Introduction to the SP Theory of Intelligence
and its realisation in the SP Computer Model

J Gerard Wolff

March 8, 2019

Abstract

This paper provides an introduction to the *SP Theory of Intelligence* and its realisation in the *SP Computer Model*. The overall goal of the SP programme of research, in accordance with long-established principles in science, has been the simplification and integration of observations and concepts across artificial intelligence, mainstream computing, mathematics, and human learning, perception, and cognition. In accordance with evidence from human learning, perception, and cognition, information compression is central in the workings of the SP System. The main mechanism for achieving information compression is the powerful concept of *SP-multiple-alignment*, borrowed and adapted from bioinformatics. This is the key to the system’s versatility in diverse aspects of intelligence, in the representation of diverse kinds of knowledge, and in the seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination. It is also the key to strengths of the SP System compared with alternatives such as deep learning, and to several potential benefits and applications of the SP System.

1 Introduction

The *SP System*—meaning the *SP Theory of Intelligence* and its realisation in the *SP Computer Model*—is a system that has been under development since about 1987, with a break between early 2006 and late 2012.
The SP System is described in outline here, in more detail in [14], and quite fully in [12]. Distinctive features and advantages of the SP System are described in [21]. Other papers in this programme of research are detailed, with download links, on www.cognitionresearch.org/sp.htm.

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

Figure 1: Schematic representation of the SP System from an ‘input’ perspective. Reproduced, with permission, from Figure 1 in [14].

In the SP System, all kinds of knowledge are represented with SP-patterns, where each such SP-pattern is an array of atomic SP-symbols in one or two dimensions. At present, the SP Computer Model works only with one-dimensional SP-patterns but it is envisaged that it will be generalised to work with two-dimensional SP-patterns as well.
1.1 Aiming for a favourable combination of conceptual Simplicity with descriptive or explanatory Power

The SP programme of research is a unique attempt to simplify and integrate observations and concepts across artificial intelligence, mainstream computing, mathematics, and human learning, perception, and cognition. This may be seen to be a process of developing concepts that combine conceptual Simplicity with high levels of descriptive or explanatory Power. Hence the name “SP” (see also Section 1.2).

This strategy accords with Ockham’s razor: a theory should be simple but not so simple that it becomes trivial, meaning that it loses descriptive or explanatory ‘power’. This top-down approach to the development of concepts contrasts with the more popular bottom-up approach which seeks to develop ideas in one area such as computer vision and then integrate it with other areas such as reasoning, and to repeat that kind of integration to create groupings of progressively increasing size.

Despite its ambition, the simplicity-with-power objective has been largely met. This is because the SP System has strengths and potential across diverse aspects of intelligence and the representation of knowledge (1.9).

1.2 Information compression in human cognition and the SP System

A central idea in the SP System comes from research into human learning, perception and cognition (HLPC): that many aspects of human intelligence may be understood as information compression (IC). Evidence for the importance of IC in HLPC is described in [25].

Compression of a given body of information, I, may be understood as a process of promoting Simplicity in I by extracting redundancy from I, whilst retaining as much as possible of its non-redundant descriptive or explanatory Power. This is a second reason for the name “SP”, additional to that mentioned in Section 1.1.

In the development of the SP System, it has proved useful to understand IC in terms of the discovery of patterns that match each other and the merging or ‘unification’ of patterns that are the same. The expression ‘information compression via the matching and unification of patterns’ may be shortened to ‘ICMUP’. Seven variants of ICMUP are outlined in [25, Section 2.1]. The seventh variant, the concept of SP-multiple-alignment, is central in the SP System and it is described in Section 1.4, below.

The emphasis on IC in the SP System is broadly in accordance with research in the tradition of Minimum Length Encoding (see, for example, [4]),
with the qualification that most research relating to MLE assumes that the concept of a universal Turing machine provides the foundation for theorising, whereas the SP System is founded on concepts of ICMUP and SP-multiple-alignment. The way in which the workings of a universal Turing machine may be understood in terms of the SP Theory is described in [12, Chapter 4].

1.3 The probabilistic nature of the SP System

Owing to the intimate relation that is known to exist between IC and concepts of inference and probability (Appendix A), and owing to the fundamental role of IC in the workings of the SP System, the system is inherently probabilistic.

That said, it appears to be possible to imitate the all-nothing-nature of conventional computing systems via the use of data where most probabilities yielded by the system, or all of them, are close to 0 or 1.

Because of the probabilistic nature of the SP System, it lends itself to the modelling of HLPC because of the prevalence of uncertainties in that domain. Also, the SP System sits comfortably within AI because of the probabilistic nature of most operations in AI.

An advantage of the SP System in those areas is that it is relatively straightforward to calculate absolute or conditional probabilities for results obtained in, for example, different kinds of reasoning [1.9.2].

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

1.4 SP-multiple-alignment

A central idea in the SP System, is the simple but powerful concept of SP-multiple-alignment, borrowed and adapted from the concept of ‘multiple sequence alignment’ in bioinformatics. As mentioned in [1.2] SP-multiple-alignment is the seventh variant of ICMUP described in [25, Section 2.1] and may be seen as a generalised version of the other six variants.

Probably the best way to explain the idea is by way of examples. Figure 2 shows an example of multiple sequence alignment in bioinformatics. Here, there are five DNA sequences which have been arranged alongside each other, and then, by judicious ‘stretching’ of one or more of the sequences in a computer, symbols that match each other across two or more sequences have been brought into line.
A ‘good’ multiple sequence alignment, like the one shown, is one with a relatively large number of matching symbols from row to row. The process of discovering a good multiple sequence alignment is normally too complex to be done by exhaustive search, so heuristic methods are needed, building multiple sequence alignments in stages and, at each stage, selecting the best partial structures for further processing.

Figure 3 shows an example of an SP-multiple-alignment, superficially similar to the one in Figure 2, except that the sequences are called SP-patterns, the SP-pattern in row 0 is New information and the remaining SP-patterns, one per row, are Old information. A ‘good’ SP-multiple-alignment is one which allows the New SP-pattern to be encoded economically in terms of the Old SP-patterns.

In this example, the New SP-pattern (in row 0) is a sentence and each of the remaining SP-patterns represents a grammatical category, including morphemes and words.

In this example, the New SP-pattern (in row 0) is a sentence and each of the remaining SP-patterns represents a grammatical category, where ‘gram-
matical categories’ include words. The overall effect of SP-multiple-alignment in this example is the parsing a sentence (‘fortunefavours thebrave’) into its grammatical parts and sub-parts.

But the SP-multiple-alignment concept is very versatile and, as described in [1.6] and [1.9], it may serve to model several different aspects of intelligence, including several kinds of reasoning, it may serve in the representation of several different kinds of knowledge, and it facilitates SIIKAC.

Although the search space for good SP-multiple-alignments is normally very large, the use of heuristic methods helps to ensure that computational complexities in the SP System are within reasonable bounds [12, Sections A.4, 3.10.6 and 9.3.1].

1.5 Unsupervised learning in the SP System

In the SP System, learning is ‘unsupervised’, deriving structures from incoming sensory information without the need for any kind of ‘teacher’, or anything equivalent (cf. [2]).

Unsupervised learning in the SP System is quite unlike ‘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.” [3, location 1500].

This form of learning, often summarised as Cells that fire together wire together, has been adopted, with variations, in most ‘deep neural networks (DNNs), and in some other AI systems.

In the SP System, unsupervised learning incorporates the building of SP-multiple-alignments but there are other processes as well. In brief, the system creates Old SP-patterns from complete New SP-patterns and also from partial matches between New and Old SP-patterns. When all the New SP-patterns have been processed like that, the system creates one or two ‘good’ SP-grammars, where an SP-grammar is a collection of Old SP-patterns, and it is ‘good’ if it is effective in the economical encoding of the original set of New SP-patterns.

As with the building of SP-multiple-alignments, the process of creating good grammars is normally too complex to be done by exhaustive search so heuristic methods are needed. This means that the system builds SP-grammars incrementally and, at each stage, it discards all but the best SP-grammars. As with the building of SP-multiple-alignments, the use of heuristic
methods helps to ensure that computational complexities in the SP System are within reasonable bounds [12, Sections A.4, 3.10.6 and 9.3.1].

The SP Computer Model has already demonstrated an ability to learn generative grammars from unsegmented samples of English-like artificial languages, including segmental structures, classes of structure, and abstract patterns, and to do this in an ‘unsupervised’ manner ([14, Section 5], [12, Chapter 9]). But there are (at least) two shortcomings in the system [14, Section 3.3]: it cannot learn intermediate levels of structure or discontinuous dependencies in grammar, although the SP-multiple-alignment framework can accommodate structures of those kinds. It appears that those two problems may be overcome and that their solution would greatly enhance the capabilities of the SP Computer Model in unsupervised learning.

1.6 Two main mechanisms for information compression in the SP System, and their functions

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

1. The building of SP-multiple-alignments. The process of building SP-multiple-alignments achieves compression of New information. At the same time it may achieve any or all of the following functions described in [12, Chapters 5 to 8] and [14, Sections 7 to 12], with potential for more:

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

(b) Pattern recognition which is robust in the face of errors of omission, commission, or substitution; and pattern recognition at multiple levels of abstraction.

(c) Information retrieval which is robust in the face of errors of omission, commission, or substitution.

(d) Several kinds of probabilistic reasoning: 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.
(e) Planning such as, for example, finding a flying route between London and Beijing.
(f) Problem solving such as solving the kinds of puzzle that are popular in IQ tests.

The building of SP-multiple-alignments is also part of the process of unsupervised learning, next.

2. Unsupervised learning. Unsupervised learning, outlined in Section 1.5 means the creation of one or two grammars which are collections of SP-patterns which are effective in the economical encoding of a given set of New SP-patterns.

1.7 SP-Neural

A potentially useful feature of the SP System is that it is possible to see how abstract constructs and processes in the system may be realised in terms of neurons and their interconnections. This is the basis for SP-Neural, a ‘neural’ version of the SP System, described in [20].

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

In this connection, it is relevant to mention that the SP System, in both its abstract and neural forms, is quite different from deep learning in neural networks [7] and has substantial advantages compared with such systems ([21, Section V], [21]).

1.8 Generalising the SP System for two-dimensional SP-patterns, both static and moving

This brief description of the SP System and how it works may have given the impression that it is intended to work entirely with sequences of SP-symbols, like multiple sequence alignments in bioinformatics. But it is envisaged that, in future development of the system, two-dimensional SP-patterns will be introduced, with potential to represent and process such things as photographs and diagrams, and structures in three dimensions as described in [15, Section 6.1 and 6.2], and procedures that work in parallel as described in [16, Sections V-G, V-H, and V-I, and C].
Figure 4: A schematic representation of a partial SP-multiple-alignment in SP-Neural, as discussed in [20, Section 4]. Each broken-line rectangle with rounded corners represents a pattern assembly—corresponding to an SP-pattern in the main SP Theory of Intelligence; each character or group of characters enclosed in a solid-line ellipse represents a neural symbol corresponding to an SP-symbol in the main SP Theory of Intelligence; the lines between pattern assemblies represent nerve fibres with arrows showing the direction in which impulses travel; neural symbols are mainly symbols from linguistics such as ‘NP’ meaning ‘noun phrase’, ‘D’ meaning a ‘determiner’, ‘#D’ meaning the end of a determiner, ‘#NP’ meaning the end of a noun phrase, and so on.
It is envisaged that, at some stage, the SP System will be generalised to work with sequences of two-dimensional ‘frames’ from films or videos.

1.9 Strengths and potential of the SP System

The strengths and potential of the SP System are summarised in the subsections that follow. Further information may be found in [14, Sections 5 to 12], [12, Chapters 5 to 9], [21], and in other sources referenced in the subsections that follow.

1.9.1 Versatility in aspects of intelligence

The SP System has strengths and potential in the ‘unsupervised’ learning of new knowledge. As noted in 1.5, this is an aspect of intelligence in the SP System that is different from others because it is not a by-product of the building of multiple alignments but is, instead, achieved via the creation of grammars, drawing on information within SP-multiple-alignments.

Other aspects of intelligence exhibited by the SP System are modelled via the building of SP-multiple-alignments. These other aspects of intelligence include: the analysis and production of natural language; pattern recognition that is robust in the face of errors in data; pattern recognition at multiple levels of abstraction; computer vision [15]; best-match and semantic kinds of information retrieval; several kinds of reasoning (next subsection); planning; and problem solving.

1.9.2 Versatility in reasoning

Kinds of reasoning exhibited by the SP System include: one-step ‘deductive’ reasoning; chains of reasoning; abductive reasoning; reasoning with probabilistic networks and trees; reasoning with ‘rules’; nonmonotonic reasoning and reasoning with default values; Bayesian reasoning with ‘explaining away’; causal reasoning; reasoning that is not supported by evidence; the inheritance of attributes in class hierarchies; and inheritance of contexts in part-whole hierarchies. Where it is appropriate, probabilities for inferences may be calculated in a straightforward manner ([12, Section 3.7], [14, Section 4.4]).

There is also potential in the system for spatial reasoning [16, Section IV-F.1], and for what-if reasoning [16, Section IV-F.2].

It seems unlikely that the features of intelligence mentioned above are the full extent of the SP System’s potential to imitate what people can do. The close connection that is known to exist between IC and concepts of inference and probability [A], the central role of IC in the SP-multiple-alignment...
framework, and the versatility of the SP-multiple-alignment framework in aspects of intelligence suggest that there are more insights to come.

As noted in section 1.3, the probabilistic nature of the SP System makes it relatively straightforward to calculate absolute or conditional probabilities for results from the system, as for example in its several kinds of reasoning, most of which would naturally be classed as probabilistic.

1.9.3 Versatility in the representation of knowledge

Although SP-patterns are not very expressive in themselves, they come to life in the SP-multiple-alignment framework. Within that framework, they may serve in the representation of several different kinds of knowledge, including: the syntax of natural languages; class-inclusion hierarchies (with or without cross classification); part-whole hierarchies; discrimination networks and trees; if-then rules; entity-relationship structures (sections 3 and 4); relational tuples (ibid., section 3), and concepts in mathematics, logic, and computing, such as ‘function’, ‘variable’, ‘value’, ‘set’, and ‘type definition’ (12, Chapter 10; 18, Section 6.6.1; 23, Section 2).

As previously noted, the addition of two-dimensional SP patterns to the SP Computer Model is likely to expand the representational repertoire of the SP System to structures in two-dimensions and three-dimensions, and the representation of procedural knowledge with parallel processing.

As with the SP System’s generality in aspects of intelligence, it seems likely that the SP System is not constrained to represent only the forms of knowledge that have been mentioned. The generality of IC as a means of representing knowledge in a succinct manner, the central role of IC in the SP-multiple-alignment framework, and the versatility of that framework in the representation of knowledge, suggest that the SP System may prove to be a means of representing all the kinds of knowledge that people may work with.

1.9.4 Seamless integration of diverse aspects of intelligence, and diverse kinds of knowledge, in any combination

An important third feature of the SP System, alongside its versatility in aspects of intelligence and its versatility in the representation of diverse kinds of knowledge, is that there is clear potential for the SP System to provide seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination. (SIIKAC) This is because diverse aspects of intelligence and diverse kinds of knowledge all flow from a single coherent and relatively simple source: the SP-multiple-alignment framework.
It appears that seamless integration of diverse aspects of intelligence and diverse kinds of knowledge, in any combination, is essential in any artificial system that aspires to the fluidity, versatility and adaptability of the human mind.

Figure 5 shows schematically how the SP System, with SP-multiple-alignment centre stage, exhibits versatility and integration.

1.10 Potential benefits and applications of the SP System

Apart from its strengths and potential in modelling aspects of the human mind (1.9), it appears that, in more humdrum terms, the SP System has
several potential benefits and applications. These include:

- **Big data.** Somewhat unexpectedly, it has been discovered that the SP System has potential to help solve nine significant problems associated with big data [17]. 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 and more; the analysis of streaming data; compression of big data; model-based coding for efficient transmission of big data; potential gains in computational and energy efficiency in the analysis of big data; managing errors and uncertainties in data; and visualisation of structure in big data and providing an audit trail in the processing of big data.

- **Autonomous robots.** The SP System opens up a radically new approach to the development of intelligence in autonomous robots [16]:

- **An intelligent database system.** The SP System has potential in the development of an intelligent database system with several advantages compared with traditional database systems [13]. In this connection, the SP System has potential to add several kinds of reasoning and other aspects of intelligence to the ‘database’ represented by the World Wide Web, especially if the SP Machine were to be supercharged by replacing the search mechanisms in the foundations of the SP Machine with the high-parallel search mechanisms of any of the leading search engines.

- **Medical diagnosis.** The SP System may serve as a vehicle for medical knowledge and to assist practitioners in medical diagnosis, with potential for the automatic or semi-automatic learning of new knowledge [11];

- **Computer vision and natural vision.** The SP System opens up a new approach to the development of computer vision and its integration with other aspects of intelligence. It also throws light on several aspects of natural vision [15];

- **Neuroscience.** As outlined in [17] 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* ([20], [12, Chapter 11]). This has potential to illuminate aspects of neuroscience and to suggest new avenues for investigation.
• **Commonsense reasoning.** In addition to the previously-described strengths of the SP System in several kinds of reasoning, the SP System has strengths in the surprisingly challenging area of “commonsense reasoning”, as described by Ernest Davis and Gary Marcus [1]. How the SP System may meet the several challenges in this area is described in [19].

• **Other areas of application.** The SP System has potential in several other areas of application including [18]: the simplification and integration of computing systems; applications of natural language processing; best-match and semantic forms of information retrieval; software engineering [23]; the representation of knowledge, reasoning, and the semantic web; information compression; bioinformatics; the detection of computer viruses; and data fusion.

• **Mathematics.** The concept of IC via the matching and unification of patterns provides an entirely novel interpretation of mathematics [22]. This interpretation is quite unlike anything described in existing writings about the philosophy of mathematics or its application in science. There are potential benefits in science from this new interpretation of mathematics.

### 1.11 Unfinished business and the SP Machine

Like most theories, the SP Theory is not complete. Four pieces of ‘unfinished business’ are described in [14 Section 3.3]: the SP Computer Model needs to be generalised to include SP-patterns in two dimensions, with associated processing; research is needed to discover whether or how the SP concepts may be applied to the identification of low-level perceptual features in speech and images; more work is needed on the development of unsupervised learning in the SP Computer Model; and although the SP Theory has led to the proposal that much of mathematics, perhaps all of it, may be understood as IC [24], research is needed to discover whether or how the SP concepts may be applied in the representation of numbers. A better understanding is also needed of how quantitative concepts such as time, speed, distance, and so on, may be represented in the SP System.

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

More generally, a programme of research is envisaged, with one or more teams of researchers, or individual researchers, to create a more mature **SP Machine**, based on the SP Computer Model, and shown schematically in
A roadmap for the development of the SP Machine is described in [5].

Figure 6: Schematic representation of the development and application of the SP Machine. Reproduced from Figure 2 in [14], with permission.

A Information compression, inference, and probability

It has been recognised for some time that there is an intimate connection between IC and concepts of inference and probability [8, 9, 10, 11]. This may suggest that there is nothing to choose between IC and concepts of probability as a foundation for the development of the SP System, or any other artificial system that aspires to human-like intelligence.

For reasons outlined in the following subsections, there is an advantage in putting the main focus on IC.

A.1 Asymmetry between ICMUP and concepts of inference and probability

The very close connection between IC and concepts of inference and probability makes sense in terms of ICMUP because:
• **IC via unification of patterns.** The unification of patterns achieves compression of information.

• **Absolute probability.** Absolute probabilities may be derived from the frequencies of patterns, and, for a given pattern, its frequency of occurrence may be derived from the number of original patterns that have been unified to create that pattern.

• **Inference and conditional probability.** Inference may be achieved via partial matching as, for example, when seeing black clouds allows us to make the inference that rain is likely, via a partial match between ‘black clouds’ and the pre-established pattern ‘black clouds rain’. Conditional probabilities for inferences may be derived from the frequencies of occurrence of patterns.

As may be seen from points just made, there is an asymmetry between ICMUP and concepts of inference and probability: absolute and conditional probabilities may be derived from the matching and unification of patterns, but the reverse is not true. This is partly because, arguably, the matching and unification of patterns is more primitive than concepts of probability. But more to the point, values for probability, in themselves, have lost information about the matches and unifications that led to their creation.

Because probabilities may be derived from unifications but not the other way round, and because the matching and unification of patterns is prominent in HLPC [25], any artificial system that aspires to the generality of human intelligence should be founded on ICMUP, not concepts of probability.

### A.2 The discovery of structure

Perhaps because of the prominence of uncertainties in the way people think, much research in AI is based on concepts of probability, especially Bayes’ Theorem.

Much can be done with this kind of probabilistic approach to AI, but something is missing: it is assumed that all of the conceptual entities in a probabilistic analysis have been created already, and there is nothing about how they may be formed. By contrast, the matching and unification of patterns opens up the possibility of isolating words as discrete entities in speech [25, Section 15.1], and likewise for phrases [25, Section 15.2]. And it

---

1In brief, Bayes’ Theory may summarised with the equation $P(h|D) = \frac{P(D|h)P(h)}{P(D)}$, where $P(h)$ = prior probability of hypothesis $h$, $P(D)$ = prior probability of training data $D$, $P(h|D)$ = probability of $h$ given $D$, and $P(D|h)$ = probability of $D$ given $h$. 

---

16
can provide a basis for the building of three-dimensional models of entities, as outlined in \[15\] Sections 6.1 and 6.2.

**A.3 Extending the scope of inference and probability**

The matching and unification of patterns opens up some interesting aspects of probability which are not seen with approaches to probability which deal with pre-formed entities. By way of explanation, consider, firstly, that it is commonly assumed in ‘mainstream’ statistics that statistical significance can only be achieved when the frequencies with which phenomena occur are relatively high. For example:

“There is a definition of probability in terms of frequency that is sometimes usable. It tells us that a good estimate of the probability of an event is the frequency with which it has occurred in the past. This simple definition is fine in many situations, but breaks down when we need it most; i.e., its precision decreases markedly as the number of events in the past (the ‘sample size’) decreases. For sample sizes of 1 or 2 or none, the method is essentially useless.” [10, pp. 74–75].

Now consider, secondly, that very often, the need for large sample sizes does not apply with the matching and unification of patterns:

- For any given size of pattern, there is a minimum frequency below which no compression can be achieved. This is because each unified chunk requires some kind of label or identifier by which it can be referenced—and the information ‘cost’ of these labels offsets the compression achieved by the unification of matching patterns. This is true even if there is some kind of optimisation via Huffman coding or the like. Unless the compression is greater than the amount of information required for the labels, there is no net saving in the number of bits that are needed.

- In this connection, there is a trade-off between sizes of patterns and the minimum frequency that is needed for compression. With small patterns, high frequencies are required. But with large patterns, useful compression can be achieved when frequencies are as low as 2 or 3.

- As an example, one can, very often, recognise with high confidence a previously-heard song or other piece of music from hearing only a smallish sample of the piece. In such cases, our brains register that it
is very unlikely that there would be any other piece of music containing the sample we have heard. Accordingly, we assign a mental probability of 1.0 to the identification we have made, a probability which corresponds to a frequency of 2, because the first ‘learning’ hearing of the music yields a frequency of 1, and the second ‘recognition’ hearing yields a frequency of 2.

A.4 Probability and causation

Another apparent shortcoming with concepts of probability arises where we wish to know about causation. For example:

“The answer [to difficulties in solving causal problems with statistics ...] has to do with the official language of statistics—namely the language of probability. This may come as a surprise to some [people] but the word cause is not in the vocabulary of probability theory; we cannot express in the language of probabilities the sentence mud does not cause rain—all we can say is that the two are mutually correlated or dependent—meaning that if we find one, we can expect the other. Naturally, if we lack a language to express a certain concept explicitly, we can’t expect to develop scientific activity around that concept.” [6, p. 342].

Although no serious attempt has yet been made to examine issues in causality in terms of the SP Theory (but see [14, Section 10.5] and [12, Section 7.9]), there are reasons to think that it may be more successful than classical statistics. This is because:

- Judea Pearl, who has studied causality and its relation to statistics in great depth, has concluded that “[An engineering diagram] is, in fact, one of the greatest marvels of science. It is capable of conveying more information than millions of algebraic equations or probability functions or logical expressions. What makes [such a diagram] so much more powerful is the ability to predict not merely how the [system] behaves under normal conditions but also how [it] will behave under millions of abnormal conditions.” [6, p. 344].

- The SP System, because of its potential to build new structures (Appendix A.2), has the potential to build the kinds of structures referred to in the quote just given. This suggests that it may also have the potential make causal inferences that may be derived from such structures, as indicated in the quote.
References


