information by compression of information".

How the whole system works, and how this paradox may be resolved, is explained in [19] (Section 4.5) and [18] (Section 3.8).

There is clear potential in the SPCM for the creation of entirely new structures which may be seen as novel or creative, but not necessarily artistic. This is an aspect of the SPTI that is waiting to be explored.

Appendix D. ICMUP Provides An Entirely Novel Perspective on the Foundations of Mathematics

In view of evidence for the importance of ICMUP in HLPC (Appendix C.2 and Appendix C.4), and in view of the fact that mathematics is the product of human brains and has been designed to help human thinking, it should not be surprising to find that ICMUP is central in the structure and workings of mathematics. This line of thinking is substantially different from any of the existing 'isms' in the foundations of mathematics, but there are potential connections with structuralism [51] (Section 4.4.4).

An important part of the arguments for the importance of ICMUP in mathematics are the versions of ICMUP outlined in Appendix C.2. For example:

- *Chunking-With-Codes*. The basic idea is that a relatively large body of information is given a relatively short identifier or 'code'. For example, with a function like '`sqrt(`x`)`', '`sqrt()`', is the code and the relatively large procedures for calculating square roots is the 'chunk'.
- *Schema-Plus-Correction*. A 'schema' is a chunk that contains one or more 'corrections' to the schema. Strictly speaking, the example just given for chunking-with-codes is an example of schema-plus-correction because the parameter x may be seen as a 'correction' that may apply a different input on different occasions. Perhaps a better example is a menu in a restaurant, where the menu and all the dishes it describes is the schema and 'corrections' to the schema are the dishes that any one diner may choose.
- *Run-Length-Coding*. This is where some entity, pattern, or operation is repeated two or more times in an unbroken sequence. Then it may be reduced to a single instance with some indication that it repeats. For example,

  - *Addition*. As an example of run-length-coding in arithmetic, an addition like $3 + 7$ may be seen as a shorthand for the addition of a single digit to the number 3, repeated 7 times.
  - *Multiplication*. Multiplication may also be seen as repeated addition, but at a higher level of abstraction. So, for example, a multiplication like 3×10 may be seen as the 10-fold repetition of the operation $x + 3$, where x starts with the value 0.

- And so on. Other examples of ICMUP in mathematics are described in [51] (Section 6.6).

Appendix E. The Benefits of a Top-Down, Breadth-First Research Strategy with Wide Scope

As its title suggests, this appendix is about how, in the quest for AGI, we may benefit from the adoption of a top-down, breadth-first research strategy with wide scope.

Appendix E.1. The Problem of Fragmentation in the Development of Theory

Allen Newell was one of the first people to draw attention to the problems of fragmentation in cognitive science in his famous paper "You can't play 20 questions with nature and win" [59]. In that paper, he exhorted researchers to tackle "a genuine slab of human behaviour" (p. 303), thus avoiding the weaknesses of micro-theories with limited scope for generalisation.

This thinking led to his book *Unified Theories of Cognition* [5] and a programme of research developing the Soar cognitive architecture [6], aiming for a unified theory of cognition.

This work chimes with Pamela McCorduck's description of fragmentation in AI:

"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." [60] (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." [60] (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." [60] (p. 424).

Although this was published in 2004, what McCorduck says is still true today. To a large extent, different aspects of AI are still developed independently of each other. Even when the overall goal is to develop AGI, it is commonly assumed that this may be approached via particular aspects of AI, and it appears that this 'bottom-up' strategy always fails (Appendix E.2).

Appendix E.2. The Seductive Plausibility of Bottom-Up Research Strategies, and Why They Fail

The reason that bottom-up research strategies are so attractive for researchers in AI seems to be that it allows researchers to concentrate on one aspect of intelligence at any one time. A bottom-up research strategy means that new papers can be produced quickly, thus helping researchers to cope with the over-strong requirement that they should 'publish or perish'.

In that connection but with reference to research in psychology, Newell wrote:

"Every time we find a new phenomenon—every time we find PI release, or marking, or linear search, or what-not—we produce a flurry of experiments to investigate it. We explore what it is a function of, and the combinational variations flow from our experimental laboratories. ... in general there are many more. Those phenomena form a veritable horn of plenty for our experimental life—the spiral of the horn itself growing all the while it pours forth the requirements for secondary experiments. ... Suppose that in the next thirty years we continued as we are now going. Another hundred phenomena, give or take a few dozen, will have been discovered and explored. ... Will psychology then have come of age? Will it provide the kind of encompassing of its subject matter—the behavior of man— that we all posit as a characteristic of a mature science? ... it seems to me that clarity is never achieved. Matters simply become muddier and muddier as we

go down through time. Thus, far from providing the rungs of a ladder by which psychology gradually climbs to clarity, this form of conceptual structure leads rather to an ever increasing pile of issues, which we weary of or become diverted from, but never really settle." [59] (pp. 2–7).

In the light of what Newell says, the reason that this kind of bottom-up strategy seems always to fail is that a theory that works in one local area rarely generalises to any other local area, or to any high-level view. Thus a persistent focus on low-level observations and concepts, with little or no attention to high-level concepts, makes it difficult or impossible to achieve simplification and integration at high levels of abstraction.

Appendix E.3. The Adoption of a Top-Down Research Strategy in the SP Research

The overarching goal of the SP research is to simplify and integrate observations and concepts in AI, mainstream computing, mathematics, and human learning, perception, and cognition.

In the quest for a general theory of those observations and concepts, the SPTI has been developed via a top-down, breadth-first research strategy with exceptionally wide scope. A clue was provided by the bioinformatics concept of 'multiple sequence alignment' which seemed to have the potential for the desired simplification and integration of concepts across a wide area.

As mentioned in Appendix A.2, the concept of multiple sequence alignment led to the development of the concept of *SP-multiple-alignment*. Despite its similarity with the concept of multiple sequence alignment, major programme of work was needed to develop the SP-multiple-alignment concept, including the creation and testing of *hundreds* of versions of the SPCM, to develop the new concept and to explore its range of potential applications.

The SP strategy should help to meet the concerns of Gary Marcus and Ernest Davis: "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." [61] (p. 15).

References

1. Minsky, M. (Ed.) *The Society of Mind*; Simon & Schuster: New York, NY, USA, 1986.
2. Reed, S.; Zolna, K.; Parisotto, E.; Colmenarejo, S.G.; Novikov, A.; Barth-Maron, G.; Gimenez, M.; Sulsky, Y.; Kay, J.; Springenberg, J.T.; et al. A generalist agent. *arXiv* **2022**. 1–40. arXiv:2205.06175v2.
3. Bostrom, N. *Superintelligence*, Kindle ed.; Oxford University Press: Oxford, UK, 2014.
4. Laird, J.E.; Rosenbloom, P.S.; Newell, A. Towards chunking as a general learning mechanism. In Proceedings of the Fourth National Conference on Artificial Intelligence, AAAI-1984, Austin, TX, USA, 6–10 August 1984, pp. 188–192.
5. Newell, A. (Ed.) *Unified Theories of Cognition*; Harvard University Press: Cambridge, MA, USA, 1990.
6. Laird, J.E. *The Soar Cognitive Architecture*; The MIT Press: Cambridge, MA, USA, 2012. ISBN: 978-0-262-12296-2.
7. Laird, J.E. *Introduction to the Soar Cognitive Architecture*; Technical Report; Center for Integrated Cognition, University of Michigan: Ann Arbor, MI, USA, 2022. arXiv:2205.03854v1.
8. Langley, P.; Laird, J.E.; Rogers, S. Cognitive architectures: Research issues and challenges. *Cogn. Syst. Res.* **2009**, *10*, 141–160.
9. Anderson, J.R.; Lebiere, C.J. *The Atomic Components of Thought*; Lawrence Erlbaum: Mahwah, NJ, USA, 1998.
10. Anderson, J.R.; Bothell, D.; Byrne, M.D. An integrated theory of the mind. *Psychol. Rev.* **2004**, *111*, 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036.
11. Wang, P. A unified model of reasoning and learning. *Proc. Mach. Learn. Res.* **2021**, *159*, 28–48.
12. Wang, P. A constructive explanation of consciousness. *J. Artif. Intell. Conscious.* **2020**, *7*, 257–275. https://doi.org/10.1142/S270507852050012
13. Wang, P. On defining artificial intelligence. *J. Artif. Gen. Intell.* **2019**, *10*, 1–37. https://doi.org/10.2478/jagi-2019-0002.
14. Wang, P. *Non-Axiomatic Logic: A Model of Intelligent Reasoning*, Kindle ed.; World Scientific Publishing Co. Pte. Ltd.: London, UK, 2013.
15. Popper, K.R. *The Logic of Scientific Discovery*, Kindle ed.; Published in 1935 as *Logik der Forschung*, and in an English translation in 1959; Routledge: London, UK, 2002.
16. Popper, K.R. *Conjectures and Refutations*, Kindle ed.; Basic Books: New York, NY, USA, 1962.
17. Legg, S.; Hutter, M. *A Collection of Definitions of Intelligence*; Technical Report, IDSIA: Lugano, Switzerland, 2007. Reference: IDSIA-07-07.
18. Wolff, J.G. *Unifying Computing and Cognition: The SP Theory and Its Applications*; CognitionResearch.org: Menai Bridge, UK, 2006. ISBN: 0-9550726-0-3.
19. Wolff, J.G. The SP Theory of Intelligence: An overview. *Information* **2013**, *4*, 283–341. https://doi.org/10.3390/info4030283.

20. Wolff, J.G. Information compression as a unifying principle in human learning, perception, and cognition. *Complexity* **2019**, *2019*, 38. https://doi.org/10.1155/2019/1879746.

21. Leivada, E.; Murphy, E.; Marcus, G. *DALL-E 2 Fails to Reliably Capture Common Syntactic Processes*; Technical Report; New York University: New York, NY, USA, 2022. arXiv:2210.12889v2.

22. Tamkin, A.; Brundage, M.; Clark, J.; Ganguli, D. *Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models*; Technical Report; Stanford University: Stanford, CA, USA, 2021. arXiv:2102.02503v1.

23. Else, H. 'Tortured phrases' give away fabricated research papers. *Nature* **2021**, *596*, 328–329. https://doi.org/10.1038/d41586-021-02134-0.

24. Cho, J.; Zala, A.; Bansal, M. *DALL-EVAL: Probing the Reasoning Skills and Social Biases of Text-to-Image Generative Models*; Technical Report; The University of North Carolina at Chapel Hill: Chapel Hill, NC, USA, 2022.

25. Wolff, J.G. Twenty significant problems in AI research, with potential solutions via the SP Theory of Intelligence and its realisation in the SP Computer Model. *Foundations* **2022**, *2*, 1045–1079. https://doi.org/10.3390/foundations2040070.

26. Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **1956**, *63*, 81–97.

27. Wolff, J.G. *Information Compression via the Matching and Unification of Patterns (ICMUP) as a Foundation for AI*; Technical Report; CognitionResearch.org: Menai Bridge, UK, 2021. Available online: tinyurl.com/2p9asr47 (accessed on 2 January 2023).

28. Wolff, J.G. Application of the SP Theory of Intelligence to the understanding of natural vision and the development of computer vision. *SpringerPlus* **2014**, *3*, 552–570. https://doi.org/10.1186/2193-1801-3-552.

29. Ford, M. *Architects of Intelligence: The Truth About AI from the People Building It*, Kindle ed.; Packt Publishing: Birmingham, UK, 2018.

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

31. Solomonoff, R.J. The discovery of algorithmic probability. *J. Comput. Syst. Sci.* **1997**, *55*, 73–88.

32. Wolff, J.G. Learning syntax and meanings through optimization and distributional analysis. In *Categories and Processes in Language Acquisition*; Levy, Y., Schlesinger, I.M., Braine, M.D.S., Eds.; Lawrence Erlbaum: Hillsdale, NJ, USA, 1988; pp. 179–215. Available online: tinyurl.com/4svmpdbf (accessed on 2 January 2023).

33. Wolff, J.G. Information compression, multiple alignment, and the representation and processing of knowledge in the brain. *Front. Psychol.* **2016**, *7*. https://doi.org/10.3389/fpsyg.2016.01584.

34. Prince, S.J.D. *Computer Vision: Models, Learning, and Inference*; Cambridge University Press: Cambridge, UK, 2012. ISBN 978-1-107-01179-3.

35. Palade, V.; Wolff, J.G. A roadmap for the development of the 'SP Machine' for artificial intelligence. *Comput. J.* **2019**, *62*, 1584–1604. https://doi.org/10.1093/comjnl/bxy126.

36. Douven, I. Abduction. In *Stanford Encyclopedia of Philosophy*; Zalta, E.N., Ed., Stanford University, 210 Panama Street, Stanford, CA, USA, 2021, https://plato.stanford.edu/archives/sum2021/entries/abduction/ (accessed on 2 January 2023).

37. Wolff, J.G. Autonomous robots and the SP Theory of Intelligence. *IEEE Access* **2014**, *2*, 1629–1651. https://doi.org/10.1109/ACCESS.2014.238

38. Wolff, J.G. Towards an intelligent database system founded on the SP Theory of Computing and Cognition. *Data Knowl. Eng.* **2007**, *60*, 596–624. https://doi.org/10.1016/j.datak.2006.04.003.

39. Wolff, J.G. The SP Theory of Intelligence: Benefits and applications. *Information* **2014**, *5*, 1–27. https://doi.org/10.3390/info5010001.

40. Wolff, J.G. *Software Engineering and the SP Theory of Intelligence*; Technical Report; CognitionResearch.org: Menai Bridge, UK, 2017. arXiv:1708.06665. Available online: bit.ly/2w99Wzq (accessed on 2 January 2023).

41. Davis, E.; Marcus, G. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Commun. ACM* **2015**, *58*, 92–103. https://doi.org/10.1145/2701413.

42. Wolff, J.G. *Commonsense Reasoning, Commonsense Knowledge, and the SP Theory of Intelligence*; Technical Report; CognitionResearch.org: Menai Bridge, UK, 2019. Available online: tinyurl.com/2rcxbu38 (accessed on 2 January 2023).

43. Wolff, J.G. Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, unification and search. *Decis. Support Syst.* **2006**, *42*, 608–625. https://doi.org/10.1016/j.dss.2005.02.005.

44. Wolff, J.G. The potential of the SP System in machine learning and data analysis for image processing. *Big Data Cogn. Comput.* **2021**, *5*, 7. https://doi.org/10.3390/bdcc5010007.

45. Gold, M. Language identification in the limit. *Inf. Control* **1967**, *10*, 447–474.

46. Schmidhuber, J. *One Big Net for Everything*; Technical Report; The Swiss AI Lab, IDSIA: Lugano, Switzerland, 2018.

47. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.00

48. Wolff, J.G. Big data and the SP Theory of Intelligence. *IEEE Access* **2014**, *2*, 301–315.

49. Wolff, J.G. How the SP System may promote sustainability in energy consumption in IT systems. *Sustainability* **2021**, *13*, 4565. https://doi.org/10.3390/su13084565.

50. Wolff, J.G. Transparency and granularity in the SP Theory of Intelligence and its realisation in the SP Computer Model. In *Interpretable Artificial Intelligence: A Perspective of Granular Computing*; Pedrycz, W., Chen, S.M., Eds.; Springer: Heidelberg, Germany, 2021. ISBN 978-3-030-64948-7.

51. Wolff, J.G. Mathematics as information compression via the matching and unification of patterns. *Complexity* **2019**, *2019*, 25.

52. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **1954**, *61*, 183–193.

53. Attneave, F. *Applications of Information Theory to Psychology*; Holt, Rinehart and Winston: New York, NY, USA, 1959.

54. Barlow, H.B. Sensory mechanisms, the reduction of redundancy, and intelligence. In *The Mechanisation of Thought Processes*; HMSO, Ed.; Her Majesty's Stationery Office: London, UK, 1959; pp. 535–559.

55. Barlow, H.B. Trigger features, adaptation and economy of impulses. In *Information Processes in the Nervous System*; Leibovic, K.N., Ed.; Springer: New York, NY, USA, 1969; pp. 209–230.

56. Chater, N. Reconciling simplicity and likelihood principles in perceptual organisation. *Psychol. Rev.* **1996**, *103*, 566–581.

57. Chater, N.; Vitányi, P. Simplicity: A unifying principle in cognitive science? *Trends Cogn. Sci.* **2003**, *7*, 19–22.

58. Hsu, A.S.; Chater, N.; Vitányi, P. Language learning from positive evidence, reconsidered: A simplicity-based approach. *Top. Cogn. Sci.* **2013**, *5*, 35–55.

59. Newell, A. You can't play 20 questions with nature and win: Projective comments on the papers in this symposium. In *Visual Information Processing*; Chase, W.G., Ed.; Academic Press: New York, NY, USA, 1973; pp. 283–308.

60. McCorduck, P. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, 2nd ed.; A. K. Peters Ltd.: Natick, MA, USA, 2004.

61. Marcus, G.F.; Davis, E. *Rebooting AI: Building Artificial Intelligence We Can Trust*, Kindle ed.; Pantheon Books: New York, NY, USA, 2019.