Article

How the SP System May Promote Sustainability in Energy Consumption in IT Systems

J. Gerard Wolff [ID]

CEng MBCS MIEEE, CognitionResearch.org, 18 Penlon, Menai Bridge, Anglesey LL59 5LR, UK; jgw@cognitionresearch.org; Tel.: +44-(0)-1248-712962 or +44-(0)-7746-290775

Abstract: The *SP System* (SPS), referring to the *SP Theory of Intelligence* and its realisation as the *SP Computer Model*, has the potential to reduce demands for energy from IT, especially in AI applications and in the processing of big data, in addition to reductions in CO_2 emissions when the energy comes from the burning of fossil fuels. The biological foundations of the SPS suggest that with further development, the SPS may approach the extraordinarily low (20 W)energy demands of the human brain. Some of these savings may arise in the SPS because, like people, the SPS may learn usable knowledge from a single exposure or experience. As a comparison, deep neural networks (DNNs) need many repetitions, with much consumption of energy, for the learning of one concept. Another potential saving with the SPS is that like people, it can incorporate old learning in new. This contrasts with DNNs where new learning wipes out old learning ('catastrophic forgetting'). Other ways in which the mature SPS is likely to prove relatively parsimonious in its demands for energy arise from the central role of information compression (IC) in the organisation and workings of the system: by making data smaller, there is less to process; because the efficiency of searching for matches between patterns can be improved by exploiting probabilities that arise from the intimate connection between IC and probabilities; and because, with SPS-derived 'Model-Based Codings' of data, there can be substantial reductions in the demand for energy in transmitting data from one place to another.

Keywords: information compression; *SP Theory of Intelligence*; *SP Computer Model*

1. Introduction

This paper describes the potential of the *SP System* (SPS), referring to the *SP Theory of Intelligence* and its realisation in the *SP Computer Model* (SPCM), for reductions in demands for energy, in comparison with current AI and IT systems. When the energy comes from the burning of fossil fuels, there is potential for corresponding reductions in emissions of CO_2. The origin of the name 'SP' is described in Section 1.1.2, below.

The SPS is introduced in Section 2, with more detail in Appendix A. Other sections describe aspects of the system with potential for substantial reductions in the energy required for AI and other IT processing.

There is some overlap of content with a previously published paper about big data [1], but the focus of this paper is on sustainability, not big data, and the audiences for the two papers are likely to be largely different.

1.1. Chasms to Be Bridged

At least two features of AI and other IT systems today raise concerns about energy consumption.

1.1.1. Energy Demands of Deep Neural Networks

The dominant paradigm in AI research today is 'deep neural networks' (DNNs), largely because of some impressive results that have been achieved with them. These include:

- Beating the world's best human players at the game of Go (See, for example: AlphaGo, https://tinyurl.com/yaxg9vz4; Mastering the game of Go with deep neural networks

and tree search, https://tinyurl.com/12ge3oht, all retrieved on 1 March 2021).
- Solving problems in the folding of proteins that are widely recognised to be very difficult. See, for example: AlphaFold: Using AI for scientific discovery, https://tinyurl.com/1bnwivkp, AI protein-folding algorithms solve structures faster than ever, https://tinyurl.com/3bt77zo4, all retrieved on 1 March 2021.
- Using DNNs for pattern recognition (see, for example [2,3]).

However, it is widely recognised that results like these have been achieved at a heavy cost in terms of energy and at a heavy cost in terms of CO_2 emissions when the energy comes from the burning of fossil fuels. For example, it has been discovered by Emma Strubell and colleagues [4] that the process of training a large AI model can emit more than 626,000 pounds of carbon dioxide, which is equivalent to nearly five times the lifetime emissions of the average American car, including the manufacture of the car itself.

This extraordinary figure for emissions of CO_2 from DNNs, and correspondingly large demands for energy, contrasts sharply with the way people can do similar things relatively fast and with a brain that runs on only about 20 watts of power [4–6].

In terms of sustainability, this chasm between the energy demands of DNNs and of the human brain represents a huge challenge, a challenge which is in urgent need of an answer in the now widely adopted quest to cut worldwide emissions of fossil carbon to zero by 2050, preferably sooner. Even if the energy to run AI systems all comes from renewable sources, we should not be unnecessarily wasteful.

This paper describes reasons for believing that the SPS has the potential to bridge this chasm and meet the challenge just described.

1.1.2. Communications and Big Data

In their book *Smart Machines* [7], John Kelly and Steve Hamm write:

> "The Square Kilometre Array is one of the most ambitious scientific projects ever undertaken. Its organizers plan on setting up a massive radio telescope made up of more than half a million antennas spread out across vast swaths of Australia and South Africa." John E. Kelly III and Steve Hamm ([7], p. 62).

> "The SKA is the ultimate big data challenge. ... The telescope will collect a veritable deluge of radio signals from outer space—amounting to fourteen exabytes of digital data per day ..." (ibid., p. 63).

These enormous quantities of data from theSKA will create huge problems in the management of data for even the smartest or most powerful of smart machines. One of the most surprising of those problems is that *the amount of energy required merely to move even a small part of these data from one place to another is proving to be a significant headache for the SKA project* ([7], p. 65, p. 92) *and other projects of that kind*. (However, see Sections 7.4 and 7.4.7).
More generally:

- "Communication networks face a potentially disastrous 'capacity crunch'." This quote summarises the conclusions of a meeting organised by the UK's Royal Society, introduced in [8], with other papers from the meeting at bit.ly/2fSy6qN, retrieved 1 March 2021.
- "Internet access may soon need to be rationed because the UK power grid and communications network can not cope with the demand from consumers." This quote is from a website of the telecomms company BT, bit.ly/2eUfMbS, retrieved 1 March 2021.

It is intended that 'SP' should be treated as a name, without any need to expand the letters in the name or explain the origin of the letters, as with names such as 'IBM' or 'BBC'. This because:

- The SPS is intended, in itself, to combine *Simplicity* with descriptive and explanatory *Power*.
- In addition, because the SPS works entirely by the compression of information, and this may be seen as a process that creates structures that combine conceptual *Simplicity* with descriptive and explanatory *Power*.

2. Introduction to the SPS

The SPS is the product of a lengthy programme of research, seeking to simplify and integrate observations and concepts across AI, mainstream computing, mathematics, and human learning, perception, and cognition.

An important unifying principle in the SPS is that all kinds of processing are achieved via IC, and IC is central in how knowledge in the system is organised. The main reason for this strong focus on IC is extensive evidence for the importance of IC in human learning, perception, and cognition [9]. Another reason which has emerged with the development and testing of the SPCM is that IC as it is incorporated in the SPCM provides for the modelling of diverse aspects of intelligence.

An important discovery from this research is the concept of *SP-multiple-alignment* (SPMA), a construct within the SPCM (Appendix A.3) which is largely responsible for: the versatility of the SPCM across aspects of AI, including diverse forms of reasoning (Appendices A.5.1 and A.5.2); the versatility of the SPCM in the representation of diverse forms of knowledge (Appendix A.5.3); and the seamless integration of diverse aspects of AI, and diverse forms of knowledge, in any combination (Appendix A.5.4).

In addition, it appears that the SPMA construct is largely responsible for the potential of the SPS for several benefits and applications (Appendix A.6).

Appendix A describes the SPS in outline, including descriptions of how the system achieves IC: via the matching and unification of patterns (Appendix A.2), via the building of SPMAs (Appendix A.3), and via the creation of 'SP-grammars' (Appendix A.4).

3. Biological Foundations

In connection with the potential in the SPS for reductions in AI-related demands for energy, it is relevant to say that in several dimensions, the system has foundations in biology, neurology, and psychology ("biological foundations" for short). As the SPS becomes progressively more mature, and especially in the development of SP-Neural (Appendix A.7), the biological foundations may be helpful in bringing energy demands of the SPS down towards the extraordinarily low 20 watts of the human brain.

Biological foundations of the SPS may be seen in the following areas:

- *Information compression*. The SP programme of research has, from the beginning, been tightly focussed on the importance of IC in the workings of brains and nervous systems, and how they organise knowledge (Section 7, [9]). Hence, IC is central in the workings of the SPS.
  The idea that IC might be significant in human perception and cognition was pioneered by Fred Attneave [10,11], Horace Barlow [12,13], and others, and has been a subject of research ever since.
- *Natural selection*. In human biology, and in the biology of non-human animals, it seems likely that IC would play a prominent role in natural selection because: (1) IC can speed up the transmission of a given body of information, **I**, in a given bandwidth; or it requires less bandwidth to transmit **I** at a given speed. (2) Likewise, IC can reduce the storage space required for a given body of information, **I**; or it can increase the amount of information that can be accommodated in a given store.
- *Research in language learning*. The SP programme of research is founded on earlier research developing computer models of language learning [14] and incorporates many insights from that research, especially the importance of IC in language learning.
- *Cell assembly and pattern assembly*. Although unsupervised learning in the SPS is entirely different from 'Hebbian' learning (See Appendix A.4), Donald Hebb's [15] concept of

a 'cell assembly' is quite similar to the SP-Neural concept of a *pattern assembly* (([16], Chapter 11), [17]).

- *Localist v distributed representation of knowledge.* The weight of evidence seems now to favour the 'localist' kind of knowledge representation adopted in the SPS and seems not to favour the 'distributed' style of knowledge representation adopted in DNNs ([18], pp. 461–463).
- *SP-Neural.* Although the SP-Neural version of the SPS (Appendix A.7) is still embryonic, there appears to be considerable potential for the expressions of IC via neural equivalents of ICMUP, SPMA, and SP-grammars, which are fundamental in the abstract version of the SPS.

4. One-Shot Learning

A striking difference between the SPS and DNNs is that the former, like people, are capable of learning usable knowledge from a single exposure or experience (see Section 8 in [19]) but a DNN needs many repetitions to learn any one concept well enough for it to enter into any other computation.

One-shot learning is illustrated schematically in Figure 1 (A). The yellow rectangle represents the system's memory store. The red disc at the top represents input after it has been compressed from the much larger size represented by the broken-line circle. The red disc below represents the same information after it has been stored. In terms of human cognition, this may be a little misleading because it is likely that some compression is done in sense organs and more is done on the way to storage and in the memory stores themselves.
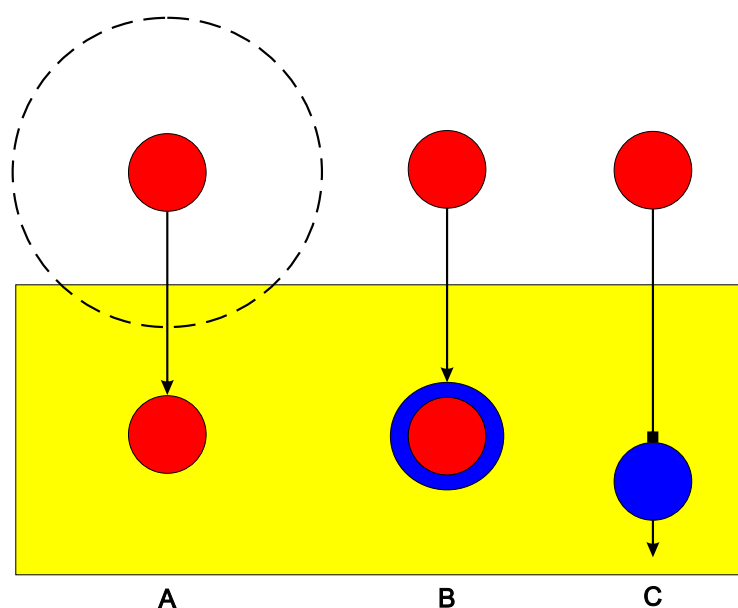

Figure 1. Schematic representation of one-shot learning (described in Section 4), transfer learning (Section 5), catastrophic forgetting (Section 6) and information compression (Section 7). How the figure illustrates those concepts is explained in each of those sections.

An example of one-shot learning is the way that one experience of touching something hot may be enough to teach a child to be careful with hot things, and that learning may persist for life.

Again, in any ordinary conversation between two people, each of them is absorbing what the other person is saying, and this is normally without the need for repetition.

When the SPS learns, the first step is to read in 'New' information from the environment in essentially the same way as a tape recorder. When it is stored, this information may enter directly in to varied kinds of analysis, as illustrated by the parsing example

in Figure A2 (Appendix A.3) without the need for any complicated preparation. This is quite different from DNNs where large-scale processing may be needed to learn each such 'simple' concepts as a buttercup, a dog, or a cow, and to bring each concept to a level of development where it may be used in other computations.

Clearly, this ability of the SPS to learn things directly without complex analysis is likely to promote parsimony in energy consumption compared with DNNs.

5. Transfer Learning

"Humans can learn from much less data because we engage in transfer learning, using learning from situations which may be fairly different from what we are trying to learn." Ray Kurzweil ([20], p. 230).

Anyone who knows how to play ice hockey is likely to adapt fairly easily to playing ordinary hockey. If you know how to ride a bicycle (two wheels) you will probably have little difficulty in riding a tricycle (three wheels). In general, we can, in 'transfer learning', use knowledge from one situation to help us learn things in other situations.

Transfer learning is illustrated schematically in Figure 1 (B), where incoming information shown in red is combined with already-stored information shown in blue. As before, the incoming information has been compressed but this is not shown in the figure.

The B figure is intended to show that in general, both the incoming New information, and the pre-existing Old information with which it is combined, would both be complete without any loss of information. This is illustrated in the example described in Appendix A.4.3 and introduced next.

Although the SPS can learn new things from scratch without any stored knowledge (Appendix A.4.2), transfer learning would normally account for the majority of learning in the SPS. A simple example is described in Appendix A.4.3. Here, an already-known Old SP-pattern, 'A 3 t h a t b o y r u n s #A', is matched with a New SP-pattern, 't h a t g i r l r u n s', with further processing outlined in that section.

In this case, the overall result is the creation of structures that give full weight to both the already-known Old SP-pattern, and the recent New SP-pattern, in accordance with the concept of transfer learning.

In terms of energy efficiency, transfer learning in both people and the SPS can mean substantial advantages because there would be little or no need to relearn things that have already been learned (Section 6). There should be corresponding savings in energy consumption.

Of course, people can and do forget things. This may be a simple defect in people compared with computers, or even records on paper, vellum, and so on. However, it may have a function in human cognition, as part of heuristic search, as described in Appendices A.3 and A.4.5.

6. Catastrophic Forgetting

An aspect of DNNs which has a bearing on energy consumption is the phenomenon called 'catastrophic forgetting'. This is because, in the standard model for DNNs, the learning of any concept wipes out any previous learning.

This is illustrated schematically in Figure 1 (C) where incoming information (shown in red) pushes out, or otherwise destroys, already-stored information (shown in blue). For the avoidance of any confusion, this is *not* a feature of the SPS but it *is* a feature of the standard DNN model.

This loss of stored information in DNNs is because learning is achieved by adjusting the weights of links between the layers of the DNN, and a set of weights which is right for one concept will not be right for any other concept. Of course one could have a new DNN for every new concept but, considering the number of concepts required in any realistic AI system, this solution is unlikely to be satisfactory.

To the extent that a new DNN for every concept might be regarded as practical, then considering the high energy demands of DNNs in learning one concept, the energy

demands of any kind of multi-DNN for the learning of many concepts would be off the scale. Since the SPS does not suffer from catastrophic forgetting and can learn many concepts without any interference among them, the SPS does not suffer from corresponding excesses in energy demands.

7. Information Compression

Aspects of how IC in the SPS may help in saving energy are discussed in the following subsections.

7.1. IC and Reducing the Size of Data

A rather obvious way in which IC might help reduce the computational demands of IT is in reducing the size of data. With regard to the difficulty of moving SKA-generated big data from one place to another (Section 1.1.2), if the size of the immovable bodies of data can be reduced, it may become feasible to move them, although it appears that this has not yet been examined in any detail.

Naturally, such a possibility will vary with the nature of the data. With the SKA, we may guess that levels of redundancy will be high, and that relatively high levels of compression may be achieved. In that case, IC may indeed be sufficient to overcome the problem of moving data from one place to another, and there would be corresponding savings in energy consumption. However, there are potentially much better answers described in the next three subsections.

7.2. IC and Probabilities

Another potential benefit of IC in the saving of energy, perhaps more surprising, arises from the intimate connection that exists between IC and concepts of inference and probability.

7.2.1. IC and Probabilities Are Two Sides of the Same Coin

These ideas were pioneered by Ray Solomonoff in the development of his *Algorithmic Probability Theory* (APT) [21,22]. It relates to *Algorithmic Information Theory* (AIT) where the information content of a body of data is the shortest length of an (idealised) computer program that anyone has been able to find which will generate those data [23]. Then, in APT, that shortest computer program provides the most probable hypothesis about the creation of the the original data.

The intimate connection that exists between IC and concepts of inference and probability makes sense in terms of the three aspects of IC described in Appendix A:

- *IC via the matching and unification of patterns)* (Appendix A.2). If two or more patterns match each other, it is not hard to see how, in subsequent processing, the beginnings of one pattern may lead to a prediction that the remainder is likely to follow. For example, if we know that 'black clouds rain' is a recurring pattern, then if we see 'black clouds', it is natural to predict that 'rain' is likely to follow.
- *IC via SP-multiple-alignment* (Appendix A.3). When the SPCM encounters a New SP-pattern like 't h e a p p l e s a r e s w e e t', it is likely to start building an SPMA like the one shown in Figure A2. As it proceeds, it is guided by the kinds of probabilistic inferences just mentioned. These apply at all and any level in the hierarchy of structures: at the level of words, at the level of phrases, and at the level of sentences.
- *IC via unsupervised learning* (Appendix A.4). Unsupervised learning in the SPCM creates one or more SP-grammars which are effective in the compression of a given set of New SP-patterns. It is this process which ensures that in IC via SPMA, the Old SP-patterns and the values for IC and probability accord with the DONSVIC principle (Appendix A.4.8).

7.2.2. Probabilities and Saving Energy

Probabilities can save energy by guiding the search for matching patterns to the areas where it can be most fruitful. As a simple example, if we are searching for strawberry jam, we are more likely to strike lucky, and to use less energy, if we search in a supermarket than if we search in a car sales showroom.

7.3. Processing in Two Stages

Another way in which IC can help to save energy may be achieved by splitting the IC process into two parts.

The general idea is to get some of the energy-demanding processing done in Phase 1, and then to carry on with less powerful computers that have been supplied with copies of the SP-grammar from Phase 1. Providing that the data that is processed in Phase 2 is substantially larger than the data for Phase 1, there is potential for an overall saving of energy.

In more detail:

1. *Phase 1: Create a grammar for data of a given type.*

    *   Choose a largish sample of data which is representative of the kinds of data to be processed.
    *   Process the sample via unsupervised learning within the SPCM to create one or two 'good' SP-grammars for those representative data. This stage is relatively demanding and may be done on a relatively powerful computer or SP Machine (Appendix A.8).

2. *Phase 2: Process one or more new streams of data. Overall, the data for Phase 2 should be very much larger than the data for Phase 1.*

    *   Use computers of relatively low power that have each been supplied with the SP-grammar from Phase 1.
    *   Providing each stream is not too large, it should be possible to achieve useful analyses of the data in terms of the SP-grammar.
    *   The analyses produced by this processing need not be simply parsings like the one shown in Figure A2. They may be any or all of the kinds of processing described in Appendix A.5.

7.4. Model-Based Coding

The problem of communication described in this section may be solved or at least reduced via a new approach to old ideas: 'analysis/synthesis' and, more specifically, the relatively challenging idea of 'Model-Based Coding'.

Analysis/synthesis has been described by Khalid Sayood like this:

> "Consider an image transmission system that works like this. At the transmitter, we have a person who examines the image to be transmitted and comes up with a description of the image. At the receiver, we have another person who then proceeds to create that image. For example, suppose the image we wish to transmit is a picture of a field of sunflowers. Instead of trying to send the picture, we simply send the words 'field of sunflowers'. The person at the receiver paints a picture of a field of sunflowers on a piece of paper and gives it to the user. Thus, an image of an object is transmitted from the transmitter to the receiver in a highly compressed form." Khalid Sayood ([24], p. 592).

This approach works best with the transmission of speech, probably because the physical structure and properties of the vocal cords, tongue, teeth, and so on, help in the process of creating an analysis of any given sample of speech and in any synthesis of speech that may be derived from that analysis. However, things are more difficult with images, especially if they are moving.

The more ambitious concept of Model-Based Coding was described by John Pierce in 1961 like this:

> "Imagine that we had at the receiver a sort of rubbery model of a human face. Or we might have a description of such a model stored in the memory of a huge electronic computer. First, the transmitter would have to look at the face to be transmitted and 'make up' the model at the receiver in shape and tint. The transmitter would also have to note the sources of light and reproduce these in intensity and direction at the receiver. Then, as the person before the transmitter talked, the transmitter would have to follow the movements of his eyes, lips and jaws, and other muscular movements and transmit these so that the model at the receiver could do likewise." John Pierce ([25], Location 2278).

At the time this was written, it would have been impossibly difficult to make things work as described. Pierce says:

> "Such a scheme might be very effective, and it could become an important invention if anyone could specify a useful way of carrying out the operations I have described. Alas, how much easier it is to say what one would like to do (whether it be making such an invention, composing Beethoven's tenth symphony, or painting a masterpiece on an assigned subject) than it is to do it." ([25], Locations 2278–2287).

Even today, Piece's vision is a major challenge. However, there appears to be a way forward via the development of the SPCM, described in the rest of this section. With some development of the SPCM, especially the generalisation of SP-patterns to accommodate information in two or three dimensions, the SPCM has potential to be very effective in the lossless transmission of big data and in lossless communications via the Internet.

7.4.1. Using an SP-Grammar for the Efficient Transmission of Data

In outline, Model-Based Coding may be made to work as shown in Figure 2. There would be two main elements to the scheme: (1) learning of an abstract description or SP-grammar ('**G**') for the kind of information to be transmitted; and (2) using **G** for the efficient transmission of information from A ('Alice') to B ('Bob').
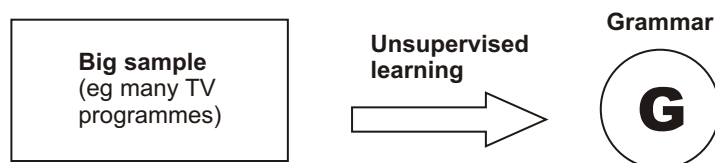
7.4.2. Unsupervised Learning of **G**

The learning in SP-based Model-Based Coding would be 'unsupervised', meaning learning directly from data without assistance of any kind of "teacher", or the labelling of examples, or rewards or punishments, or anything equivalent.

A strength of the SPS in this connection is that the SP programme of research grew out of earlier research on the *unsupervised* learning of language [14], and the entire SPCM revolves around learning that is *unsupervised*.

As in Section 7.3, learning would normally be done independently of any specific transmission, it would be done by a relatively powerful computer, and with a relatively large sample of the kind of data that is to be transmitted, such as a large collection of TV programmes.
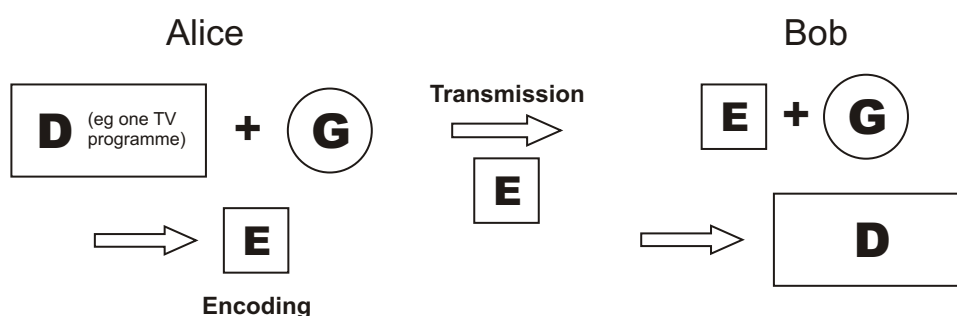
Unsupervised learning

Transmission

Figure 2. A schematic view of how, with Model-Based Coding, information may be transmitted efficiently from Alice to Bob.

7.4.3. Alice and Bob Both Receive Copies of **G**

Alice and Bob would each receive a copy of **G**. For example, **G** may be installed on every new computer, every new smartphone, and in every TV set, and it may also be made available for downloading.

In transmission of any one body of information ('**D**'), such as one TV programme, **D** would first be processed by Alice in conjunction with **G** to create an 'encoding' ('**E**') which would describe **D** in terms of the entities and abstract concepts in **G**. The encoding, **E**, would then be transmitted to Bob who would use it, in conjunction with his own copy of **G**, to reconstruct **D**. Provided that Alice and Bob have the same **G**, the version of **D** that is created by Bob should be exactly the same as the version of **D** that was transmitted by Alice, without loss of information.

7.4.4. **E** for Any Given **D** Would Normally Be Very Small Compared with **D**

Since **E** would normally be very small compared with **D**, there would, with one qualification, normally be a large saving in the amount of information to be transmitted compared with the transmission of raw data. Also, for reasons given below, it is likely that **E** would, with the same qualification, normally be very small compared with what would be transmitted using ordinary compression methods such as LZ, JPEG or MPEG, without the benefit of Model-Based Coding.

The qualification is that any given **G** would be used for the transmission of many different **D**s. If **G** is only used once or twice, any saving is likely to be relatively small.

7.4.5. Model-Based Coding Compared with Standard Compression Methods

The main differences between Model-Based Coding and alternative schemes using ordinary compression methods are these:

1. Any 'learning' with ordinary compression methods is part of the encoding stage, not an independent process.
2. Any such learning with ordinary compression methods is normally relatively unsophisticated and designed to favour speed of processing on low-powered computers rather than high levels of information compression.

3. In addition, if there is any 'learning' with ordinary compression methods, Alice transmits both **G** and **E** together, not **E** by itself. As we shall see, this is likely to mean much smaller savings than if **E** is transmitted alone.

4. In some versions of MPEG compression, Alice and Bob may be provided with some elements of **G**—such as the structure of human faces or bodies—but these are normally hard coded and not learned.

    Any learning in this case appears to be within a framework that lacks generality, is restricted to such things as faces or bodies, and is without the potential for unsupervised learning of a wide variety of entities and concepts (see, for example [26–28]).

In general, there are likely to be relatively large gains in efficiency in transmission with Model-Based Coding compared with transmission with ordinary methods for information compression. However, since the year 2000, few if any researchers have been conducting research on Model-Based Coding, perhaps because of the difficulties that John Pierce anticipated.

7.4.6. The Potential of the SPS for Model-Based Coding

To develop transmission of information via Model-Based Coding as outlined above, the SPS provides a promising way forward. This system has clear potential to provide the main functions that are needed: unsupervised learning of **G**; encoding of **D** in terms of **G** to create **E**; and lossless recreation of **D** from **E** and **G**.

If the SPS is being used by Alice as a means of transmitting information economically to Bob, then, with a previously learned **G** playing the part of Old knowledge and a given body of information (**D**) playing the part of New information, the encoding created by the SPS may play the part of **E** in the transmission of **D**, as described above.

Regarding the first of the functions mentioned above—unsupervised learning of **G**—the SP computer model has already demonstrated unsupervised learning of plausible generative grammars for the syntax of English-like artificial languages, including the learning of segmental structures, classes of structure, and abstract patterns ([16], Chapter 9)). With non-linguistic or 'semantic' forms of knowledge, the system has clear potential to learn such things as class hierarchies, class heterarchies (meaning class hierarchies with cross classification), part-whole hierarchies, and other forms of knowledge ([16], Section 9.5)).

7.4.7. Concluding Remarks about Model-Based Coding

Model-Based Coding has great potential to reduce the volumes of data that need to be transmitted in moving big data from one place to another or in communications via the Internet.

Instead of transmitting a 'grammar' and, at the same time, an 'encoding' of the data to be transmitted in terms of the grammar—which, with minor deviations, is what is needed with ordinary compression methods—it is only necessary, most of the time, to transmit only a relatively small encoding of the data. This advantage of Model-Based Coding arises from the fact that in contrast with the use of ordinary compression methods in the transmission of data, both Alice and Bob are equipped with a grammar for the kind of data that is to be transmitted.

Preliminary trials indicate that the volume of information to be transmitted with Model-Based Coding may be less than 6% of the volume of information to be transmitted with ordinary compression methods.

Until recently, it has not been feasible to convert John Pierce's vision into something that may be applied in practice. Now, with the development of the SPS, there is clear potential to realise the three main functions that will be needed: unsupervised learning of a grammar for the kind of data that is to be transmitted; the encoding of any one example of such data in terms of the grammar; and decoding of the encoding to retrieve the given example.

It appears now to be feasible to develop at least an approximation to these capabilities within the foreseeable future. By contrast with other work on Model-Based Coding,

unsupervised learning in the SPS has the potential to learn what will normally be the great diversity of entities and concepts that are implicit in the data.

With these developments, big data may glide quickly and efficiently from one place to another, without the need for massive bandwidth, and without needing the output of a small power station to haul it on its way. In addition, there may be less need to worry about possible shortages of bandwidth on the Internet or shortages of energy to power the Internet.

8. Conclusions

The *SP System* (SPS), meaning the *SP Theory of Intelligence* and its realisation in the *SP Computer Model*, is an AI system under development with potential to cut energy demands in IT systems such as those for deep neural networks (DNNs) and the management of big data. Likewise for CO_2 emissions when the energy comes from the burning of fossil fuels.

The biological foundations of the SPS suggest that with further development, the SPS may approach the extraordinarily low (20 watt) energy demands of the human brain.

Any such achievement would be partly because the SPS, like people, may learn usable knowledge from a single exposure or experience (one-shot learning). In comparison, deep neural networks (DNNs) need many repetitions for the learning of one concept.

Again, the SPS, like people, can incorporate old learning in new (transfer learning), in contrast to DNNs where new learning wipes out old learning ('catastrophic forgetting').

Other ways in which the mature SPS is likely to prove relatively parsimonious in its demands for energy arise from the central role of information compression (IC) in the organisation and workings of the system:

- *IC makes data smaller*, so there is less to process.
- *The close connection between IC and concepts of probability* mean that there are probabilities that can be exploited to improve the efficiency of searching for matches between patterns.
- *Model-Based Coding*, described by John Pierce in 1961, may become a reality with the development of an industrial-strength *SP-Machine*:
  - With a relatively powerful computer, create an *SP-grammar* from, for example, a collection of TV programmes.
  - Distribute the SP-grammar to TV transmitters and many computerised TV receivers.
  - For each programme to be transmitted, Alice first encodes it in terms of the SP-grammar, the relatively small encoding is then transmitted, and, finally, Bob decodes the encoding in terms of the SP-grammar to recreate the programme exactly.
  - This makes the greatest savings when the creation and distribution of the SP-grammar is relatively infrequent compared with the uses of the SP-grammar in the encoding and decoding of TV programmes and the like.

Taking a global view of the SPS, there is considerable potential in future versions for substantial economies in energy demands.

Abbreviations

The following abbreviations are used in this manuscript:

AI   Artificial Intelligence
DNN   Deep Neural Network
IC   Information Compression
ICMUP   Information Compression via the Matching and Unification of Patterns
SPCM   SP Computer Model
SPMA   SP-multiple-alignment
SPS   SP System

Appendix A. Outline of the SPS

This paper, like most other papers about the SPS, provides an appendix like this one, describing the SPS in outline. Although this means repetition across different papers, it is necessary to ensure that, for readers who are not already familiar with the SPS, each paper can be read without the need to consult any other publication.

The SPS 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 stored as *Old* information (red), stored in the brain.
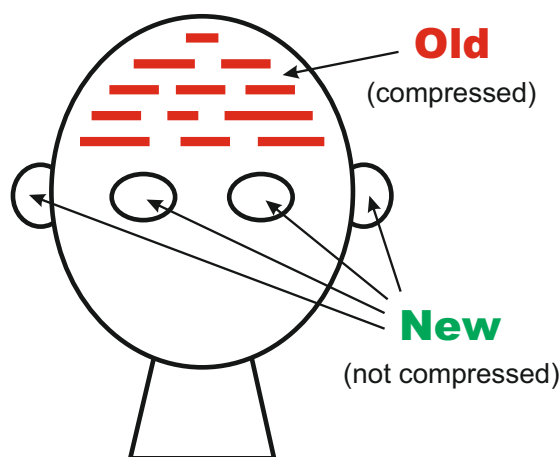

Figure A1. Schematic representation of the SPS from an 'input' perspective. Reproduced, with permission, from Figure 1 in [29].

As mentioned in Section 2, the SPS is the product of a lengthy programme of research, seeking to simplify and integrate observations and concepts across AI, mainstream computing, mathematics, and human learning, perception, and cognition.

In addition, an important discovery from this research is the concept of *SP-multiple-alignment* (SPMA), a construct within the SPCM (Appendix A.3) which is largely responsible for: the strengths and potential of the SPCM in aspects of AI (summarised in Appendix A.5) and the potential benefits and applications of the SPS (summarised in Appendix A.6).

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

Appendix A.1.1. SP-Patterns and SP-Symbols

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

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.

Appendix A.1.2. Redundancy

The *redundancy* in a body of information **I** means the amount of information within **I** which is repeated within **I**. In a typical **I**, much of that redundancy may be captured via the IC methods in the SPCM.

In discussions of IC, the concept of redundancy can be a useful way of referring to the information which is the main target for removal from **I**. Normally, it is not possible to remove all redundant information from **I**. In some applications other than the SPCM (which is designed for lossless IC), some non-redundant information is removed at the same time as redundant information.

In the SPS, IC is achieved in three main ways, described in the following three subsections. Each of these three methods except the first incorporates the previous method.

Appendix A.2. IC via the Matching and Unification of Patterns

This and the next two subsections discuss the three main methods in the SPS for compressing any given body of information **I**.

In contrast to techniques for IC such as Huffman coding ([24], Chapter 3) or wavelet compression ([24], Chapter 16), IC in the SPS is achieved largely by searching for patterns that match each other and merging or 'unifying' patterns that are the same. The expression "Information Compression via the Matching and Unification of Patterns" may be abbreviated as 'ICMUP'.

An important point in this connection is that there can be good matches between patterns that are not exactly the same. For example, there can be a successful match between two patterns such as 't h e b l a c k c a t w a l k s' and 't h e c a t w a l k s'. As described in Appendix A.4, this capability is important in the unsupervised learning of new structures.

ICMUP is often successful in finding much of the redundancy in **I**, but typically not all of it. More is normally encoded via the encoding methods associated with the building of SPMAs and SP-grammars, as described in subsections that follow.

Appendix A.3. IC via SP-Multiple-Alignment

A major part of the SPS is IC via the concept of *SP-multiple-alignment* (SPMA). This incorporates ICMUP but super-drives it by finding higher-level patterns as well. An example of an SPMA is shown in Figure A2.
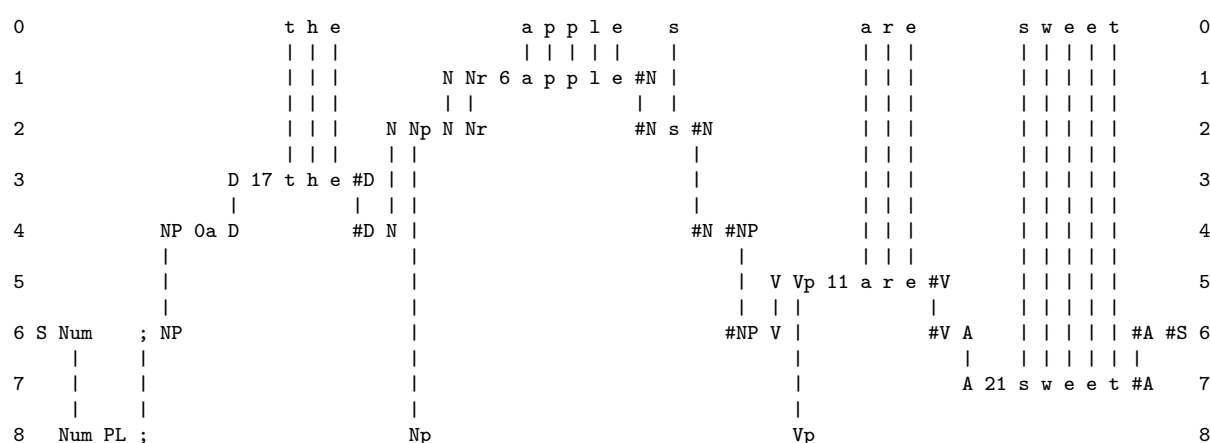
```
0                          t h e            a p p l e   s            a r e       s w e e t     0
                           | | |            | | | | |   |            | | |       | | | | |
1                          | | |      N Nr 6 a p p l e #N |           | | |       | | | | |     1
                           | | |          | |            |  |         | | |       | | | | |
2                          | | |    N Np N Nr            #N s #N      | | |       | | | | |     2
                           | | |        | |                 |        | | |       | | | | |
3                   D 17 t h e #D | |                        |        | | |       | | | | |     3
                        |      | |                           |        | | |       | | | | |
4          NP 0a D          #D N |                     #N #NP         | | |       | | | | |     4
            |                   |                       |            | | |       | | | | |
5           |                   |                       |  V Vp 11 a r e #V      | | | | |     5
            |                   |                       | | |            |       | | | | |
6 S Num   ; NP                  |                     #NP V |          #V A      | | | | | #A #S 6
     |    |                     |                           |            |       | | | | | |
7    |    |                     |                           |          A 21 s w e e t #A      7
     |    |                     |                           |
8   Num PL ;                    Np                         Vp                                  8
```

Figure A2. The best SPMA created by the SPCM that achieves the effect of parsing a sentence ('t h e a p p l e s a r e s w e e t'), as described in the text.

Here is a summary of how SPMAs like the one shown in the figure are formed:

1.　At the beginning of processing, the SPCM has a store of Old SP-patterns including those shown in rows 1 to 8 (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 a p p l e s a r e s w e e t'.

3.　Then the program searches for 'good' matches between SP-patterns, where 'good' means matches that yield relatively high levels of compression of the New SP-pattern in terms of Old SP-patterns with which it has been unified. The details of relevant calculations are given in (see ([29], Section 4.1) and ([16], Section 3.5)).

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', 'N Nr 6 a p p l e #N', 'V Vp 11 a r e #V', and 'A 21 s w e e t #A'.

5.　Each of these matches may be seen as a partial SPMA. For example, the match between 't h e' in the New SP-pattern and the Old SP-pattern 'D 17 t h e #D' may be seen as an SPMA between the SP-pattern in row 0 and the SP-pattern in row 3.

6.　After unification of the matching symbols, each such SPMA may be seen as a single SP-pattern. So 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.　As processing proceeds, similar pair-wise matches and unifications eventually lead to the creation of SPMAs like that shown in Figure A2. At every stage, all the SPMAs that have been created are evaluated in terms of IC (details of the coding are described in (see (Section 4.1 in [29]) and (Section 3.5 in [16])), and then the best SPMAs are retained and the remainder are discarded. In this case, the overall 'winner' is the SPMA shown in Figure A2.

8.　This process of searching for good SPMAs 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 much to be achieved via exhaustive search within a reasonable time. 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.

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 construct is that it can model many more aspects of intelligence besides the parsing of a sentence. These are summarised in Appendix A.5, although unsupervised learning (Appendix A.4) is a little different from the others.

Appendix A.4. IC via 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 ([20], pp. 11–12), emphasis added.

Appendix A.4.1. Unsupervised Learning in the SPS Is *Entirely Different* from 'Hebbian' Learning

Before we go on, it as well to stress that unsupervised learning in the SPS is *entirely different* from the well-known 'Hebbian' learning:

"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes

place in one or both cells such that A's efficiency, as one of the cells firing B, is increased." ([15], Location 1496).

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

In brief, the difference is that: (1) In Hebbian learning, all learning is slow; (2) In the SPS, some learning is relatively fast (eg one-shot learning (Section 4) and some transfer learning (Section 5)), but with complex knowledge and skills, SPS-learning is likely to be slow (Appendix A.4.6).

Appendix A.4.2. Learning with a *Tabula Rasa*

"... the first time you train a convolutional network you train it with thousands, possibly even millions of images of various categories." Yann LeCun ([20], p. 124).

"We can imagine systems that can learn by themselves without the need for huge volumes of labeled training data." Martin Ford ([20], p. 12).

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 s w e e t #A' in 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 (Section 4), 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 take account of what other people are saying.

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

Appendix A.4.3. 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.

At this stage, unsupervised learning in the SPS is intimately combined with processes of interpretation, as outlined in Appendix A.3. At this stage, the main possibilities are as follows:

Appendix A.4.4. No Match between New and Old SP-Patterns

If New information is received but there is no match with anything in the store of Old SP-patterns, then learning is much the same as in learning with a *tabula rasa* (Appendix A.4.2): the New information is simply added to the Old information, with the addition of ID SP-symbols.

Partial matches between New and Old SP-patterns

In the SPS, as with people, partial matches between New and Old SP-patterns may lead to the creation of additional Old SP-patterns, as outlined next.

Consider, first, an Old SP-pattern like this: 'A 3 t h a t b o y r u n s #A'. With a New SP-pattern like this: 't h a t g i r l r u n s', it is likely to form an SPMA as shown in Figure A3.

```
0        t h a t g i r l r u n s       0
         | | | |           | | | |
1    A 3 t h a t b o y   r u n s #A 1
```

Figure A3. The best SPMA created by the SPCM with a New SP-pattern 't h a t g i r l r u n s' and an Old SP-pattern 'A 3 t h a t b o y r u n s #A'.

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 best results in this case are SP-patterns for the words 'that', 'girl', 'runs', and 'boy', and a higher-level SP-pattern for 'that girl/boy runs'. This kind of process is a major part of unsupervised learning in the SPCM (Appendix A.4.5).

There are many other partial alignments that can be formed in cases like this, and much more complexity in the structures that may be formed. However, normally, most of these 'bad' structures are weeded out in the process of forming SP-grammars (next).

Appendix A.4.5. 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.

As with the creation of 'good' SPMAs (Appendix A.3), the development of 'good' SP-grammars requires heuristic search: a step-wise processes of selection through an abstract space of possibilities.

Appendix A.4.6. Speeds of Learning

When the abstract space mentioned in Appendix A.4.5 is small, learning can be fast. However, with complex knowledge and skills, such as the knowledge and skills required by a medical doctor, a top-level sportsperson, or a concert pianist, the complexity of the search space is likely to mean that learning will be slow, in accordance with human experience.

This contrasts with the speed of one-shot learning in people and in the SPS (Section 4 and Appendix A.4.3) and the speed with which transfer learning can occur in people and the SPS when the New knowledge to be learned is very similar to the Old knowledge with which it is to be combined.

Appendix A.4.7. Current and Future Developments

The SPCM as it is now can learn plausible grammars from examples of two-word and three-word English-like artificial languages without spaces or other markers between words. Its main weakness is that at present it cannot learn intermediate levels of structure such as phrases and clauses (see Section 3.3 in [29]). It appears that this problem is soluble and solving it will greatly enhance the capabilities of the system.

It is envisaged that similar principles may apply to the learning of non-syntactic 'semantic' structures, and to the integration of syntax with semantics. Such developments are likely to be facilitated by generality in the way in which the SPCM represents and processes all kinds of knowledge.

With generalisation of the concept of SP-pattern to include two-dimensional SP-patterns, there is potential for the SPCM to learn 3D structures, as described in (see Sections 6.1 and 6.2 in [30]).

Appendix A.4.8. The DONSVIC Principle

A general principle in the workings of the SPCM, and the earlier SNPR and MK10 computer models of language learning [14], is the *Discovery of Natural Structures via Information Compression* (DONSVIC) (see Section 5.2 in [29]), where 'natural' structures are things that people recognise naturally such as words and objects.

Why is this principle not more fully recognised in processes for IC such as LV, JPEG, and MPEG? It seems that this is probably because processes like LV etc have been designed to achieve speed on relatively low-powered computers, with corresponding sacrifices in IC efficiency. The DONSVIC principle is likely to become more important with computers that can achieve higher levels of efficiency in IC.

Appendix A.5. Strengths and Potential of the SPS in AI-Related Functions

The strengths and potential of the SPS are summarised in the subsections that follow. Further information may be found in Sections 5–12 in [29] and Chapters 5–9 in [16], and in [31] and other sources referenced in the subsections that follow).

Appendix A.5.1. Versatility in Aspects of Intelligence

As we have seen in Appendix A.4, the SPS has strengths and potential in the 'unsupervised' learning of new knowledge.

The SPS also has strengths and potential in other aspects of intelligence including: the analysis and production of natural language; pattern recognition that is robust in the face of errors in data; pattern recognition at multiple levels of abstraction; computer vision [30]; best-match and semantic kinds of information retrieval; several kinds of reasoning (next subsection); planning; and problem solving.

Appendix A.5.2. Versatility in Reasoning

An aspect of intelligence where the SPS has strengths and potential is probabilistic reasoning in several varieties including: one-step 'deductive' reasoning; chains of reasoning; abductive reasoning; reasoning with probabilistic networks and trees; reasoning with 'rules'; nonmonotonic reasoning and reasoning with default values; Bayesian reasoning with 'explaining away'; causal reasoning; reasoning that is not supported by evidence; the inheritance of attributes in class hierarchies; and inheritance of contexts in part-whole hierarchies.

There is also potential in the system for spatial reasoning (see Section IV-F.1 in [32], and for what-if reasoning Section IV-F.2 in [32]).

Appendix A.5.3. Versatility in the Representation of Knowledge

Within the framework of SPMA, SP-patterns may serve in the representation of several different kinds of knowledge, including: the syntax of natural languages; class-inclusion hierarchies; part-whole hierarchies; discrimination networks and trees; if-then rules; entity-relationship structures; relational tuples, and concepts in mathematics, logic, and computing, such as 'function', 'variable', 'value', 'set', and 'type definition'.

There will be more potential when the SPCM has been generalised for two-dimensional SP patterns.

Appendix A.5.4. Seamless Integration of Diverse Aspects of Intelligence, and Diverse Kinds of Knowledge, in Any Combination

Because diverse aspects of intelligence and diverse kinds of knowledge all flow from a single coherent and relatively simple framework—the SPMA framework—it is likely that the SPCM will support the seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination. It appears this is *essential* in any artificial system that aspires to the fluidity, versatility and adaptability of the human mind.

Figure A4 shows schematically how the SPS, with SPMA centre stage, exhibits versatility and integration across diverse aspects of intelligence.
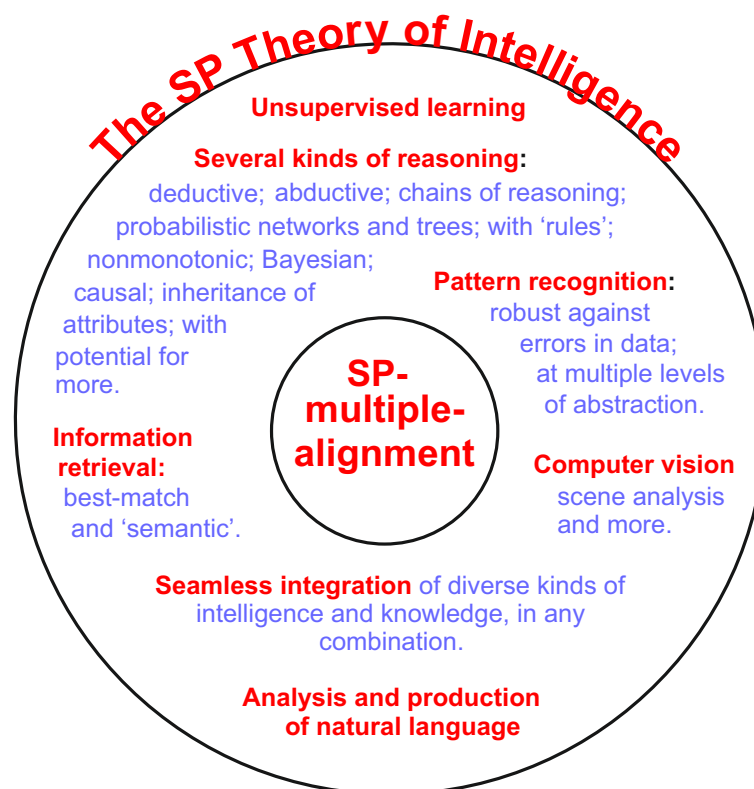
Figure A4. A schematic representation of versatility and integration in the SPS, with SPMA centre stage.

Appendix A.6. Potential Benefits and Applications of the SPS

Apart from its strengths and potential in modelling AI-related functions (Appendix A.5), it appears that in more humdrum terms, the SPS has several potential benefits and applications, several of them described in peer-reviewed papers. These include:

- *Big data*. Somewhat unexpectedly, it has been discovered that the SPS has potential to help solve nine significant problems associated with big data [1]. These are: overcoming the problem of variety in big data; the unsupervised learning of structures and relationships in big data; interpretation of big data via pattern recognition, natural language processing; the analysis of streaming data; compression of big data; Model-Based Coding for the efficient transmission of big data; potential gains in computational and energy efficiency in the analysis of big data; managing errors and uncertainties in data; and visualisation of structure in big data and providing an audit trail in the processing of big data.
- *Autonomous robots*. The SPS opens up a radically new approach to the development of intelligence in autonomous robots [32];
- *An intelligent database system*. The SPS has potential in the development of an intelligent database system with several advantages compared with traditional database systems [33]. In this connection, the SPS has potential to add several kinds of reasoning and other aspects of intelligence to the 'database' represented by the World Wide Web, especially if the SP Machine were to be supercharged by replacing the search mechanisms in the foundations of the SP Machine with the high-parallel search mechanisms of any of the leading search engines.
- *Medical diagnosis*. The SPS 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 [34];

- *Computer vision and natural vision*. The SPS opens up a new approach to the development of computer vision and its integration with other aspects of intelligence. It also throws light on several aspects of natural vision [30];
- *Neuroscience*. Abstract concepts in the SP Theory of Intelligence map quite well into concepts expressed in terms of neurons and their interconnections in a version of the theory called *SP-Neural* ([17], ([16], Chapter 11)). This has potential to illuminate aspects of neuroscience and to suggest new avenues for investigation.
- *Commonsense reasoning*. In addition to the previously described strengths of the SPS in several kinds of reasoning, the SPS has potential in the surprisingly challenging area of "commonsense reasoning and commonsense knowledge" [35]. How the SPS may meet the several challenges in this area is described in [36].
- *Other areas of application*. The SPS has potential in several other areas of application including ones described in [37]: the simplification and integration of computing systems; best-match and semantic forms of information retrieval; software engineering [38]; the representation of knowledge, reasoning, and the semantic web; information compression; bioinformatics; the detection of computer viruses; and data fusion.
- *Mathematics*. The concept of ICMUP provides an entirely novel interpretation of mathematics [39]. This interpretation is quite unlike anything described in existing writings about the philosophy of mathematics or its application in science. There are potential benefits in science and beyond from this new interpretation of mathematics.

Appendix A.7. SP-Neural

The SPS has been developed primarily in terms of abstract concepts such as the SPMA construct (Appendix A.3). However, a version of the SPS called SP-Neural was also proposed, expressed in terms of neurons and their inter-connections and inter-communications. Current thinking in that area is described in [17].

Two points of interest are noted here:

- *Neural validation*. Although SP-Neural is derived from the SPS, an abstract model of information processing, it maps quite well on to known features of neural tissue in the brain. This may be seen as a kind of neural validation of the abstract model from which it derives.
- *The role of inhibition in the brain*. In view of the importance of IC as a unifying principle in the SPS, and in view of the prevalence of inhibitory tissue in the brain, the known role of inhibitory neurons in some parts of the nervous system (([40], p. 505), [41]), suggests that inhibition could prove to be the key to understanding how IC may be achieved in SP-Neural, and hence in real brains.

Appendix A.8. Development of an SP Machine

In view of the strengths and potential of the SPS (Appendix A.5) and its potential benefits and applications (Appendix A.6), the SPCM appears to have promise as the foundation for the development of an *SP Machine*, as described in [42].

It is envisaged that the SP Machine wills feature high levels of parallel processing and a good user interface. It may serve as a vehicle for further development of the SPS by researchers anywhere. Eventually, it should become a system with industrial strength that may be applied to the solution of many problems in government, commerce, and industry. A schematic view of this development is shown in Figure A5.
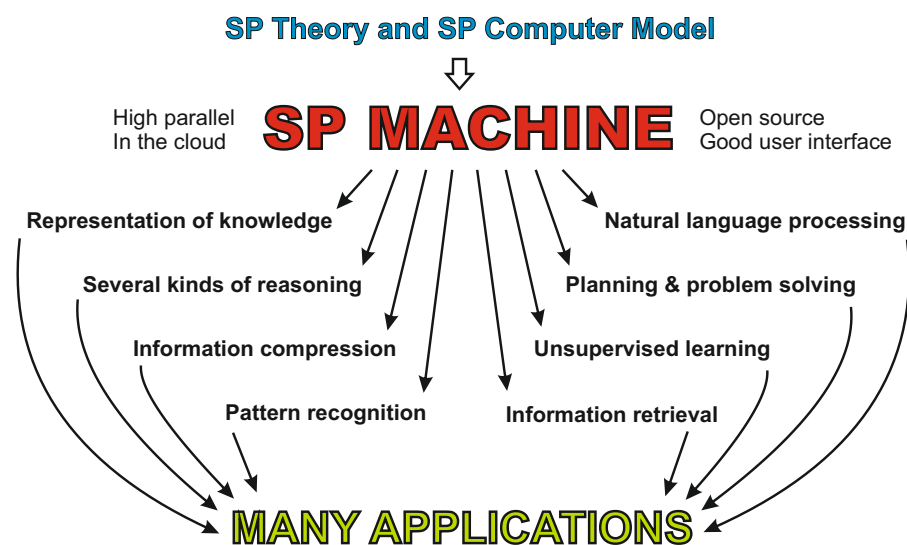
Figure A5. Schematic representation of the development and application of the SP Machine. Reproduced from Figure 2 in [29], with permission.

References

1. Wolff, J.G. Big data and the SP Theory of Intelligence. *IEEE Access* **2014**, *2*, 301–315. [CrossRef]
2. Guokun, L.; Chang, W.-C.; Yang, Y.; Liu, H. Modeling long- and short-term temporal patterns with deep neural networks. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18), Ann Arbor, MI, USA, 8–12 July 2018.
3. Yun, K.; Huyen, A.; Lu, T. Deep neural networks for pattern recognition. *arXiv* **2018**, arXiv:1809.09645.
4. Strubell, E.; Ganesh, A.; McCallum, A. Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), Florence, Italy, 28 July–2 August 2019.
5. Furber, S. To build a brain. *IEEE Spectr.* **2012**, *49*, 44–49. [CrossRef]
6. Jabr, F. Does Thinking Really Hard Burn More Calories? Available online: https://www.scientificamerican.com/article/thinking-hard-calories/ (accessed on 20 April 2021).
7. Kelly, J.E.; Hamm, S. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*; Columbia University Press: New York, NY, USA, 2013.
8. Ellis, A.D.; Suibhne, N.M.; Saad, D.; Payne, D.N. Communication networks beyond the capacity crunch. *Philos. Trans. R. Soc. A* **2015**, *374*, 191. [CrossRef]
9. Wolff, J.G. Information compression as a unifying principle in human learning, perception, and cognition. *Complexity* **2019**, *2019*. [CrossRef]
10. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **1954**, *61*, 183–193. [CrossRef] [PubMed]
11. Attneave, F. *Applications of Information Theory to Psychology*; Holt, Rinehart and Winston: New York, NY, USA, 1959.
12. Barlow, H.B. Sensory mechanisms, the reduction of redundancy, and intelligence. In *The Mechanisation of Thought Processes*; Her Majesty's Stationery Office: London, UK, 1959; pp. 535–559.
13. 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.
14. 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 Associates: Mahwah, NJ, USA, 1988; pp. 179–215.
15. Hebb, D.O. *The Organization of Behaviour*; John Wiley & Sons: New York, NY, USA, 1949.
16. Wolff, J.G. Unifying Computing and Cognition: The SP Theory and Its Applications. 2006. Available online: https://www.cognitionresearch.org/ (accessed on 19 April 2021).
17. Wolff, J.G. Neural Mechanisms for Information Compression by Multiple Alignment, Unification and Search. Technical Report; 2002. Available online: http://arXiv.org/abs/cs.AI/0307060 (accessed on 7 March 2021).
18. Page, M. Connectionist modelling in psychology: A localist manifesto. *Behav. Brain Sci.* **2000**, *23*, 443–512. [CrossRef] [PubMed]
19. Wolff, J.G. Problems in AI research and how the SP System may solve them. *arXiv* **2021**, arXiv:2009.09079v3.
20. Ford, M. *Architects of Intelligence: The Truth About AI from the People Building It*; Packt Publishing: Birmingham, UK, 2018.
21. Solomonoff, R.J. A formal theory of inductive inference. Parts I and II. *Inf. Control* **1964**, *7*, 1–22. 224–254. [CrossRef]
22. Solomonoff, R.J. The discovery of algorithmic probability. *J. Comput. Syst. Sci.* **1997**, *55*, 73–88. [CrossRef]
23. Li, M.; Vitányi, P. *An Introduction to Kolmogorov Complexity and Its Applications*, 4th ed.; Springer: New York, NY, USA, 2019.
24. Sayood, K. *Introduction to Data Compression*; Morgan Kaufmann: Amsterdam, The Netherlands, 2012.

25. Pierce, J.R. *Symbols, Signals and Noise*, 1st ed.; Harper & Brothers: New York, NY, USA, 1961.

26. Aizawa, K.; Harashima, H.; Saito, T. Model-based analysis synthesis image coding (mbasic) system for a person's face. *Signal Process. Image Commun.* **1989**, *1*, 139–152. [CrossRef]

27. Feng, W.U.; Peng, G.A.O.; Wen, G.A.O. Model-based coding. *Chin. J. Comput.* **1999**, *12*, 1239–1245.

28. Moghaddam, B.; Pentland, A. An automatic system for model-based coding of faces. In Proceedings of the IEEE Data Compression Conference (DCC '95), Snowbird, UT, USA, 28–30 March 1995; pp. 362–370.

29. Wolff, J.G. The SP Theory of Intelligence: An overview. *Information* **2013**, *4*, 283–341. [CrossRef]

30. 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. [CrossRef]

31. Wolff, J.G. The SP Theory of Intelligence: Its distinctive features and advantages. *IEEE Access* **2016**, *4*, 216–246. [CrossRef]

32. Wolff, J.G. Autonomous robots and the SP Theory of Intelligence. *IEEE Access* **2014**, *2*, 1629–1651. [CrossRef]

33. Wolff, J.G. Towards an intelligent database system founded on the SP theory of computing and cognition. *Data Knowl. Eng.* **2007**, *60*, 596–624. [CrossRef]

34. 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. [CrossRef]

35. Davis, E.; Marcus, G. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Commun. ACM* **2015**, *58*, 92–103. [CrossRef]

36. Wolff, J.G. Commonsense reasoning, commonsense knowledge, and the SP Theory of Intelligence. *arXiv* **2016**, arXiv:1609.07772.

37. Wolff, J.G. The SP Theory of Intelligence: Benefits and applications. *Information* **2014**, *5*, 1–27. [CrossRef]

38. Wolff, J.G. Software engineering and the SP Theory of Intelligence. *arXiv* **2017**, arXiv:1708.06665.

39. Wolff, J.G. Mathematics as information compression via the matching and unification of patterns. *Complexity* **2019**, *2019*. [CrossRef]

40. Squire, L.R.; Berg, D.; Bloom, F.E.; du Lac, S.; Ghosh, A.; Spitzer, N.C. (Eds.) *Fundamental Neuroscience*, 4th ed.; Elsevier: Amsterdam, The Netherlands, 2013.

41. Békésy, G. *Sensory Inhibition*; Princeton University Press: Princeton, NJ, USA, 1967.

42. Palade, V.; Wolff, J.G. A roadmap for the development of the 'SP Machine' for artificial intelligence. *Comput. J.* **2019**, *62*, 1584–1604. [CrossRef]