

Solutions to problems with deep learning

J Gerard Wolff*

March 14, 2018

Abstract

Despite the several successes of deep learning systems, there are concerns about their limitations, discussed most recently by Gary Marcus. This paper discusses Marcus's concerns and some others, together with solutions to several of these problems provided by the *SP theory of intelligence* and its realisation in the *SP computer model*. The main advantages of the SP system are: relatively small requirements for data and the ability to learn from a single experience; the ability to model both hierarchical and non-hierarchical structures; strengths in several kinds of reasoning, including 'commonsense' reasoning; transparency in the representation of knowledge, and the provision of an audit trail for all processing; the likelihood that the SP system could not be fooled into bizarre or eccentric recognition of stimuli, as deep learning systems can be; the SP system provides a robust solution to the problem of 'catastrophic forgetting' in deep learning systems; the SP system provides a theoretically-coherent solution to the problems of correcting over- and under-generalisations in learning, and learning correct structures despite errors in data; unlike most research on deep learning, the SP programme of research draws extensively on research on human learning, perception, and cognition; and the SP programme of research has an overarching theory, supported by evidence, something that is largely missing from research on deep learning. In general, the SP system provides a much firmer foundation than deep learning for the development of artificial general intelligence.

Key words: deep learning, artificial neural networks, the SP theory of intelligence, SP-multiple-alignment, information compression.

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

1 Introduction

Deep learning has received a great deal of attention, largely because of what it can do well, but there are concerns about its limitations, discussed most recently by Gary Marcus [10].

This short paper discusses Marcus’s concerns and some others, together with solutions to several of these problems provided by the *SP theory of intelligence* and its realisation in the *SP computer model* [19, 18], outlined in [25, Appendix A] with pointers to where fuller information may be found.¹

Several of these solutions have been described in [24, Section V]. References to this paper are made at appropriate points below.

2 What deep learning does well

As Marcus says: “Deep learning, as it is primarily used, is essentially a statistical technique for classifying patterns, based on sample data, using neural networks with multiple layers.” [10, p. 3]. Even in applications such as the playing of games, it seems that the primary function of deep learning is in the recognition of patterns.

3 Problems with deep learning and how they may be solved

In this section, the first 10 subheadings are the same as in [10], with the relevant section number shown at the end of each heading. The remaining headings are drawn largely from [24, Section V], unless they are already covered by the first 10 headings.

3.1 Deep learning thus far is data hungry (3.1)

The data-hungry nature of deep learning and how this may be overcome in the SP system has been discussed quite fully in [24, Section V-E]. There is also relevant discussion in [24, Section V-D].

In brief, the SP system, like a person, can learn from a single exposure or experience, and it can begin to form meaningful new structures and generalisations with exposure to a mere handful of other examples.

¹Publications in the SP programme of research, most with download links, may be found on www.cognitionresearch.org/sp.htm.

In this connection, the SP system is much more like a child than are deep learning systems: neuroscientist David Cox has been reported as saying: “To build a dog detector [with a deep learning system], you need to show the program thousands of things that are dogs and thousands that aren’t dogs. My daughter only had to see one dog.” and, the report says, she was happily pointing out puppies ever since.²

What about the slow learning of complex skills like speaking and understanding a new language or how to play a piano? With deep learning, it is assumed that this kind of slow acquisition of skills may be explained by the gradual strengthening of links in Hebbian-style learning. By contrast, in the SP system, the slow learning of complex skills may be explained by the complexity of the search that is required to find good structures.

In summary, the SP system can explain learning from a single exposure or experience and it can explain the slow learning of complex skills. By contrast, a deep learning system can explain the slow learning of complex skills, but it fails to explain how learning may be achieved with a single exposure or experience, and more generally, it is excessively demanding in its requirements for data.

3.2 Deep learning thus far is shallow ... (3.2)

In [10, Section 3.2], Marcus points out quite rightly that “it is important to realize that the word ‘deep’ in deep learning refers to a technical, architectural property (the large number of hidden layers used in a modern neural networks, where their predecessors used only one)” (p. 7).

He goes on to say that it is easy to over-interpret the results from a deep learning system. For example, “according to a widely-circulated video of the system learning to play the brick-breaking Atari game Breakout, ‘after 240 minutes of training, [the system] realizes that digging a tunnel through the wall is the most effective technique to beat the game’. But the system has learned no such thing; it doesn’t really understand what a tunnel, or what a wall is; it has just learned specific contingencies for particular scenarios.” (p. 8).

The SP system certainly does not yet provide a comprehensive solution to issues like those just described. In brief, it seems fair to summarise the strengths and potential of the SP system, and to compare it with deep learning systems, as follows:

- The SP system has strengths in the representation of diverse kinds of

²“Inside the moonshot effort to finally figure out the brain”, *MIT Technology Review*, 2017-10-12, bit.ly/2wRxsOg.

knowledge, in diverse aspects of intelligence, and in the seamless integration of diverse kinds of knowledge and diverse aspects of intelligence, in any combination. These strengths, which all flow from the powerful concept of *SP-multiple-alignment*, are summarised in [25, Sections 3, 4, and 5], with pointers to where fuller information may be found.

- Although the SP system has strengths in the representation of diverse kinds of knowledge, it seems likely that more research will be required to understand how the system may learn and represent the great range of concepts employed by people. There is some discussion in [20, Sections 6.1 and 6.2] about how the system may develop a concept of a three-dimensional object, and in [20, Section 5.3], there is brief discussion of the development of concepts like motion and speed.
- At a ‘deep’ level, it seems likely that *all* kinds of learning, both in deep learning systems and in the SP system, may be understood as the learning of statistical contingencies.

3.3 ... And has limited capacity for transfer (3.2)

A point that is related to the shallowness of much learning by deep learning systems (Section 3.2) is that such systems appear to have limited capacities for the transfer of learning from one situation to another:

“*Transfer tests*—in which the deep reinforcement learning system is confronted with scenarios that differ in minor ways from the one ones on which the system was trained show that deep reinforcement learning’s solutions are often extremely superficial.” ([10, Section 3.2], emphasis in the original).

That said, it appears that the ‘IMPALA’ system [6] “exhibits positive transfer between tasks” (Abstract).

Although learning in the SP system is not yet well enough developed to prove the point, it seems likely that the system will have relatively robust abilities to transfer learning from one situation to another:

- Since new learning does not disturb old learning in the SP system (Section 3.12, below), the learning of a concept like ‘ball’ in a game like tennis, is likely to transfer relatively directly to learning about a game like baseball.

- Even in cases where the ball in one sport is rather different from the ball in another—such as a ball in tennis compared with a ball in football—the system is designed to abstract the elements that are the same and to record the difference, so that the two kinds of ball may be recognised as being similar, and, at the same time there will be no risk of confusing the two kinds of ball in the two kinds of game.

3.4 Deep learning thus far has no natural way to deal with hierarchical structure (3.3)

Computer models that I developed in a programme of research on language learning (summarised in [19]) were designed to work by ‘hierarchical chunking’. As one might expect, these models were good at representing hierarchical structures.

But in the ‘SP’ programme of research—where the aim has been to simplify and integrate observations and concepts across artificial intelligence, mainstream computing, mathematics, and human learning, perception, and cognition—hierarchical chunking would not do. The challenge has been to discover or create a framework that would serve equally well for the representation of both hierarchical and non-hierarchical structures.

What has proved to be a good solution is the concept of *SP-multiple-alignment*, borrowed and adapted from the concept of ‘multiple sequence alignment’ in bioinformatics. Although the basic idea is to create alignments of two or more sequences,³ the framework lends itself very well to the representation of the kinds of hierarchical structures recognised in linguistic analysis, as can be seen in Figure 1.

There is more discussion in [24, Section V-K].

3.5 Deep learning thus far has struggled with open-ended inference (3.4)

No work has yet been done to explore whether or how the SP system can model ‘open ended’ inferences but, unlike deep learning systems, it has strengths in several different forms of reasoning 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

³It is envisaged that at some stage the SP system will be adapted work with two-dimensional patterns as well as on-dimensional sequences.

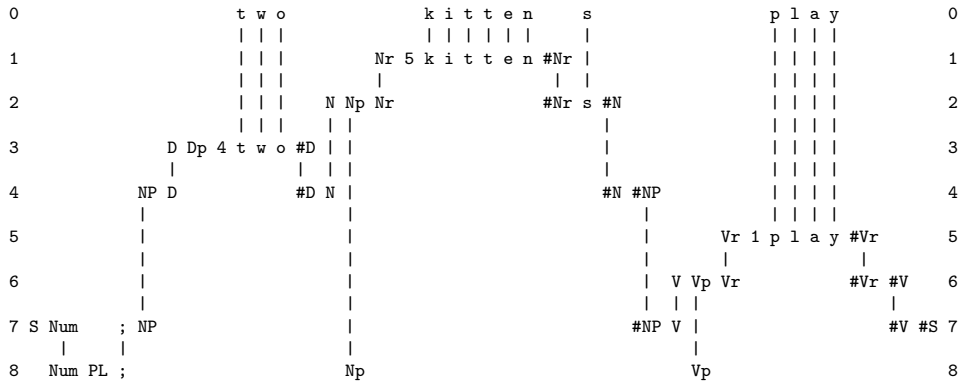


Figure 1: The best SP-multiple-alignment created by the SP computer model with a store of SP-patterns like those in rows 1 to 8 (representing grammatical structures, including words) and an SP-pattern representing a sentence to be parsed shown in row 0.

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 ([18, Chapter 7], [19, Section 10]). Where it is appropriate, probabilities for inferences may be calculated in a straightforward manner ([18, Section 3.7], [19, Section 4.4]). There is also potential for spatial reasoning [21, Section V-F.1], and for what-if reasoning [21, Section V-F.2].

There is more discussion in [24, Section V-L].

3.6 Deep learning thus far is not sufficiently transparent (3.5)

Marcus [10, Section 3.5] writes that “Although some strides have been [made] in visualizing the contributions of individuals nodes in complex networks ..., most observers would acknowledge that neural networks as a whole remain something of a black box.” (p. 11).

In this respect, there is a sharp contrast with the SP system [24, Section V-J]:

- All knowledge stored by the SP system is transparent and open to inspection.
- In general, knowledge in the SP system is likely to be comprehensible by people but problems may arise with concepts that have not yet been well studied in the SP programme of research (Section 3.2).

- There is an audit trail for all processing performed by the SP system and all conclusions it may reach.

3.7 Deep learning thus far has not been well integrated with prior knowledge (3.6)

Marcus [10, Section 3.6] writes that “Work in deep learning typically consists of finding a training database, sets of inputs associated with respective outputs, and learn all that is required for the problem by learning the relations between those inputs and outputs, using whatever clever architectural variants one might devise, along with techniques for cleaning and augmenting the data set. With just a handful of exceptions, ..., prior knowledge is often deliberately minimized.” (p. 11).

Later, he writes that “It also not straightforward in general how to integrate prior knowledge into a deep learning system:, in part because the knowledge represented in deep learning systems pertains mainly to (largely opaque) correlations between features, rather than to abstractions like quantified statements (e.g. all men are mortal), see discussion of universally-quantified one-to-one-mappings in Marcus (2001), or generics” (p. 11).

Similar things may be said about the SP system but, here, the problems are likely to be less severe. This is because, in general, prior knowledge may be represented in the same format as knowledge that the system learns for itself.

Later in the same section, Marcus writes:

“Problems that have less to do with categorization and more to do with commonsense reasoning essentially lie outside the scope of what deep learning is appropriate for, and so far as I can tell, deep learning has little to offer such problems. In a recent review of commonsense reasoning, Ernie Davis and I [5] began with a set of easily-drawn inferences that people can readily answer without anything like direct training, such as *Who is taller, Prince William or his baby son Prince George? Can you make a salad out of a polyester shirt? If you stick a pin into a carrot, does it make a hole in the carrot or in the pin?*

“As far as I know, nobody has even tried to tackle this sort of thing with deep learning.

“Such apparently simple problems require humans to integrate knowledge across vastly disparate sources, and as such are a long

way from the sweet spot of deep learning-style perceptual classification. Instead, they are perhaps best thought of as a sign that entirely different sorts of tools are needed, along with deep learning, if we are to reach human-level cognitive flexibility.” (p. 12).

It appears that the SP system has considerable potential with the kinds of commonsense reasoning discussed by Davis and Marcus [5]. This is described in [22], a detailed response to the issues raised in their paper.

3.8 Deep learning thus far cannot inherently distinguish causation from correlation (3.7)

Marcus [10, Section 3.7] writes: “If it is a truism that causation does not equal correlation, the distinction between the two is also a serious concern for deep learning. Roughly speaking, deep learning learns complex correlations between input and output features, but with no inherent representation of causality.” (p. 12–13).

It cannot be claimed that there is a comprehensive analysis, within the SP system, of the difference between causation and correlation. But the system does support useful demonstrations of how such concepts may be modelled in the system:

- *Causation.* In [19, Section 10.5] and [18, Section 7.9], there is a demonstration of how the SP system may serve in the causal diagnosis of faults in an electronic circuit.
- *Correlation.* In [19, Section 10.2] and [18, Section 7.8], there is an example showing how Bayesian reasoning, with conditional probabilities centre stage, may be modelled in the SP system.

3.9 Deep learning presumes a largely stable world, in ways that may be problematic (3.8)

Marcus [10, Section 3.8] writes: “The logic of deep learning is such that it is likely to work best in highly stable worlds, like the board game Go, which has unvarying rules, and less well in systems such as politics and economics that are constantly changing.” (p. 13).

Much the same may be said of the SP system, or any other system that models the world via its statistical structure, which is the rule for most AI systems.

The trick, of course, for people and for artificial systems, is to be prepared for constant changes in the phenomena that one is modelling. And with predictions in areas such as economics, there is the possibility that each prediction itself may change what it is that is being predicted!

3.10 Deep learning thus far works well as an approximation, but its answers often cannot be fully trusted (3.9)

Marcus [10, Section 3.9] notes that there are now many examples where deep learning systems are fooled into making misclassifications which appear eccentric or bizarre to people. In [24, Section V-G] I have described similar examples.

In the latter section I have written:

“With regard to the first kind of error—failing to recognise something that is almost identical to what has been recognised—there is already evidence that the SP computer model would not make that kind of mistake. It can recognise words containing errors of omission, commission and substitution [18, Section 6.2.1], and likewise for diseases in medical diagnosis viewed as pattern recognition [17, Section 3.6] and in the parsing of natural language [2, Section 4.2.2].

“No attempt has been made to test experimentally whether or not the SP computer model is prone to the second kind of error—recognising abstract patterns as ordinary objects—but a knowledge of how it works suggests that it would not be.”

In general, it seems that the SP system is unlikely to make the strange errors of classification to which deep learning systems are prone.

3.11 Deep learning thus far is difficult to engineer with (3.10)

Marcus [10, Section 3.10] writes:

“Another fact that follows from all the issues raised above is that is simply hard to do robust engineering with deep learning. As a team of authors at Google put it in 2014, in the title of an important, and as yet unanswered essay [13], machine learning is ‘the

high-interest credit card of technical debt’, meaning that is comparatively easy to make systems that work in some limited set of circumstances (short term gain), but quite difficult to guarantee that they will work in alternative circumstances with novel data that may not resemble previous training data (long term debt, particularly if one system is used as an element in another larger system).

“In an important talk at ICML, Leon Bottou [4] compared machine learning to the development of an airplane engine, and noted that while the airplane design relies on building complex systems out of simpler systems for which it was possible to create sound guarantees about performance, machine learning lacks the capacity to produce comparable guarantees. As Google’s Peter Norvig [11] has noted, machine learning as yet lacks the incrementality, transparency and debuggability of classical programming, trading off a kind of simplicity for deep challenges in achieving robustness.

“Henderson and colleagues have recently extended these points, with a focus on deep reinforcement learning, noting some serious issues in the field related to robustness and replicability [7].” (p. 14).

Evidence to date suggests that these remarks are unlikely to apply to the SP system. Although development of the system has taken a long time, the SP computer model now exhibits considerable stability and robustness, and promises to provide a sound basis for scaling up with parallel processing, and for further developments as noted in [19, Section 3.3].

3.12 Catastrophic forgetting

A weakness of deep learning that was overlooked in the writing of [24] is a problem called “catastrophic forgetting”, meaning the way in which new learning in a deep learning system wipes out old memories. A solution has been proposed in [9] but it appears to be partial, and it is unlikely to be satisfactory in the long run.

The SP system is immune to any such influence in its learning. All learning is achieved by the addition of new SP-patterns to the system’s store of SP-patterns, and, while there may be some merging of new information with old information that is similar, new learning does not disturb old learning.

Of course, there may be a case for introducing some kind of forgetting into the system, but the system as it is now does not forget.

3.13 Under- and over-generalisations and their correction, and learning from ‘dirty data’

A general problem in any kind of learning system, especially the learning of a natural language, is how to generalise ‘correctly’ from the finite sample of information which is the basis for learning—avoiding both over-generalisations and under-generalisations [10, Section 3]. A related problem is learning from ‘dirty data’: how to learn ‘correct’ structures despite the fact that most samples of data contain ‘errors’—and this in the face of evidence that learning may be successful without the opportunity for children to receive correction from adults or older children.

Several solutions have been proposed for deep learning systems, some of which are referenced in [24, Section V-N]. But I believe it is fair to say that none of them are entirely satisfactory.

What I believe is a much better solution has been described in [19, Section 5.3] and in [24, Section V-N]. In brief:

1. Given a finite sample of data, \mathbf{I} , compress it as much as possible to yield a *grammar*, \mathbf{G} , and an *encoding*, \mathbf{E} , of \mathbf{I} in terms of \mathbf{G} .
2. In general, it will be found that \mathbf{G} represents the ‘essence’ of \mathbf{I} , without over- or under-generalisations, and without ‘errors’ from the ‘dirty data’.

There is already evidence in support of the framework outlined here. I believe, with further research, it will prove to be sound.

3.14 Avoiding reliance on human intelligence

Although it is often not mentioned in news reports about deep learning, that kind of system often relies on people and their intelligence to prepare the data which the AI system uses for learning, in a form that by-passes weaknesses in the AI system.⁴

Although the SP system falls short of human-level intelligence, it differs from deep learning systems in its founding on a coherent body of theory, and its versatility, which promise to deliver an AI machine with robust, human-level intelligence, with little or no need for the kind of assistance which is currently needed by deep learning systems.

⁴See, for example, “Artificial intelligence has a dirty little secret: It’s powered by people”, *Stars & Stripes*, 2018-03-05, bit.ly/2pb1hrv.

3.15 Psychology and neuroscience

It has long been recognised that deep learning systems are only loosely related to any kind of structure or processing in the brain. More generally, research on deep learning has been conducted largely without reference to what has been learned about human learning, perception, and cognition, or what is known about neuroscience.

By contrast, the SP system draws extensively on my own background in cognitive psychology, and my extensive programme of research on the learning of a first language or languages by children (summarised in [16]).

From its beginnings, the SP programme of research has been influenced by a long-running theme, beginning with research by Fred Attneave [1], Horace Barlow [2, 3] and others, pointing to the importance of information compression in human learning, perception, and cognition.

It is also relevant to mention that abstract structures and processes in the SP system, map quite neatly on to what appear to be plausible structures of neurons and their interconnections, and how they may function, described in [23].

3.16 Overarching theory

A variety of sources, perhaps most notably the work of Ray Solomonoff [14, 15] point to the importance of information compression in learning.

In research on deep learning in artificial neural networks, well reviewed by Jürgen Schmidhuber [12], there is some recognition of the importance of information compression [12, Sections 4.2, 4.4, and 5.6.3], but it appears that the idea is not well developed in deep learning systems.

By contrast, as readers may have learned from [19, 18], and may guess from Section 3.13, the SP system is devoted to the compression of information, and more precisely compression of information via the powerful concept of SP-multiple-alignment, illustrated in Figure 1. There is much evidence in support of this theory presented in [19, 18] and elsewhere.

In general, the SP system has a coherent overarching theory, something which is largely missing from research in deep learning.

4 Potential risks of excessive hype

In [10, Section 4], Marcus writes:

“My own largest fear is that the field of AI could get trapped in a local minimum, dwelling too heavily in the wrong part of

intellectual space, focusing too much on the detailed exploration of a particular class of accessible but limited models that are geared around capturing low-hanging fruit—potentially neglecting riskier excursions that might ultimately lead to a more robust path.”

This chimes very much with my own experience. Despite many respectable publications in the SP programme of research, and many useful results, it has proved very difficult to get a hearing for this work by those engaged in research on deep learning. It seems that the extraordinary enthusiasm for deep learning, perhaps coupled with the large amounts of money being channelled into this area, has made it very difficult for researchers to divert any of their attention to anything other than deep learning. One senior researcher, who was kind enough to reply to one of my emails, said that although the SP research may be interesting, he does not have the time to look at it.

Good science and engineering is not like this. To solve difficult problems it is necessary to maintain several paths through the search space, and for researchers in any one area to be prepared to keep abreast of developments in other areas. In keeping with that approach, the central aim of the SP programme of research is simplification and integration of observations and concepts across artificial intelligence, mainstream computing, mathematics, and human learning, perception, and cognition.

Overspecialisation is a phenomenon noted by John Kelly and Steve Hamm, both of IBM, in their book *Smart Machines* [8]. In connection with research on different sensory modalities they write:

“In order to make this great leap and become true thinking machines, the cognitive systems of the future will integrate information from multiple sensing technologies. Today, as scientists labor to create machine technologies to augment our senses, there’s a strong tendency to view each sensory field in isolation as specialists focus only on a single sensory capability. Experts in each sense don’t read journals devoted to the others senses, and they don’t attend one another’s conferences. Even within IBM, our specialists in different sensing technologies don’t interact much. Yet if machines are to help humans understand the world, they have to make sense of it and communicate about it in a way that’s familiar and comprehensible to humans. This integration of data from various sensing technologies is beginning to happen in multimedia and visual analytics, where vision and sound are

correlated. But that’s just the start of what will be required in the next era of computing. (p. 74).

5 What would be better?

In [10, Section 5], Marcus writes: “Despite all of the problems I have sketched, I don’t think that we need to abandon deep learning. Rather, we need to reconceptualize it: not as a universal solvent, but simply as one tool among many, a power screwdriver in a world in which we also need hammers, wrenches, and pliers, not to mention chisels and drills, voltmeters, logic probes, and oscilloscopes.” (p. 18).

Yes, of course, in the spirit of maintaining several paths through the search space, it would be wrong to abandon all research on deep learning. But there is certainly a need to open up other areas and I believe the SP framework is one of them.

Regarding Marcus’s remarks about symbolic and sub-symbolic systems [10, Section 5.2], I believe the SP system bridges that divide. In principle, it may work at any level of granularity.

In connection with this: “The power and flexibility of the brain comes in part from its capacity to dynamically integrate many different computations in real-time. The process of scene perception, for instance, seamlessly integrates direct sensory information with complex abstractions about objects and their properties, lighting sources, and so forth.” (p. 20), a major strength of the SP system, due largely to the powerful concept of SP-multiple-alignment, is the ability of the system to integrate diverse kinds of knowledge and diverse aspects of intelligence, in any combination [25, Section 5].

As may be seen from Section 3.15, I agree very much that we should build “models that are motivated not just by mathematics but also by clues from the strengths of human psychology.” (p. 21).

6 Conclusion

The gist of this paper is that, while research on deep learning should certainly continue, the SP programme of research also merits attention. In that connection, the SP system has several advantages compared with deep learning systems. The main ones are:

- *Quantities of data and one-trial learning (Section 3.1).* By contrast with deep learning systems, the SP system can produce meaningful results with quite small amounts of data. It provides a model for the

way in which people can learn from a single exposure or experience. At the same time it provides an explanation for why it takes time to learn complex skills.

- *Hierarchical and non-hierarchical structures (Section 3.4)*. Unlike deep learning systems—which do not lend themselves well to the representation of hierarchical structures—the SP system, via the concept of SP-multiple-alignment, accommodates such structures very well. At the same time, it also provides for the representation of non-hierarchical structures.
- *Reasoning (Sections 3.5 and 3.7)*. Although no attempt has yet been made to explore whether or how the SP system may perform ‘open ended’ inference, the SP system—unlike deep learning systems—has strengths in several different kinds of reasoning. It also has strengths in ‘commonsense’ reasoning, as described in [22].
- *Transparency (Section 3.6)*. By contrast with deep learning systems, the SP system provides complete transparency in the way in which it represents knowledge and much of the knowledge is likely to be comprehensible by people. The SP system also provides a full audit trail for all its processing.
- *It is unlikely that an SP system would be easily fooled (Section 3.10)*. Deep learning systems can misclassify stimuli in ways that people find eccentric or bizarre. Experience with the SP system, and a knowledge of how it works, suggests that it would be unlikely to make that kind of error.
- *Catastrophic forgetting (Section 3.12)*. A striking weakness of deep learning systems is ‘catastrophic forgetting’—the way in which new learning wipes out old learning. The SP system does not suffer from this problem because new learning does not disturb old learning.
- *Correction of over- and under-generalisations, and learning from ‘dirty data’ (Section 3.13)*. With deep learning systems, a variety of solutions have been proposed for how to generalise correctly from a sample of data, without over- or under-generalisation, but it appears that none of them are entirely satisfactory. By contrast, the SP system proposes a theoretically-coherent solution that flows from the core theory in the system. Those same principles provide an explanation of how learning of ‘correct’ knowledge can be successful, despite errors in the data which is the basis for learning.

- *Psychology and neuroscience (Section 3.15)*. By contrast with research on deep learning systems, which is largely independent of what is known about human learning, perception, and cognition, and neuroscience, the SP system draws extensively on research in those areas. The SP system also suggests how abstract structures and processes in the system may be realised in terms of neurons and their interconnections.
- *Overarching theory (Section 3.16)*. The central principle in the SP theory, supported by much evidence, is that much of human learning, perception, and cognition, may be understood as compression of information via the concept of SP-multiple-alignment. By contrast, deep learning systems have little or no over-arching theory.

In general, I believe that SP system provides a much firmer foundation than deep learning for the development of artificial general intelligence.

References

- [1] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954.
- [2] H. B. Barlow. Sensory mechanisms, the reduction of redundancy, and intelligence. In HMSO, editor, *The Mechanisation of Thought Processes*, pages 535–559. Her Majesty’s Stationery Office, London, 1959.
- [3] H. B. Barlow. Trigger features, adaptation and economy of impulses. In K. N. Leibovic, editor, *Information Processes in the Nervous System*, pages 209–230. Springer, New York, 1969.
- [4] L. Bottou. Two big challenges in machine learning. In *Proceedings from 32nd International Conference on Machine Learning*, 2015.
- [5] E. Davis and G. Marcus. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9):92–103, 2015.
- [6] L.e Espeholt, H. Soyer, R. Munos, K. Simonyan, V. Mnih, T. Ward, Y. Doron, V. Firoiu, T. Harley, I. Dunning, S. Legg, and K. Kavukcuoglu. Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. Technical report, DeepMind Technologies, London, 2018. arXiv:1802.01561v2 [cs.LG].

- [7] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. Deep reinforcement learning that matters. Technical report, McGill University and Microsoft Maluuba, 2017. arXiv, cs.LG.
- [8] J. E. Kelly and S. Hamm. *Smart machines: IBM’s Watson and the era of cognitive computing*. Columbia University Press, New York, Kindle edition, 2013.
- [9] J. Kirkpatrick. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, 114(13):3521–3526, 2017.
- [10] G. Marcus. Deep learning: a critical appraisal. Technical report, NYU Psychology, 2018. arXiv:1801.00631v1 [cs.AI].
- [11] P. Norvig. State-of-the-art AI: building tomorrow’s intelligent systems. In *Proceedings from EmTech Digital, San Francisco*, 2016.
- [12] J. Schmidhuber. Deep learning in neural networks: an overview. *Neural Networks*, 61:85–117, 2015.
- [13] D. Sculley, T. Phillips, D. Ebner, V. Chaudhary, and M. Young. Machine learning: The high-interest credit card of technical debt. In *Proceedings from SE4ML: Software Engineering for Machine Learning (NIPS 2014 Workshop)*, 2014.
- [14] R. J. Solomonoff. A formal theory of inductive inference. Parts I and II. *Information and Control*, 7:1–22 and 224–254, 1964.
- [15] R. J. Solomonoff. The discovery of algorithmic probability. *Journal of Computer and System Sciences*, 55(1):73–88, 1997.
- [16] J. G. Wolff. Learning syntax and meanings through optimization and distributional analysis. In Y. Levy, I. M. Schlesinger, and M. D. S. Braine, editors, *Categories and Processes in Language Acquisition*, pages 179–215. Lawrence Erlbaum, Hillsdale, NJ, 1988. bit.ly/ZIGjyc.
- [17] J. G. Wolff. Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, unification and search. *Decision Support Systems*, 42:608–625, 2006. bit.ly/1F366o7, arXiv:1409.8053 [cs.AI].

- [18] J. G. Wolff. *Unifying Computing and Cognition: the SP Theory and Its Applications*. CognitionResearch.org, Menai Bridge, 2006. ISBNs: 0-9550726-0-3 (ebook edition), 0-9550726-1-1 (print edition). Distributors, including Amazon.com, are detailed on bit.ly/WmB1rs.
- [19] J. G. Wolff. The SP theory of intelligence: an overview. *Information*, 4(3):283–341, 2013. bit.ly/1NOMJ6l, arXiv:1306.3888 [cs.AI].
- [20] J. G. Wolff. Application of the SP theory of intelligence to the understanding of natural vision and the development of computer vision. *SpringerPlus*, 3(1):552–570, 2014. bit.ly/2oIpZB6, arXiv:1303.2071 [cs.CV].
- [21] J. G. Wolff. Autonomous robots and the SP theory of intelligence. *IEEE Access*, 2:1629–1651, 2014. bit.ly/18DxU5K, arXiv:1409.8027 [cs.AI].
- [22] J. G. Wolff. Commonsense reasoning, commonsense knowledge, and the SP theory of intelligence. Technical report, CognitionResearch.org, 2016. Submitted for publication. bit.ly/2eBoE9E, arXiv:1609.07772 [cs.AI].
- [23] J. G. Wolff. Information compression, multiple alignment, and the representation and processing of knowledge in the brain. *Frontiers in Psychology*, 7:1584, 2016. bit.ly/2esmYyt, arXiv:1604.05535 [cs.AI].
- [24] J. G. Wolff. The SP theory of intelligence: its distinctive features and advantages. *IEEE Access*, 4:216–246, 2016. bit.ly/2qgq5QF, arXiv:1508.04087 [cs.AI].
- [25] J. G. Wolff. Strengths and potential of the SP theory of intelligence in general, human-like artificial intelligence. Technical report, CognitionResearch.org, 2017. bit.ly/2z3rM8b, viXra:1711.0292, hal.archives-ouvertes.fr/hal-01633376.