Multiple alignment could be the “double helix” of intelligence

The SP theory of intelligence promises to provide the human-like intelligence that is missing from the universal Turing machine

Gerry Wolff

It is true, as a White-House-sponsored conference has said, that AI is still far from matching the flexibility and learning capability of the human mind.[1] But there is now substantial evidence that a concept of *multiple alignment*, borrowed and adapted from research in biochemistry, can provide a productive way forward. This concept has the potential to be as significant for an understanding of “intelligence” in a broad sense as is DNA for biological sciences.

For several years, I have been working to develop a new theory of computing—a successor to Alan Turing’s “universal Turing machine”—that would provide the human-like intelligence which, as Turing recognised, is missing from his famous abstract machine, with its endless tape and read-write head. The result of this work is the *SP theory of intelligence*, with *multiple alignment* at centre stage.

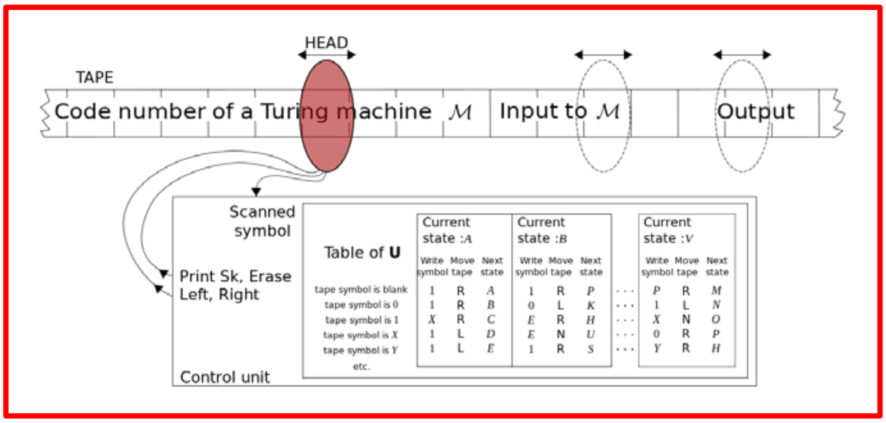


Table of **U**	Current state :*A*			Current state :*B*				Current state :*V*		
	Write symbol	Move tape	Next state	Write symbol	Move tape	Next state		Write symbol	Move tape	Next state
tape symbol is blank	1	R	*A*	1	R	*P*	. . .	*P*	R	*M*
tape symbol is 0	1	R	*B*	0	L	*K*	. . .	1	L	*N*
tape symbol is 1	*X*	R	*C*	*E*	R	*H*	. . .	*X*	N	*O*
tape symbol is *X*	1	L	*D*	*E*	N	*U*	. . .	0	R	*P*
tape symbol is *Y*	1	L	*E*	1	R	*S*	. . .	*Y*	R	*H*
etc.										

The SP computer model—the form in which the theory has been developed and tested—would not (yet) pass the “Turing test” by fooling people into thinking it is a person if they were to converse with it via typed messages. But it exhibits a lot of the versatility and adaptability of human intelligence, all flowing largely from the way in which the multiple alignment framework can accommodate diverse kinds of knowledge and perform diverse kinds of processing.

Since people often ask what “SP” means or stands for, it is really just a name, adopted because the SP system works by compressing information, and compression of information may be seen as a process of making a body of information as *Simple* as possible but retaining as much as possible of its descriptive *Power*. The name also connects with Occam’s Razor—the idea that scientific theories should combine simplicity with descriptive or explanatory power.

---

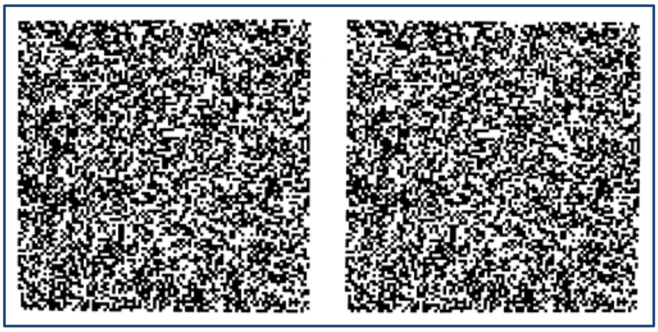
[1] See “Artificial intelligence is far from matching humans, panel says”, *New York Times*, 2016-05-25, nyti.ms/1TQ1xpb.

Contact: Dr Gerry Wolff PhD CEng MBCS MIEEE, CognitionResearch.org, Menai Bridge, UK; *Phone:* +44 (0) 1248 712962; +44 (0) 7746 290775; *Skype:* gerry.wolff; *E-mail:* **jgw@cognitionresearch.org**; *Web:* **www.cognitionresearch.org**.

My personal journey with these ideas began when I was a student at Cambridge University in England, attending lectures by Dr Horace Barlow, who later became a professor and Fellow of the Royal Society. What really caught my attention was the idea—founded on Claude Shannon's "theory of communication", later called "information theory"—that much of the workings of brains and nervous systems may be understood as compression of information.

A simple example, mentioned by Barlow, is how, when we view the world with two eyes, we are seeing two almost-identical views, but we merge them into a single view, thus compressing the information.

An illustration of how, with two eyes, we merge the two views is the way in which we interpret a "random-dot stereogram" of the kind developed by Béla Julesz at Bell Labs, an example of which is shown in the figure.

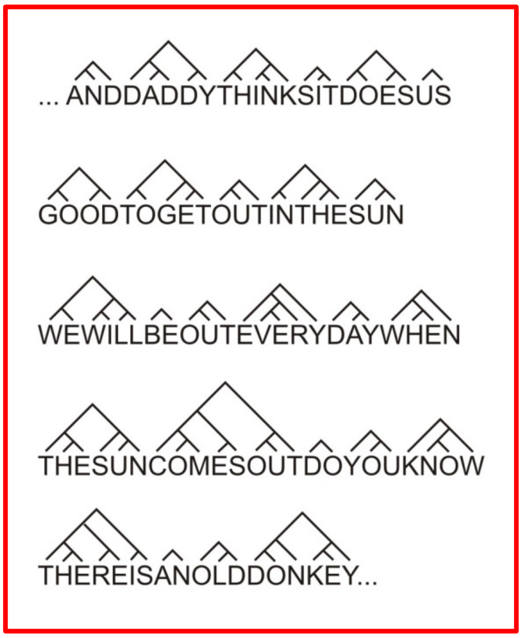


Here, each of the two images is an entirely random pattern of black and white pixels, prepared so that they are the same, except that a square area near the middle of the left image is shifted to the left in the right image.

When we view the two images in a stereoscope (illustrated in the figure), so that one's left eye sees the left image and one's right eye sees the right image, the central square takes shape gradually as a discrete object which appears to stand out in relief against the background.

Since each image is, in itself, entirely random—meaning that the central square exists only in the combination of the two images—the only way in which we can see that central square is by merging the two images, pixel-by-pixel, to create a single view—and this means compression of information.

Information compression came to the fore again later when I was looking at the problem of segmentation in the learning of a first language: how is it that young children, apparently without the benefit of reading, can develop a sense that language is composed of discrete entities like words when it is clear from how speech appears on an oscilloscope or sound spectrogram, that we mostly speak in "ribbons" of sound without gaps or anything else to separate the words?



After a good deal of experimentation, with false starts, I came up with a computer program that could largely solve an analogous problem with text, prepared without any punctuation or spaces between the words. This program, without being given any dictionary, could analyse a sample of unsegmented text and do a fairly good job of discovering the structure of the text in terms of words, as shown in the figure. The groupings that the program has worked out are shown above the text.

With other samples, prepared as sequences of word classes, it did quite well at discovering the structure of the text in terms of phrases.

Boiled down to its essentials, what the program does is to build up its own dictionary of repeating segments of text, concentrating on those that provide the most effective means of compressing the text.

This evidence suggests that it is at least feasible for children to learn the segmental structure of language in a similar way.

This work led to the development of another computer program that, without any “teacher” or similar assistance, is quite effective at discovering the grammatical structure of samples of English-like artificial languages, including the structure of the text in terms of words and sequences of words, and the discovery of grammatical classes of words. Again, information compression is the key.

For me, a change of employment, into the software industry, seems to have been the “grit in the oyster” that led to a new programme of work developing the SP theory itself. A project to develop an “integrated project support environment” (IPSE) for software engineers raised the question of how to integrate the many versions of a typical software product with the parts and sub-parts of the product, each of which typically has several versions.

This, with several other things I had learned about software, connected in my mind with the problem of integrating the syntax of natural language with its meanings or semantics, in brains or in artificial systems. I began to see the possibility of developing a new kind of computing system, which might also be a new model of human cognition, that would solve these and other problems. To develop these ideas, I returned to academic work.

The challenge with this new programme of research has been to create a model or framework that would simplify and integrate a range of observations and concepts in artificial intelligence, mainstream computing, and human perception and cognition. Because the developing ideas seemed also to apply to mathematics, that was added to the list later.

The motivation, in the SP research, for simplification and integration across a broad canvass is partly because this is what science is meant to be about, and partly because of the way computer science, artificial intelligence, and cognitive science have become fragmented.

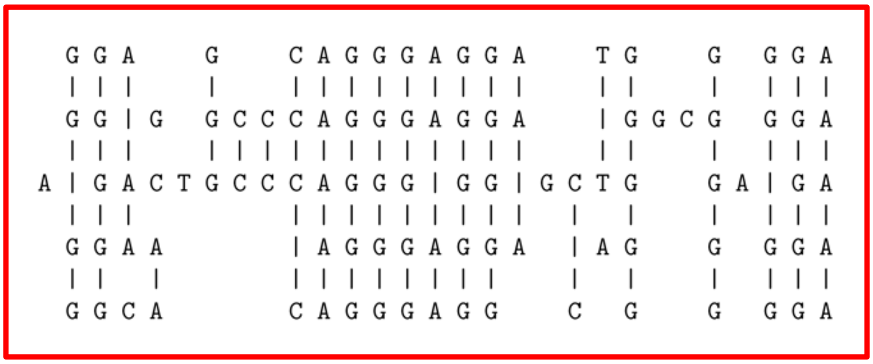
The fragmentation of AI has been well described by Pamela McCorduck: “The goals once articulated with debonair intellectual verve by AI pioneers appeared unreachable ... Subfields broke off—vision, robotics, natural language processing, machine learning, decision theory—to pursue singular goals in solitary splendor, without reference to other kinds of intelligent behaviour.”

In connection with research in cognitive psychology, there has been criticism of research with “microtheories”, with only a very narrow range of application. And Allen Newell exhorted psychologists to develop theories that would deal with “a genuine slab of human behaviour”.

What became clear at an early stage was that my earlier computer models would not do. They build structures in a predominantly hierarchical manner and there seemed to be little chance of adapting them to the non-hierarchical aspects of things like pattern recognition. The new theory, and its expression in a new computer model, should accommodate both hierarchical and non-hierarchical aspects of intelligence, and other aspects as well.

Another thing that become increasingly clear as work progressed was that it was possible to side-step much of the mathematical complexity associated with information compression by concentrating on a simple, “primitive” idea: that any given body of information (***I***) may be compressed by finding patterns in ***I*** that match each other and merging or ‘unifying’ them. This idea, dubbed “information compression via the matching and unification of patterns”, (ICMUP) is bedrock in the SP theory.

Early work, developing methods in the manner of “dynamic programming” to find good full and partial matches between pairs of patterns, led me to an interest in *multiple alignment*, something that has been part of the information-processing tool kit for analysing biochemical structures for many years. This has turned out to be a major breakthrough in developing the SP theory, so I’ll described what it is and why—with some modifications—it can do so much for us.



```
  G G A     G     C A G G G A G G A     T G     G   G G A
  | | |     |     | | | | | | | | |     | |     |   | | |
  G G | G   G C C C A G G G A G G A     | G G C G   G G A
  | | |     | | | | | | | | | | | |     | |     |   | | |
A | G A C T G C C C A G G G | G G | G C T G     G A | G A
  | | |           | | | | | | | | |   |   |     |   | | |
  G G A A         | A G G G A G G A   | A G     G   G G A
  | |   |         | | | | | | | |     |   |     |   | | |
  G G C A         C A G G G A G G     C   G     G   G G A
```

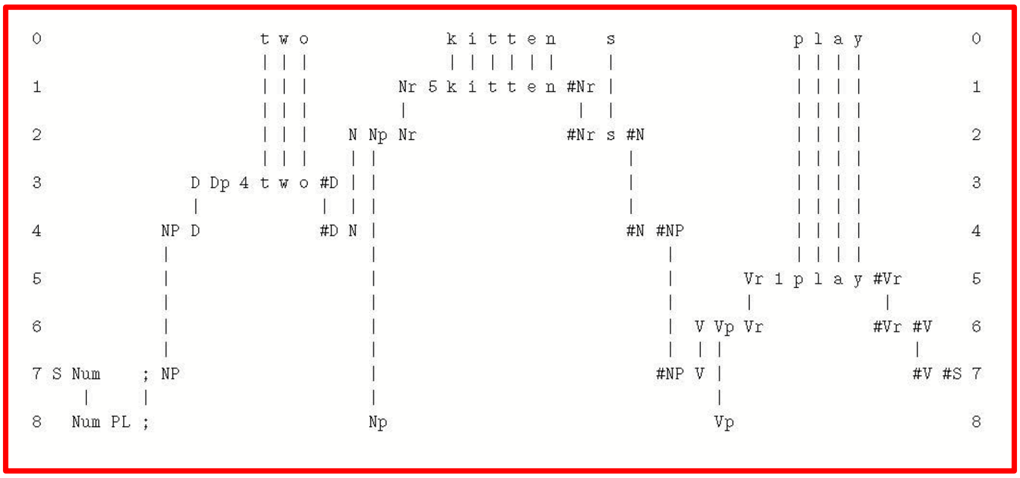
In biochemistry, it is often useful to make comparisons between DNA sequences or sequences of amino-acid residues, not merely two at a time but with three, four, or more sequences, all together. The figure shows how this has been done with five DNA sequences, each one expressed with a four-letter alphabet, one letter for each of four bases: adenine (A), thymine (T), guanine (G) and cytosine (C).

In the figure, there has been judicious stretching of the sequences to bring matching letters into vertical lines. There is normally an astronomically large number of ways in which this can be done, so a computer is needed that can sort through the many possibilities quickly, gradually narrowing the field to one or two multiple alignments that score well in terms of the numbers of letters that have been brought into line. Even with a computer, it is necessary to use “heuristic” techniques to cut down the amount of searching, by choosing. at every stage, only the most promising paths to explore. That means that, normally, we cannot be sure of finding the best possible answer but must be content with answers that are “good enough”.

A clue that multiple alignment, as it has been developed in biochemistry research, might be relevant to the SP programme of research is that bringing matching symbols into alignment has clear potential for ICMUP—a potential that is realised in multiple alignment as it has been developed in the SP theory.

The three examples described next are intended to show how the multiple alignment concept has been adapted in the SP theory and some of the versatility of the concept in modelling different aspects of intelligence. The examples are simple, partly because complicated examples would not fit easily on the page and partly because such examples would be hard to follow. But it should be emphasised that the simplicity of the examples should not be interpreted as naivety or weakness in the SP system—its potential is very large.

The main difference between multiple alignment in biochemistry and multiple alignment in the SP system is that, in the latter case, one sequence or “pattern” (sometimes more than one) represents incoming information and is classified as “New”, while the rest represent stored information and are classified as “Old”. Incidentally, the current SP computer model works only with one-dimensional patterns but it is envisaged that, at some stage, it will be generalised to work with two-dimensional patterns—to represent images, other 2D kinds of knowledge, and 3D knowledge in the manner of commercial applications that create 3D models from photographs.



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

In the figure, which shows the best of a set of multiple alignments created by the SP computer model, the sentence “`t w o k i t t e n s p l a y`” is a New pattern which, by convention, is always shown in row 0. Letters and strings of letters in this and later examples are intended to represent any kind of basic perceptual element in vision, hearing, touch, and so on.

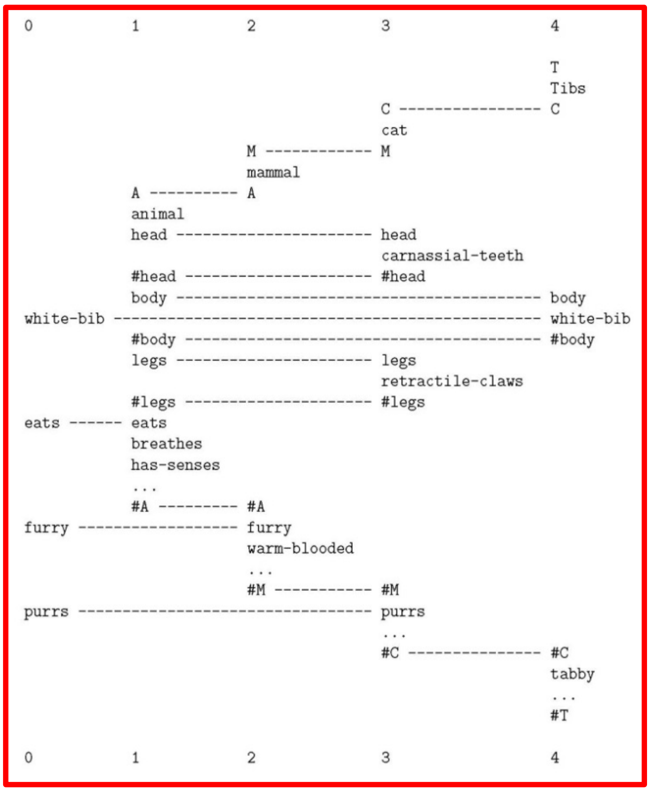
Rows 1 to 8 show Old patterns, one pattern per row, representing grammatical structures, including words. The whole multiple alignment shows how the sentence may be analysed or “parsed” into a noun phrase (“`t w o k i t t e n s`”) and a verb (“`p l a y`”), with the pattern is row 4 representing the noun-phrase structure (“`NP`”) and the pattern in row 6 representing the verb (“`V`”).

In rows 1 to 3, the noun phrase is broken down in a similar way into a ‘determiner’ (“`t w o`”) and a ‘noun’ (“`k i t t e n s`”), with the relevant grammatical categories marked in rows 3 and 2, respectively.

The pattern in row 8 shows how the plural subject of the sentence (marked with “`Np`”) must be followed by a plural verb (marked with “`Vp`”).

This example illustrates some of the versatility of the multiple alignment concept as it has been adapted in the SP theory—showing how a framework which is not, in itself, hierarchical, may be used to model something that is hierarchical: the parsing of a sentence into its parts and subparts.

A key feature of the SP system is that it is fundamentally probabilistic. It can calculate absolute and relative probabilities for any multiple alignment and for any inference that may be drawn from a multiple alignment. The probabilistic nature of the system flows directly from the intimate connection that is known to exist between information compression and concepts of prediction and probability.

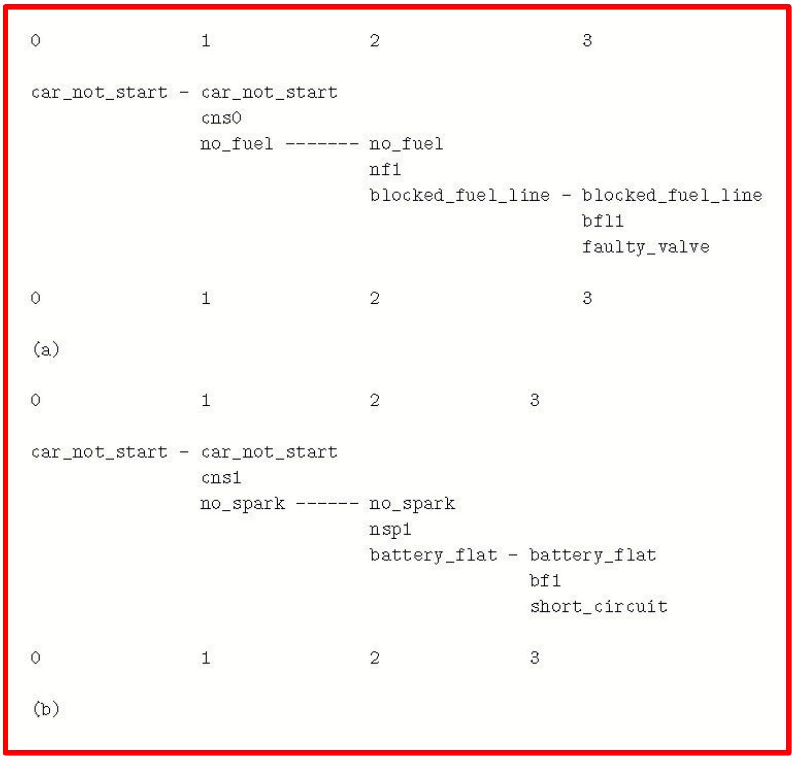


```
0           1            2              3                  4

                                                           T
                                                           Tibs
                                        C ---------------- C
                                        cat
                         M ------------ M
                         mammal
            A ---------- A
            animal
            head ---------------------- head
                                        carnassial-teeth
            #head --------------------- #head
            body ----------------------------------------- body
white-bib ------------------------------------------------ white-bib
            #body ---------------------------------------- #body
            legs ---------------------- legs
                                        retractile-claws
            #legs --------------------- #legs
eats ------ eats
            breathes
            has-senses
            ...
            #A --------- #A
furry ------------------ furry
                         warm-blooded
                         ...
                         #M ----------- #M
purrs --------------------------------- purrs
                                        ...
                                        #C --------------- #C
                                                           tabby
                                                           ...
                                                           #T

0           1            2              3                  4
```

A second example from the SP computer model shows, in a simplified form, how the system may provide a means of recognising something at several different levels of abstraction. This example, like the previous one, is the best of a set of multiple alignments created by the SP computer model, but it is rotated by 90° because that makes it fit better on the page.

In the figure, the New pattern in column 0 represents the features of some unknown creature, while the Old patterns in columns 1 to 4 show that the creature has been recognised as an animal (column 1), a mammal (column 2), a cat (column 3), and a specific cat ‘Tibs’ (column 4).

Not shown in this example is how recognition may be achieved despite errors of various kinds in the New pattern, how the parts and sub-parts of something may be represented and combined with classes and sub-classes and how the system may accommodate ‘family resemblance’ or ‘polythetic’ categories (where recognition does not depend on the presence or absence of any specific feature or combination of features).



```
0              1              2                  3

car_not_start - car_not_start
               cns0
               no_fuel ------- no_fuel
                              nf1
                              blocked_fuel_line - blocked_fuel_line
                                                 bfl1
                                                 faulty_valve

0              1              2                  3

(a)

0              1              2              3

car_not_start - car_not_start
               cns1
               no_spark ------ no_spark
                              nsp1
                              battery_flat - battery_flat
                                             bf1
                                             short_circuit

0              1              2              3

(b)
```

Our third example shows, in each of two multiple alignments, how the SP system may model a simple chain of reasoning in diagnosing what's wrong with a car that will not start.

The multiple alignment in (a) shows that "`car_not_start`" in column 0 may be due to there being no fuel, a possible cause of which is a blocked fuel line, which may itself be due to a faulty valve.

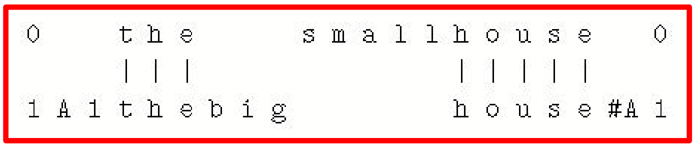
The multiple alignment in (b) shows another line of reasoning: that the same fault may be due to the absence of a spark, which may be caused by a flat battery, which may be due to a short circuit.

It turns out that the multiple alignment framework can not only model chains of reasoning as in this example but several other kinds of reasoning as well, as detailed later.

So far, we have been assuming that the store of Old patterns has been given to the system, ready-made, and that is how, using the SP computer model, the examples have been produced. But a key part of the SP system is that it can learn Old patterns for itself.

Unlike my previous models for learning, which built up patterns in an hierarchical manner, the SP computer model derives Old patterns largely from multiple alignments, except for the early stages when Old patterns may be derived directly from New patterns or parts thereof.

At the beginning of processing by the system, when the repository of Old patterns is empty or nearly so, each New pattern is stored fairly directly as an Old pattern, much as it is received, but with the addition of "ID" symbols at the beginning and end.



```
0     t h e       s m a l l h o u s e   0
      | | |                 | | | | |
1 A 1 t h e b i g           h o u s e #A 1
```

Later, when some Old patterns have been stored, the system may start to recognise full or partial matches between New and Old patterns, like the partial match between "`t h e s m a l l h o u s e`" and "`A 1 t h e b i g h o u s e #A`" shown in the figure. From the parts that match in a multiple alignment like this, and the parts that don't match, the system can begin to form Old patterns like "`B 1 t h e #B`", "`C 1 h o u s e #C`", "`D 1 s m a l l #D`", and "`D 2 b i g #D`". In addition, the system forms an abstract pattern which records the sequence of words in terms the ID-symbols of the constituent patterns.

Of course, things are often much less tidy than this example may suggest. But, as in the building of multiple alignments, the system uses heuristic search to sift out the structures that are useful in terms of information compression, and to discard the rest. The SP system, with some further development, has clear potential to build up "grammars" for the New information it receives, including structures at various levels of complexity and abstraction.

What has been described so far is intended to give an indication of how one relatively simple system may exhibit the kind of versatility and adaptability needed for human-like artificial intelligence and to model human perception, cognition, and learning.

To fill out the picture, the SP system has strengths and potential in the representation and processing of several different kinds of knowledge, including: the syntax of natural language; class hierarchies, part-whole hierarchies, and their integration; trees and networks, including Bayesian networks; entity-relationship structures; relational knowledge (tuples); if-then rules, associations, and other knowledge in support of reasoning; patterns and images; structures in three dimensions; and sequential and parallel procedures.

Because all these kinds of knowledge are represented in one simple format (SP patterns) and because they are all processed in one simple framework (multiple alignment), they can work together seamlessly in any combination, an integration that appears to be essential in any system that aspires to human-like intelligence.

The SP system also has strengths and potential in several different aspects of intelligence, including: unsupervised learning (meaning learning without the aid of a “teacher” or similar assistance); natural language processing (including the parsing and production of language, with potential for the understanding of natural language and the production of language from meanings); fuzzy pattern recognition and recognition at multiple levels of abstraction; computer vision and the modelling of natural vision; best-match and semantic forms of information retrieval; planning; problem solving; and several kinds of reasoning (including one-step ‘deductive reasoning’, abductive reasoning, probabilistic networks and trees, reasoning with ‘rules’, nonmonotonic reasoning, ‘explaining away’, causal reasoning, and reasoning that is not supported by evidence).

As with different kinds of knowledge, the use of one simple framework as a foundation for diverse aspects of intelligence facilitates their seamless integration in any combination, an integration which, as before, appears to be essential for human-like versatility and adaptability in intelligence.

Distance (*m*)	*Time* (*sec*)
0.0	0
4.9	1
19.6	2
44.1	3
78.5	4
122.6	5
176.5	6
240.3	7
313.8	8
397.2	9
490.3	10
593.3	11
706.1	12
Etc	Etc

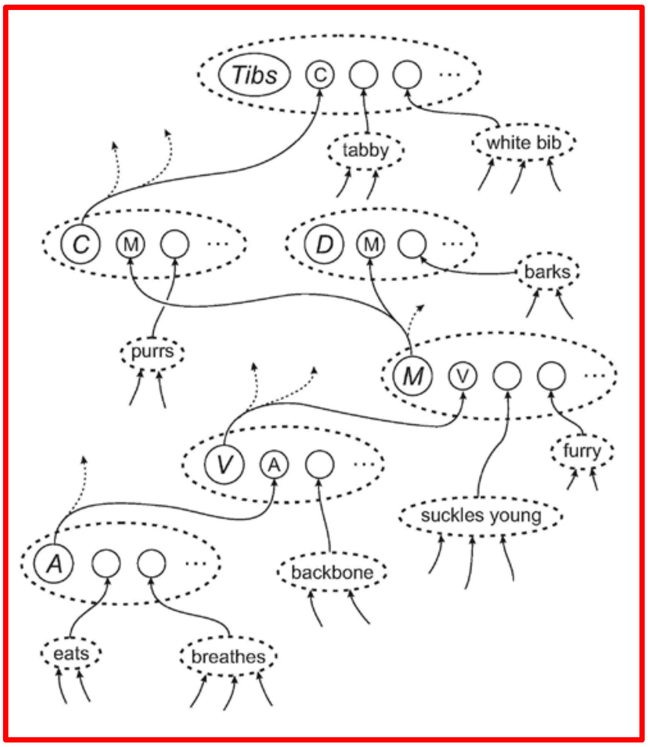
As I mentioned earlier, it became increasingly clear as this work progressed, that much of the thinking might also apply to mathematics. Consider, for example, Newton’s equation for his second law of motion: $s = (gt^2)/2$, where s is the distance travelled by a falling object in a given time (t) since it started to fall, and g is a measure of acceleration due to gravity (9.8 m/s^2 at sea level). This is an extremely compact way of representing any realistically-large table of distances and times for a falling object, like the one shown, or much bigger. In other words, the equation achieves high levels of information compression.

A little analysis shows that several aspects of mathematics may be understood as ICMUP. It is tempting to suggest, as a working hypothesis, that all of the constructs used in mathematical analysis may be understood as devices for the compression of information and, as a working hypothesis, that all of them may be seen to work via ICMUP.

Since multiple alignment incorporates ICMUP and since it has proved to be very versatile, it seems possible that much of mathematics, perhaps all of it, may be understood in terms of multiple alignment as it has been developed in the SP programme of research.

Although the development of human-like intelligence has been a main focus in the SP research, the SP theory also has some useful things to say about “mainstream” computing. It appears that concepts from ordinary programming and from software engineering like *procedure*, *variable*, *value*, *type*, *function with parameters*, *conditional statements*, *recursion*, *object-oriented design*, and *parallel processing*, can all be modelled in the multiple alignment framework.

A key point about the SP system is that there would be no “programming” in the sense of hand-crafting instructions to make the SP machine do what one wants. Instead, the system works by compressing data—but those data may include examples of how things are done, like baking a cake or buying something in a shop. The system’s strengths and potential in unsupervised learning suggests that there is potential for the automatic creation of programs from appropriate data, reducing or eliminating much of the labour in ordinary programming.



One aspect of the system that has not yet been mentioned is that it appears to be possible, in a fairly straightforward manner, to map abstract concepts in the SP theory into concepts in a version of the theory—*SP-neural*—that is couched in terms of neurons and their interconnections.

One such mapping is shown in the figure. Here, ellipses with broken lines represent neural structures called *pattern assemblies* that correspond to patterns in the main SP theory, and neural connections between pattern assemblies correspond to alignments or potential alignments between matching symbols in the main theory.

At present, the main value of SP-neural, in conjunction with the main SP theory, is to be a source of hypotheses about the representation and processing of knowledge in the brain. Looking further ahead, it may provide the basis for a new architecture for computers which would be quite different from the current generation of ‘artificial neural networks’ and with potential advantages over conventional computers.

So far, so good. On the strength of what has been said so far, one could say that the SP system is “interesting”. But it is natural to ask how the system compares with other approaches to artificial intelligence and related topics, what the system might do for us in more concrete terms, and what needs to be done to realise the benefits.

Regarding the first question, the SP system has been compared in some detail with several other AI-related systems and, on the strength of this analysis, it appears to provide a firmer foundation for developing an understanding of “intelligence”, broadly construed, than the main alternatives.[2] In particular, several problems associated with “deep learning in artificial neural networks” (the subject of much current interest) may be solved with the SP system.

---

[2] See “The SP theory of intelligence: distinctive features and advantages”, PDF, J G Wolff, IEEE Acess, 4, 216-246, 2016, bit.ly/21gv2jT.

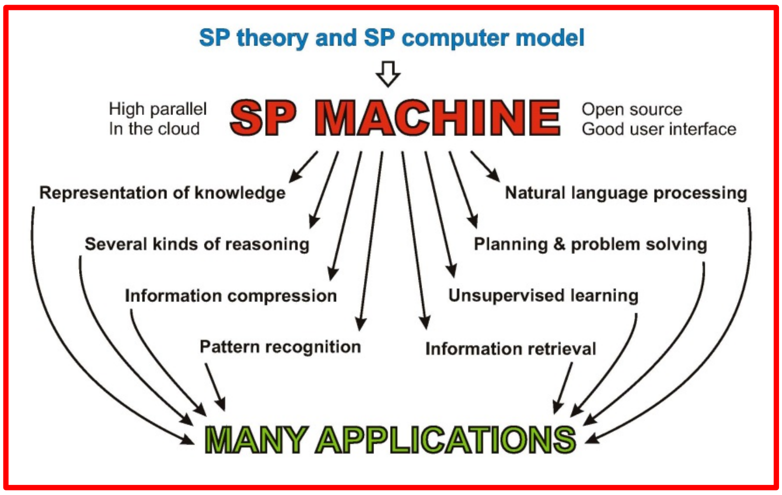
Some of the possible applications of the SP system have been described in peer-reviewed papers. They include: the processing of natural language; the system may function as a new and improved kind of database with intelligence, and it has potential applications in areas such as crime detection and prevention; it has potential to help develop human-like versatility and adaptability in autonomous robots; it may help in the development of computer vision, including the analysis of scenes, and the integration of vision with other aspects of intelligence; it has potential applications in medical diagnosis viewed as a process of pattern recognition, and also medical diagnosis via causal reasoning; it may provide a new basis for software engineering, including the automatic creation of software from relevant data; since information compression lies at the heart of the SP system, it may serve as a means of reducing the size of any body of data; it may also be applied in the analysis of DNA sequences or sequences of amino-acid residues, in the detection of computer viruses, in ‘data fusion’ (the integration of separate but related bodies of information), and, as with SP-neural, the main SP theory may provide the basis for a radically new architecture for computers with advantages over computers as they are now.

Somewhat unexpectedly, the SP system has potential to help solve nine problems associated with “big data”, the humongous quantities of data that now flow from industry, commerce, social media, science, and more. Those problems, very well described in a book called *Smart Machines*[3] by John Kelly and Steve Hamm, both of IBM, are important because, unless good solutions can be developed, it will be difficult or impossible, even with supercomputers, to extract anything more than a small part of the value in the truly humungous quantities of data that are now coming from science, industry, commerce, public administration, health services, and elsewhere.

As an example of how the SP theory may help gain value from big data, it may help to tame the great variety of formalisms and formats that are used in the computing industry to represent different kinds of knowledge, each with their own mode of processing. SP patterns in the multiple alignment framework have potential as a “universal framework for the representation and processing of diverse kinds of knowledge” (UFK).

Another possibility with the SP system and big data is making big increases in the speed of transmitting data from one place to another and with big cuts in costs, by judicious separation of the “grammar” of the data and its “encoding”.

There is also potential for big cuts in the amount of energy needed to process big data by taking advantage of the statistical information that the SP system gathers as a by-product of how it works.



The SP research programme has now reached the stage where there are far more avenues to be explored than could be tackled by any one research group. Things need to be opened up to many other research groups and individual researchers.

It is envisaged that, from the SP computer model, the *SP machine* will be created as a software virtual machine hosted on an existing high-performance

---

[3] *Smart Machines: IBM's Watson and the Era of Cognitive Computing*, John E Kelly III and Steve Hamm, New York: Columbia University Press, 2013.

computer and enhanced with high levels of parallel processing and a user-friendly user interface.

This facility will be a means for researchers anywhere in the world to explore what can be done with the SP system and to create their own versions of it.

Details of publications in the SP programme of research are given, in many cases with download links, on www.cognitionresearch.org/sp.htm, and on linked web pages.