Title: Grandfather of AI not

Subtitle: Yet still able to infer LLMs are not AGI

Tags: AI, LLM, AGI, X4, Inference

Grandfather of AI not

When I showed my youngest son the first cut of my latest video he said, “oh, I see what you're up to. You know, you're never going to convince anybody that you're the grandfather of AI”. And of course, he's right.

But the damage that he does to my ego is more than compensated for by the pride I have in his capacity for critical thinking. Anyway, although I'm no Geoffrey Hinton I think I can lay claim to having had a little insight into artificial intelligence (AI) that anticipated some key aspects of recommendation algorithms and large language models (LLMs). These insights, based on findings from psychology, have not yet been fully exploited but could take us further towards artificial general intelligence (AGI), especially when combined with some of the virtues of both machine learning and symbolic approaches.

To explain this, I take you on something of a historical tour, both of the transition from symbolic to machine learning approaches to AI and of my personal journey of using findings from psychology to inform the development of AI systems. I also spend some time noting the differences between human and machine intelligence, particularly large language models (LLMs). This takes stock of some of the major gaps between them and indicates directions to take in filling these gaps.

I only came to the full realisation that I had something to say on this subject recently when I moved house. I was unpacking boxes and came across a manual that I wrote and marketed in 1984 for a computer program that experimented with models of human inference. I was astonished!

Let me give you an example of what I mean by human inference. Let's say that I have a friend who has just come out of prison. Close your eyes for a moment. What image does this conjure up? Now, without saying any more you will probably infer all sorts of things about them. Some of these inferences may pop into consciousness while many may lie lurking under the surface. You may, for example, see in your mind's eye a young man in his teens or 20s, looking a bit dishevelled and probably untrustworthy. Alternatively, you might think of a specific, famous or fictional criminal you heard about recently or someone you know. The inferences you make about my friend won't stop there. You may make all sorts of inferences about their background, whether they have a partner, whether they have children, their education, how attractive and intelligent they are, and so on.

When I tell you that my friend is a 68-year-old woman who was put in prison because she refused to pay her TV license, you may be surprised. And this is understandable because my friend would not fit your stereotyped impression of a criminal. You may also be surprised that the criminal justice system would do such a thing. And probably they would not do it now. They have less time and probably more humanity.

In 1984, if you were to ask people working in artificial intelligence to write a program that would identify criminals from their characteristics, they would have probably gone about it by writing a set of rules - rules like 'IF male, young, uneducated and from a poor background THEN possibly criminal'. Alternatively, they may have produced a decision tree that would capture much the same logic. Others might have used 'Bayes' theorem', a probabilistic approach, that would update beliefs about whether a person might be in a particular category like 'criminal' or not as new evidence accumulated. There were some major problems with this approach.

Formally Defined Representations

One major problem comes about because these approaches tend to be based on formally defined representations. An assumption in logic and mathematics is that the symbols used (i.e. numbers, logical operators and so forth) have strictly defined meanings and that these meanings are the same independent of context. The number 7 is the number 7 independent of context whereas 'criminal' is not (differing, for example, on time in history and legal jurisdiction, let alone your personal views).

This is not true of words. Words like 'young', 'uneducated' and 'poor' are at least relative to context, and other words like 'attractive' are largely subjective. Nevertheless, the rules that were created treated these words like mathematical or logical symbols with well-defined and fixed meanings. When you had a rule like 'IF red spots THEN suspect measles' and asked a doctor to precisely define what they meant by 'red' or 'spots' they tended to flounder. What shade of red? How many spots? What size and density of spots? The answers tended to vary depending on the case. There were no necessary and defining features that would imply a conclusion. Rather it was a pattern of symptoms that would more or less match this disease or that one. This was a well-documented issue known as the 'knowledge-elicitation problem'.

And that, of course, is where these symbolic approaches to AI became unstuck. You can't divide the world into 'criminals' and 'not criminals'. You could say it was 'criminal' to try, but they wouldn't lock you up for it. In fact, given changes in culture, you can hardly divide the world into the distinct categories of male and female. Words are slippery.

I am reminded of a story that I often tell about my eldest son. At about the age of 7, he came up to me and said 'Dad, can I ask you about the future?'. I was somewhat taken aback but thought he must be referring to a discussion we'd recently had about stars and planets. Then he said, 'You know that picture on the wall, and the piano'. 'Yes', I said and then it dawned on me. 'Sam', I asked, 'do you want to talk about the future or the furniture?'. 'Ah, yes', he said, 'that's it – the furniture'. What he was puzzled about was what exactly made something a piece of furniture. It turns

out that this is an issue that pervades not just language but is at the very heart of human inference and cognition. It has also attracted much thought by philosophers.

Wittgenstein had argued in the 1930s and 40s (published in Philosophical Investigations in 1953) that objects and concepts as commonly used in language do not in general have necessary and sufficient features by which they could be defined. Anything you can sit on, you can say is a 'chair', and what we think of as a 'chair' can be used for purposes other than sitting on (like standing on or putting bags on). It is difficult the find characteristics that are true of all chairs and not true of things that are not chairs. The issue of what is furniture is an even further level of abstraction.

Instead of necessary and sufficient features, Wittgenstein argued that what we think of as being in the same category, like, say, chairs, criminals or games have what he termed 'family resemblance'. Chairs in general may have many attributes. Any particular chair will have some, and another chair may have others, and some attributes may overlap, but no single set of attributes definitively defines 'chair'. This is a much more fuzzy view of the world than classical logic is equipped to deal with.

Part of the problem was that logic already had a robust approach to inference. In logic, if we say that 'all men are mortal' and 'Socrates is a man', then it follows that 'Socrates is mortal'. We say that this must be true, or is proven. This is an example of deductive logic. Its appeal is the certainty that the conclusion follows from the premises. The mistake of symbolic approaches to AI was to be lured into thinking that this kind of deductive logic could be applied to all the complexities of the real world, and that just because you could prove Socrates is mortal you could equally infer that red spots reliably indicated measles. The desire for certainty and simplicity blinded the school of symbolic AI to the complexity and subtlety of a real world and the way in which language only loosely maps on to it.

Filtering the Evidence

When we bring a mindset of wanting to be 'correct' and 'certain', then symbolic representations have a strong appeal. This is what philosophers call a 'rationalist' approach. It maintains that knowledge is primarily derived from logic and reason. Not only will derived knowledge be certain but it should also be untainted by bias. There is a tendency to 'prescribe' rather than 'describe'. The rationalist approach steers towards what should be the case rather than what is the case. It also provides a theoretical framework into which to fit the facts.

By contrast, an empirical approach, collects the facts first and then constructs a theory to account for them. Knowledge comes from observation and sensory experience as opposed to being derived. I notice that I am using terms like evidence, facts and data interchangeably. Let's leave aside for now that, just like other words and concepts, the meanings of these terms are not fixed.

There is now a considerable body of findings in psychology to show that we are not neutral with respect to evidence. We not only cherry pick evidence but we also interpret it through lenses that are heavily and inevitably biased.

The rationalist tendency is not confined to logic and maths. All theories whether they are logical, medical, anthropological, economic, personal or whatever provide a framework and the overwhelming temptation is to use the framework to structure the facts. This is the case even when the theory has been developed empirically. We just can't help filtering and interpreting the 'evidence' through the theoretical frameworks that we have often spent many years learning and developing. This is what philosophers call 'theory-laden observation'. It's not that these frameworks cannot be useful. After all they may have persisted because they have been useful. It's more that they tend to fix your assumptions and blind you to alternative ways of understanding the evidence. A theory, even if it is derived empirically, once it becomes 'received wisdom' in society generally, or in your own head, can become akin to a rationalist straitjacket.

There is an approach in psychotherapy called 'radical presence' (McNamee, 2015) where the therapist tries to drop all their theoretical training and simply listen to what their clients are saying and understand them on their own terms. It is often an approach that deals better with the complexity of the real world than trying to fit everything into a pre-existing structure. In a way this is what machine learning and neural net technologies in particular do. They construct a model from their training set in a way that is relatively atheoretical.

Bridging between Evidence and Conclusions

Another major problem common to logical, symbolic and Bayesian approaches, is that they attempt to bridge between evidence and conclusion in discrete steps. Collect the evidence, feed it into a system, and out pops a conclusion. Nothing inherently wrong with that. After all, it is the conclusion that you want. The problem with this approach lies in the way in which the evidence and the conclusion are related to each other. The process of going from evidence to conclusion does not align with the way that people characteristically make spontaneous inferences. They can make inferences this way, but for the most part they do not.

The symbolic school of AI made the default assumptions about the process of human inference - not so much where it ends up as the route it takes to get there. Anybody who has phoned an organisation and encountered a voice response system that offers menu options, knows that going from evidence to conclusion in a series of independent sequential decisions is fraught with difficulties. An implicit assumption of the symbolic approach was that the inference proceeds in this kind of sequential manner from premises/evidence/data to conclusions. In a rule-based system, for example, the inference mechanism would move in discrete logical steps towards its conclusion – IF red spots THEN measles.

Another common technique was to start with the conclusion – i.e. diagnose measles (known as backward chaining), as opposed to starting with symptoms and seeing where they lead. The epitome of this approach was use of Prolog, a logic programming language, where everything was structured around proving a proposition. I spent several years trying to get Prolog to do things that could have been done in more conventional programming languages in weeks or days. I would say that this use of Prolog on various government-funded programmes during the late 1980s may have significantly slowed the progress of AI for practical applications

in some research labs. But who was to know? – that is 'blue skies research' for you. Most of the time it doesn't work out as you hoped.

Logic and symbolic AI do not capture how people tend to make inferences. Rather than working step by step, a doctor, for example, will tend to match whole patterns. They process the input in parallel rather than in sequence. Many of the cues are not medical at all. An older woman walks into the surgery with a slight limp and wearing old clothes and already this simple pattern alerts the doctor to some likely and unlikely possibilities. The patient's demographics (gender, age, locality) and the way they present, already considerably narrow down the possibilities even before any symptoms are discussed. The doctor has seen so many patients that this information is taken in at a glance and inferences from this are automatic. It hardly touches the sort of high-level conscious processing that symbolic AI tried to capture. A doctor I once interviewed described this process as much more like recognising a picture than making a deduction. This is both parallel rather than sequential and more passive recognition rather than active recall. Chess grand masters have been shown in studies to use much the same process (de Groot, 1965; Chase and Simon, 1973).

The step at a time approach is not how people make inferences or retrieve memories. In the party game 'Botticelli' you may be asked whether Van Gogh is a Dutch painter or simply whether he is a painter. Experimental evidence shows that you respond faster to 'Dutch painter' than simply to 'painter' even though the category of painters is far larger. A 'collie' is more 'dog' than 'animal'. It is counter-intuitive, but it is well supported by research. The phenomenon is called the 'category size effect' (Landauer and Meyer, 1972) and may be due to a 'collie' being a closer match to a prototypical dog than it is to a prototypical animal.

The Key Mistaken Assumptions

These three assumptions:

1. That categories are clear and fixed
2. That you need a framework to interpret evidence
3. That you derive conclusions from evidence sequentially

are all mistaken when it comes to human inference and held back AI from being able to do the things that people find easy like recognising objects and understanding speech.

A Personal History

My background was in psychology and, although familiar with logic and mathematics, my instinct was to look towards the way people seem to reason and make inferences, with all its complexity and apparent inefficiency. My inclination was to be descriptive rather than prescriptive. It didn't take much observation to realise that people did not in general infer things by using deductive logic. Indeed, using

deductive logic was quite hard. You had to be taught it. It did not come naturally. So, if people did not use deductive logic to make inferences then how exactly did they do it? Answering this question (or rather proposing hypotheses / models about the answer) back in 1984 is my claim to a place in the history of AI (though few people took much notice at the time despite its relevance today).

I can easily trace the development of my own ideas from the moment, when in a lecture on individual differences during my psychology degree, I first encountered George Kelly's personal construct theory (Kelly 1955). For some reason, this resonated with me so much that my focus on understanding how people construe the world remains central to my thinking 50 years later. Kelly saw people, especially children, as kind of 'scientists' trying to understand the world, testing theories about it and firming up ways of construing the world that made sense to them. If you should doubt this, then question your assumption as to why a baby throws their toys out of the pram. Are they attention seeking or testing hypotheses about cause and effect - for example watching to see how their parents respond, as well as perhaps confirming something about the physical properties of the world?

During my teens and 20s, I had a series of girlfriends. The significance of this is not exactly what you may think. You will have your own ideas about young males, and this just helps to prove my later point about stereotypes. Anyway, I was as interested in the minds of these girlfriends as I was in anything else. In particular, I was interested in how they thought about the other people in their lives.

I would ask them to write a list of somewhere between 10 and 20 people that they knew well. This could include family members, old boyfriends, teachers and all manner of people that they encountered. Then I would use a procedure that I had learnt from a method called Repertory Grid Technique (Fransella and Bannister, 1977). This involved taking 3 of the people at a time at random and asking the girl to say which person was the odd one out and why. It was 'the reason why' that I was most interested in. Doing this several times and noting down the reasons helped build a picture of the distinctions they made between people. These were their 'personal constructs'.

At the end of this exercise, one would end up with perhaps 12 to 20 different distinctions. They might be straightforward ones like, 'these two are male, while the other one is female', or they may be slightly more complicated like 'this one is good looking while the others are not', 'this one has a sense of humour while the others are serious', or they might be more personal like 'I went on holiday with this person but not the others, or 'I like this person while I am indifferent to the others'.

Some of these distinctions put people into categories (male/female) while others could potentially order people along a scale (e.g. from stupid to intelligent or kind to unkind).Nevertheless, at the end of the exercise, one would have a pretty good idea about how this girl would understand and categorise the people in her world. And then when I tried it out on the next girl, some of the categories and scales would be similar while others might be quite idiosyncratic.

One way or the other I was very happy with this technique as a structured way in which one could gain insight into how another person categorised. I suppose I could

have gone through the same exercise with boys but trying to understand the distinction between people was something that, in my own experience, girls seemed to like to do and gave them some insights as well as me.

My friends were doing much the same thing but in a less structured way and using different. subject matter. So, while they might have been having long casual conversations about tastes in music, I was taking my findings to the computer and doing mathematical analysis on them. This is how I got into such things as multidimensional scaling and principal component analysis.

Basically, what all these techniques do is allow you to plot things, whatever they might be – people, football teams, animals, biscuits, trees, musical genres etc. into a multi-dimensional space. I seem to have been born with a passion for systematically understanding how people think – possibly partly because of the many anomalies I observed between what they said on one occasion and another, and what they said and what they did. I appeared to have had the default assumption that people would be logical and consistent and they were not. This was an interesting observation that I was curious to unravel.

Expert 4 – An Inferencing Spreadsheet

By the 1980s, I had read a number of books that had developed my thinking about human inference. Books such as ‘Dimensions of Interpersonal Space’ by Patrick Slater (1977), ‘Judgement under Uncertainty, Heuristics, and Biases’ (Kahneman, Slovic and Tversky 1982) - Daniel Kaheman, Nobel Prize winner and much later of ‘Thinking Fast and Slow’ fame, and ‘Categories and Concepts’ by Smith and Medin (1981) .

In1984, I wrote a computer programme that I called an inferencing spreadsheet and went on to market it under the name ‘Expert 4’ (Rivers, 1984). I wrote this in Basic on an IBM PC over the course of a few weeks while inspired by these ideas. It was a little rough and ready, but it worked and the documentation was thorough and well structured. Now, you would call it ‘proof of concept’. To regard this as an achievement it is worth noting that this provided a computational model of human inductive inference 30 years or so before the breakthroughs in machine learning that led to Large Language Models. Probably a bit cheaper too!

Other spreadsheet programmes existed by that time but were not as ubiquitous as they are today, so it was novel in its use of the spreadsheet format to explore models of human inference. Imagine a spreadsheet that lists all the animals in the 1st column, and then in subsequent columns describes these in terms of their attributes. In fact, each column would be an attribute, and in the cells would be the value on that attribute that an animal had. So, for example, if the attribute was ‘number of legs’, then the possible values might have been 2, 4, 6, and 8.

Say you had a spreadsheet of 50 animals, all described in terms of their attributes and values. You would have a little database. Now, let's say that there are some cells in the spreadsheet where you don't know the values. So long as the spreadsheet is reasonably well populated it is possible to make inferences about

these missing values.

For example, if you didn't know what cows ate, you might consider animals that were most similar to a cow and make an inference from that. The algorithms that I developed to do inferencing in Expert 4 were based on the psychological theories of category and concept formation that I had read about in the psychology literature. And indeed, these seemed to work very well.

I had already understood many of the difficulties with symbolic AI in 1984 before writing Expert 4. They had a significant influence on steering the design of the program away from symbolic AI. I was already convinced from my forays into personal construct theory that collecting cases (examples) in the subject area (domain) that one was interested in, be it people, medical diagnosis, trees, animals etc was a good idea. It was a short step to see that cases within a domain tended to share attributes (i.e. in the case of animals, their relationship to people, what they preferred to eat, common diseases etc.) but that the cases could be distinguished by the values on these attributes (i.e. wild or domestic, in the case of 'relationship to people').

The spreadsheet format seemed obvious – the case in column 1, the attributes across the remaining columns and the values in the cells. This was no more than a database except that the spreadsheet format made everything more visible. For both users and developers, this approach to the user interface opened up what had previously been a black box. The main innovation for me was the mechanism for inferring the values in the cells where the data was missing.

To make the inference of a missing value I drew on what I knew from the psychology books I had read. This led me to try a variety of algorithms like calculating a similarity measure between the case where you were trying to make the inference and the other cases in the spreadsheet. This is not something that spreadsheets typically do, even now. They are still passive calculators as opposed to actively looking for interesting patterns in the data. The calculation of similarity in Expert 4 was key in the sense that it was effectively positioning and calculating the distance between every case in the multi-dimensional space defined by the attributes.

The approach of calculating similarity quickly led to the realisation that it was possible to identify prototypes – that is to say, a case that is typical of its class (Rosch 1973 and 1975, Rosch and Mervis,1975). A typical animal, a typical disease or whatever. More than this, the spreadsheet format makes it easy to identify (i.e. derive) prototypes for cases that have particular attributes. For example, you might find a prototype for a farm animal as opposed to a domestic animal.

This was all a decade before the world wide web came into common use. Early search engines like Alta Vista used similarity algorithms to match search queries to results. This was until Google.com was registered in September 1997 and used algorithms based on page links to rank order results. Recommendation algorithms like those used by Amazon from the early 2000s and those that suggest the next Youtube (registered February 2005) video to watch, the next music track to play and so on, are based on matching mechanisms similar to those developed in Expert 4. In Expert 4 terms they are each 'domains' that have example cases with attributes

and attribute values. The calculation of similarity and prototypes based on (sometimes) limited evidence, allows these algorithms to select out items you are likely to want. Every time you pick one to use, the extra data refines its model. Recommendation algorithms are amongst the most commercially significant technologies in recent years and companies can be quite secretive about how they work to preserve some commercial advantage.

Full documentation for the Expert 4 programme was never published but I still have a comprehensive manual, and I have put together a video that describes it. The programme appeared in the '*Catalogue of Artificial Intelligence Tools*' (Bundy, A. and Wallen, 1984) and was review by the New York Times (Anderson, 1985). The Home Office trialled it on an application to do with inferring the time of death in police investigations and published its findings (Spiehler 1988). There is a link to these papers, the manual and more in Rivers (2026).

Machine Learning

Even 50 years ago when I was messing about with personal construct theory, techniques like multidimensional scaling had been around for some time. Now, here we are more than a quarter of the way through the 21st century, and these techniques are arguably the bedrock of artificial intelligence as we know it today.

Artificial intelligence these days tends to be most closely associate with the technology of neural nets. Neural nets fall within a subclass of artificial intelligence called machine learning. This is something of a misnomer because they only learn in a mathematical sense (the optimisation of a loss function) rather than anything resembling the sense by which people integrate new knowledge into what they already know. As such, they are trained (i.e. taught) but let's leave the distinction between teaching and learning aside for now.

The history of artificial intelligence is often characterised as a battle between symbolic and neural network approaches. But the same distinction can be characterised in other ways. For example, as the distinction between prescriptive and descriptive approaches or as the distinction between deductive and inductive inference. It is also the distinction between rationalism and empiricism or understanding meaning in terms of definitions and understanding it in terms of context. It is worth noting the rise in psychology of an approach called 'situated action' (Suchman 1987) in which behaviours are better understood in the precise context in which they arise rather than through other pre-defined frameworks. Neural nets are also sometimes characterised as doing analogical reasoning or interpolation.

What is actually going on in the training of neural nets for large language models? (Feel free to skip a few paragraphs if you are an AI expert and familiar with these techniques). The training of a neural net is essentially training the weights on connections between nodes in the network to represent the associations between words. Large language models rely on a technique called word embedding. Each word (or to be more accurate word token - i.e. the stem 'play' as opposed to 'plays' or 'played' or 'playing', or a part word or punctuation mark) is described in terms of typically hundreds or thousands of parameters. This description is called a vector.

The vector is no more than a long list of numbers (parameter values) that describe a word in terms of its degree of association with other words. Imagine a dictionary in which the entry for each word is a long list of numbers.

This is like giving each word a score on thousands of different dimensions. In other words, it locates the word in a large multidimensional space. King and queen are very close to each other in this multi-dimensional space. Whilst cabbage and king are somewhat farther apart. Cabbage is probably snuggling up to Brussels sprouts, cauliflower, and the like. So, you see, I feel justified that I was onto something in about 1972 when I first became attached to Kelly's Personal Construct Theory, Repertory Grid technique, Multi-dimensional Scaling and the like. It turned out that these were more viable for capturing aspects of human cognition than the symbolic techniques. They were able to cope with the complexities and subtleties of meaning in a way that symbolic representations with their fixed and defined meanings, simply could not.

There are some important tweaks to this account of how large language models work. The initial training of a model is just the start. After that, the models are refined by human and other input. A technique known as Reinforcement Learning from Human Feedback (RLHF) is often used to sharpen up the basic model. This injection of human intelligence into the process, together with the fact that the vast corpuses of text that the original models are trained on has also captured and consolidated human intelligence, tells us something. It shows that the technology of Large Language Models is doing rather less than you think. It has found a way in which human intelligence can be captured and presented back to users, but a good deal of the intelligence attributed to it is actually human. You do not call a book intelligent just because it captures and presents human knowledge. That's not to say LLMs are not useful. They provide ridiculously easy access to a vast repository of knowledge presented in a way that seems to understand what you want - even more so than most people might. That's the intelligence that the technology contributes. It's in the dialogue.

The most significant component is the mechanism by which these models respond to prompts when they are run. In addition to using the neural nets as trained they contextualise the word 'embeddings' in the light of the prompts and subsequent discussion. This is known as the 'context window'. Each word token adapts its meaning as it passes through the 'attention layers' of the neural net according to the other words in the context window. So, for example, the word 'bank' may change its meaning when the context is to do with money as opposed to fishing by the river. The dynamic reinterpretation based on surrounding context is what developers call the 'attention' mechanism (Vaswani eta., 2017).

So what?

Let's now start unwinding the significance of all this. First, we are now in a position to better understand why symbolic approaches to AI did not work for the type of everyday intelligence that people have. Basically, the proponents of symbolic AI overestimated the power of deductive inference in addressing the real world together with underestimating the difficulties in acquiring knowledge within this paradigm. This is something like using a tool that is admittedly extremely powerful but for the

wrong job. It's like using a forklift truck to move a viscous fluid like honey. It is not the power of the tool that matters; it is its appropriateness to the context of its use. Then it turned out that while natural language is an excellent tool for addressing the subtleties of the real world, theory driven computational linguistics overestimated the role of grammar and underestimated the role of context in capturing meaning. Context and statistical association seem to have worked much better, as demonstrated by the remarkable success of large language models.

A more general insight relates to the way in which the constructs through which we see the world heavily influence our interpretation of it. The categories and constructs we use to see and understand the world is what philosophers refer to as 'ontology'. If you are on the left in politics, you may see the 'benefits system' as a method of equalising the unfair allocation of resources in society, whilst if you are on the right, you may see the same thing as unfairly rewarding people for doing nothing. If the main constructs through which you understand people are to do with whether they are victims or perpetrators, good or evil, left or right or whatever, then that will shape not only your thoughts but potentially also your behaviours. Similarly, if you are a doctor, you may look at people in terms of their health while if you are trading fruit on the market, you may see people in terms of their propensity to buy apples. Your constructs shape what you look for, what you notice and how you interpret it. I argue elsewhere that the ontological layer is a foundational layer on which other forms of intelligence are built.

Your ontology is very much a two-edged sword. It can be exceptionally useful in facilitating relevant inferences, but it can also blind you to alternative explanations. It is too easy to jump to conclusions based on partial evidence. A specialist doctor may diagnose your disease in seconds whereas everybody else will thrash around a huge set of possibilities getting nowhere. But the same doctor can easily miss a disease that another might spot, just because they are not looking for it. You must be prepared to look at the same thing in different ways and instead of always using the same tool, get better at selecting the right tool for the job. When stuck it can be better to put down all the tools and just let the raw data tell you the story as best it can while acknowledging that all data collection is necessarily either deliberately or inadvertently theory laden.

Neural networks are successful because they make fewer assumptions about the interpretation of the data. They let the data speak for itself. The ontology emerges from the data rather than being forced into a pre-determined structure. Their training reflects the patterns in the data rather than imposes a pattern on it. They are not constrained by the structural requirements of symbolic logic, for example.

In the case of LLMs the ontology is already to some extent defined by the word tokens in the training set. But whereas a single book or theory will structure the words into a particular relationship to each other, the learnt connections in a neural network form an abstract model of the relationship between all tokens in the training set regardless of specific usage. They are raw statistical associations untainted by the imposition of any specific meaning. For example, the meaning of the word 'Elizabeth' in a Jane Austen novel specifically refers to a particular character. It's meaning is 'constrained' by its appearance amongst the other words in the novel. In a trained neural network the word 'Elizabeth' is not constrained in that way and will

be associated with whatever words it statistically relates to across the whole training set.

But, and it is an important but, even these raw statistical associations reflect more abstract characteristics about their patterns of usage. For example, they capture the historical patterns of usage related to gender and race. In other words, they are culturally biased. They contain presumptions that can be difficult to identify because we are already so familiar with them. So, whilst a trained neural network may be neutral with respect to specific usages of words they may not be neutral with respect to overall patterns of usage.

There is an interesting parallel with human development. Some of the success of humans (and animals) is because so much learning is passed from one generation to the next during a long period of nurturing, language, education, socialisation and other cultural influences. Each generation must learn for themselves the lesson of the age in which they are born. They start with just the basic level of assumptions necessary to survive then learn the rest on the streets, so to speak. This makes the young vulnerable, but this evolutional adaption is highly effective in shaping each generation to its circumstances. By adulthood we have crystallised those ontological assumptions and find it more difficult to drop them and adapt to change.

We also tend to underestimate the extent to which much of our cognition is motivated by our goals and the future. For example, our memories are not a recording of events but are largely a re-construction to meet the needs of the moment. Certain presidents of the United States have illustrated this very well. Some of us try to reconcile the anomalies and contradictions that this creates while others either lack the mental capacity or do not care.

Large language models have a similar tendency to meet the need of the moment rather than deal with anomalies. Their tendency to hallucinate betrays their lack of a mechanism to buffer and reflect on what they are about to utter. They are optimised for plausible response generation rather than factual accuracy. They have no qualms about reconstructing the world to meet the need to respond to their user. This mimics the tendency for motivated memory and reasoning in people.

When we as people generate, whether it be thoughts, speech or behaviours, our motivations bring forward data from our experience that suits us best in the moment. It can look as if we are consciously and actively biased in retrieving memories and formulating arguments, but this is mostly happening below the level of consciousness. We do not strive for truth or accuracy so much as strive to achieve our objectives. I am talking in general terms, of course. Some of us strive for accuracy more than others, but all in all, we have underestimated the extent to which our default operation is to achieve goals rather than accuracy.

Spreading Activation and Priming

What have we learnt about human inference that both accounts for the relative success of technologies like Large Language Models and illustrates their deficiencies? Let's revisit an example of human inference and tease out in more

detail what is going on. We will quickly discover that inference, expectations, predictions and context are all intimately related.

If I tell you that I have a pet, without other evidence you might assume that it is a dog (i.e. the prototypical pet depending on where you live), but you may also be prepared for cat or maybe hamster. You may imagine a dog with a bone in a basket. You have inferred the basket and the bone based on your knowledge of dogs generally, and all that knowledge can be thought of as collapsed into a prototype – a typical example. It may be an actual example or even an example that does not exist but has all the most likely values on its attributes. Then when I tell you it spends most of the day looking for things to eat in the garden, you will probably update your inferences to rule out 'dog' and consider typical domestic animals that spend most of their time outside. You may think of a dog in a kennel or more likely perhaps you now think of 'cat'.

Can you see where I am going with this? I am demonstrating that rather than collecting lots of evidence and deducing what the animal is, you are making inferences from the very start on mere fragments of evidence. Even if you try to suspend judgement, you cannot help yourself orienting to the animals that it might be.

What is going on is what psychologists and neurologists call 'spreading activation' (Collins and Loftus, 1975). I say the word 'pet' and a wave of activation spreads out to concepts that are closely associated with it. It's like the game where I say a word and you must respond as quickly as you can with the first word that comes into your head. The words to which activation spreads are said to be 'primed'. If I say 'pet', you think 'dog'. Alternatively, if you have just come back from seeing your friend where you fed the 'goldfish', then 'goldfish' will already have been recently activated and that could be the first word you think of when hearing 'pet'. There are many findings from psychology that support this view. For example, primed words are recognised faster than unprimed ones (Meyer and Schvaneveldt,1971).

To put it another way, you are walking around with your head full of primed concepts that have been activated by the recent history of your experience (or thought). It is not as if your head is empty when you hear 'pet'. You are a sentient being, continuously living in and experiencing the world. You are not just responding. You are already 'primed' to expect certain things. You expect 'dog' and 'cat'. You 'half expect' 'guineapig', 'rabbit' and 'goldfish', but you will be surprised by 'polar bear' and 'monkey' (if you live in UK, anyway). Your expectations will be guided by the frequency with which concepts have been associated with each other in the past. These 'patterns' of association stand behind your expectations and your ability to predict.

Neural nets are doing much the same thing. They have been trained in these patterns of association through repeated exposure to examples. They are not learning the specific examples but are assimilating repeated specific experiences into generalised semantic representations. Psychologists would describe this as consolidation of episodic into semantic memory (Tulving, 1972).

It is the context of your past, heavily weighted towards the recent past (but also interestingly, your very early experiences of a concept) that defines how you interpret it and the places where activation spreads to. This aligns with what Daniel Kahneman (2011) in his book '*Thinking, Fast and Slow*' calls 'System 1' thinking. 'System 1' is fast, often unconscious and can be prone to bias and error (so needs correcting by the slower 'System 2' thinking'). Very approximately, 'System 1' is what neural nets do, while 'System 2' is more akin to symbolic approaches to AI. And, incidentally, this is why symbolic approaches to AI may make a comeback as a mechanism for correcting the errors in inferencing made by neural network approaches. There is a and area of neurosymbolic AI (Hitzler et. al, 2022) that has the potential to bridge between machine learning representations of knowledge in neural networks and symbolic representations that have greater integrity when it comes to reasoning. Indeed, Large Language Models can be thought of as a basis for neuro-symbolic approaches (Fang et. al, 2024). This is too big a subject to address here, but I hope to return to it in a later piece.

I wrote Expert 4 with the understanding that in inference you could start anywhere and that wherever you started it could have implications everywhere - an approach driven by the notion of spreading activation. This had the consequence that anything could be a missing value and that, in principle (subject to computational constraints) all the other values in the spreadsheet would contribute to its inference – in particular the degree to which the case matched others in the spreadsheet.

Although Expert 4 was marketed as a product at the time, there was more research that could have been done refining the inference algorithm so that it better modelled experimental results in human inference. The object was not to get an inferencing algorithm that produced 'right' answers but to get algorithms that modelled the way in which human inference works. This was partly driven by the belief that despite its errors, human inference might represent an optimal evolutionary solution that trades-off accuracy with all the constraints of resource management and real-time operation.

In particular, the whole approach illuminates how and why people tend to think in stereotypes. As activation spreads from one concept to others that are associated with it, you are primed to expect certain things, while others that are not associated are not primed. You are prepared. The mechanism of spreading activation confers evolutionary advantage. You hear a twig snap in the African savanna and a whole host of possibilities are primed. Your senses are heightened such that you become hyper-aware of any small clue that might indicate whether a threat is imminent. From just a small fragment of information a whole set of inferences can be made.

Unfortunately, this is a recipe for prejudice and biased thinking generally. Basing expectations on fragmentary information can be highly mis-leading. Hearing a twig snap does not always mean that a predator is lurking. You cannot tell without confirmation. Nevertheless, it is best to err on the side of caution. It could be anything. The less that is known the greater the spread of the priming. Know less, infer more.

Other small fragments of information converge to reduce the ambiguity. The pattern of indicators builds and is automatically matched with prototypical patterns that imply

threats, such as an approaching predator. At some point sufficient evidence will have accumulated to trigger a course of action. This mechanism for updating beliefs is similar to the Bayesian model of inference mentioned above but has psychological rather than mathematical underpinnings.

It is not so much that more evidence provides a more precise pointer into an inference (akin to a symbolic logic approach). It's more that the first piece of evidence opens up a multitude of possibilities that later evidence then modulates. New evidence re-shapes the search space rather than pinpointing a location within it. Instead of indication a specific option, it changes the pattern of likelihood of options. It's a Sherlock Holmes strategy of ruling out the impossible and then whatever is left, however unlikely, must be the solution. System 1 spreads activation far and wide then system 2 tames it. Finding a specific point in the space is easier when multiple routes are opened up rather than relying on a single path. It's a process of 'brainstorm' then 'evaluate. Storm then norm.

This is borne out by everyday experience. When you know very little about somebody there is a tendency to let your imagination run wild inventing all sorts of possibilities. Spreading activation and inference are the same thing. We think that we live in the world and occasionally infer things but, in fact, it's as much the other way round – we live in the world that we infer in our minds. that is then 'held back' or constrained by the real world.

Not only are spreading activation and inferencing the same thing but expectation and prediction can be added to the list. Go to meet somebody at a busy railway station and you may see them several times in the crowds only to discover that, on closer inspection, it's not them at all. A match on hair colour or body shape may be sufficient to give you the initial impression it is the person you are searching for. There is also a neat experiment that demonstrates the same thing. Play an audio recording of a sentence to somebody but take out one word and substitute it with white noise. They will hear that word (or at least some word that fits the context). This is known as 'Auditory Pareidolia' or 'Top-down Auditory Processing'. (Davis and Johnsrude, 2007).

As we go about the world, our senses and our thoughts are engaged in the never-ending process of spreading activation to, and priming various concepts. As we change from place to place, context to context and thought to thought, the concepts that are primed, and those that decay or habituate, are in a constant flux, ebbing and flowing.

The highly activated concepts are not just more receptive. They actively project onto the world. They set expectations, they define what we perceive and attend to. They shape our interpretations. What we experience is at the interface of what we are primed for and what exist in the world. The two are in continuous negotiation to define our reality. For the most part we see what we expect to see but sometimes there is a mis-match – an anomaly - a 'surprise' as Kahneman would put it. That's when system 2 steps in to try and resolve it. This is how we live in the world – partly experiencing it as it is and partly as we expect it to be, and not necessarily as it is.

Both Expert 4 and LLMs live in this world that is partly grounded and partly inferred. Let loose unconstrained the inferences can spread and lose touch with reality. It seems that the context window in large language models is performing a role like the inference mechanism captured within the Expert 4 program. The context window acts as the priming mechanism. It updates the way in which new input is interpreted. It acts to guide the inference of 'missing values', that in the case of an LLM is the next word in the generated response. It then updates the context with the newly generated word.

Interestingly, it also illustrates how easily phenomena like hallucination and model collapse can creep in. It is easy to see in Expert 4 how in a sparsely populated spreadsheet, if inferred values were put back into the spreadsheet and treated as if they were raw data the integrity of subsequent inferences would start to lose touch with reality – what is often now referred to as the 'ground truth'. When this happens at runtime this can lead to hallucinations. When it occurs across successive generations of model training, the same phenomenon is called model collapse. When people are deprived of sensory input, for example, when placed in an 'isolation tank' of water at body temperature and shielded from light and sound, they also very quickly start to hallucinate (Zubek,1969).

Beyond Large Language Models

But do neural nets, and LLMs in particular, really make inferences in the way in which people do? It seems likely that both LLMs and Expert 4 are making 'System 1' type inferences in a way that simulates mechanisms of human inference. Expert 4 can be thought of as a neural net where the weights between nodes are represented in the statistical patterns within the data. However, there is far more to human inference than what either Expert 4 or LLMs capture.

First, neither does what Kahneman described as 'System 2' 'thinking'. According to Kahneman (2011) system 2 is triggered when system 1 produces a result that is a surprise. But how does such a surprise come about? What is the mechanism that triggers a surprise? An LLM will simply carry on generating text even when this does not align with reality – it will hallucinate. The mechanisms used to generate the text do not have a way of registering surprise. This must come from some other mechanism external to the generation process - a mechanism that is monitoring it in some way, comparing its output with some other source of expectation about what its output should be. In people the trigger is a mismatch between expectation and experience and is mediated by an emotional reaction.

When as a human being you see or hear something you do not expect, even from something you yourself 'generate', the first indication is usually some kind of emotion like puzzlement, discomfort, amazement, disbelief or whatever. Work on 'somatic markers' supports this view (Damasio, 1994). These are all types of surprise. The surprise then fires up system 2 mechanisms that seek to find an explanation for the difference or anomaly between what you expected and what you observed. LLMs do not have a mechanism like emotions to act as a trigger. Nor do they have any systematic system 2 mechanisms to deal with it. However, LLMs do have ad hoc mechanisms, for example, 'guardians' or 'safety filters' that detect and prevent malicious usage, and these are arguably what might be counted as system 2. The

problem is that while LLMs currently have just a few of these mechanisms, humans have a vast array of them from strategies to steer system 1 in various directions through to levels of higher-order self-monitoring (i.e. metacognitive resource management and control).

Can LLMs reason and plan? There are many claims that they can and this is sometimes presented as taking a 'step by step' or 'chain of reasoning' approach. This may in some cases be little more than the usual type of context-window-driven generation, but it can be effective and give what might be considered the illusion of reasoning. For the last few days, I have wanted to go out for a walk but have been thwarted by the weather. Every time I have been ready to leave it has started raining. So, today I had to reason about whether I was going to try going out again. I looked at the weather radar and saw that there was probably a window of reasonable weather if I left within the hour. I took a light raincoat just in case and headed off in the car to the park where I wanted to walk. On the way I ran into traffic and decided that I would explore a different route. At every stage I was integrating new information, revising expectations, and adapting my plan — all driven by goals and emotions.

An LLM could not do this. Why? Because I was engaged with the real world while an LLM is only engaged with words. Both words and the real world have statistical patterns but whereas an LLM, given a prompt, launches something akin to a 'ballistic' reply, my reasoning and planning is 'guided' by events in the real world.

You may have noticed that it is a characteristic of all plans that they get thwarted. Almost every step of the way something unexpected happens – it rains, there's a traffic jam, you can't find a place to park… and on it goes. And every time that happens you must reason your way out of it. Will taking an alternative route in a traffic jam be quicker? Is being quicker the only criterion? What if I opportunistically decide that there will be benefit in the discovery of an alternative route even if it takes longer. There is a lot going on here. There is the recognition of a problem, the generation of a variety of alternative solutions, the reassessment of the criteria to evaluate solutions, and various metacognitive controls to determine whether I have the knowledge and energy to explore various decision paths.

While an LLM can sometimes simulate reasoning and planning it is not really doing them. It is not monitoring its own output, constructing scenarios and testing them in its 'mind' before deciding which to go with. It is not coming up against barriers in the real world and working out how to overcome them.

The approach that the AI industry is taking to increase the intelligence of LLMs is to scale up model size and training data. There are some good arguments for this. After all, broadly speaking if you look at the cortex it does seem to be built up of layers with each layer responding to patterns in the layers below. Could building neural networks with more layers be the answer? It could be that the System 2 thinking goes on in the higher layers. The neurological evidence would seem to suggest this. But it cannot be the only answer for several important reasons:

LLMs are not embodied. They do not have a range of sensory modalities like humans (vision, hearing, touch) that cross-validate knowledge of the real world.

There are no integrated models of the real world being built inside the LLM to consult and test. There is no physiology and hence no feeling on which to trigger surprise, and hence no monitoring of complex internal states. Now, robots – that may be another story.

LLMs do not reflect on their own thinking. Yes, there is a feedback loop from generated text back into the context window but there is no independent representation with which to compare and check what they are generating. There is no metacognition.

A human brain, despite being a large lump of neurones and connections between them, is in fact, highly differentiated in its operation. Different parts of the brain specialise in different types of processing. Although there is plasticity (i.e. the ability for one brain area to take over the processes of another, e.g. after brain injury) there is still a high degree of localisation of function. Long range connections between brain structures keep the whole show coordinated. This type of complexity is not replicated in artificial neural network architectures. And this says nothing of the complexity of synaptic connections and the impacts of hormones and other chemical influences on the brain's operation.

LLMs have no persistent memory. Although some implementations use external memory mechanisms such as 'Retrieval Augmented Generation' (RAG) that allow access to previous interactions or external knowledge bases, these remain fundamentally different from the integrated, continuous memory of human experience. They operate only within the current context window (even though these are increasing in size). They may be able to retrieve previous context, but the contexts are not integrated. A person has continuous experience and is engaged in a constant task of integrating new knowledge with old – of consolidating episodic memory into semantic memory.

Why do LLMs have difficulty doing basic arithmetic or counting the number of 'r's there are in 'strawberry'? It partly goes back to the distinction between teaching and learning (and partly to the distinction between machine learning and symbolic AI). If it's not 'taught' somewhere in the training set, the knowledge will not be accessible in the same way as a person can access this knowledge. A child learns and 'caches' the principle so that it can be re-used in a different context. This goes beyond having a persistent memory because knowledge becomes integrated with other knowledge. Context acts as a reminder to consider drawing on that knowledge and metaknowledge helps decide its relevance to the context. Neither of these mechanisms is available to the default generative processes of LLMs. In people, that metaknowledge may be captured in higher layers of the cortex – it is a layer that is capable of 'thinking about thinking' (and thinking about thinking about thinking recursively as necessary).

Current LLMs do not have this layer of reflection. However, some aspects of human metacognitive reflection might be better achieved by combining machine learning and symbolic methods, as in the neurosymbolic approach mentioned earlier. This could represent one of the most promising frontiers in current AI research.

Let's wrap up on this now by making a few more general summary points:

Progress in AI was held back by thinking that intelligence was something to do with abstract logic and symbol manipulation rather than dealing directly with the complex and probabilistic nature of the real world. It is important to use the right tool for the job. With the early focus on symbolic approaches, we mis-judged the trade-off between the certainty of prediction and the ability to handle complexity. We may yet find that being articulate (like an LLM) is also an abstraction that has only a modest bearing on performing physical tasks in the real world. Competency in language does not always translate into competency in task performance as many organisations have found to their cost.

We started to make more progress in AI when we gave up idealised interpretive frameworks (like logic) and assumptions and let the data speak for itself. If there is a mismatch between a theory or interpretive framework and the data, the signal in the data can be obscured. The same principle can apply in science and journalism – go back to the raw data / source and be wary of others filtering and interpretations of it. This always seems especially important to me when looking at politics (and advertising) where often the main purpose is to add spin or inject noise to conceal the signal.

Perception, meaning and interpretation are heavily context dependent. The mechanisms of spreading activation and priming work well to enable people to retrieve relevant memories, reason and act appropriately in different situations.

LLMs do not learn. They are taught in their training and their post-training but outside the context window they have no persistent memory or mechanism of assimilating new knowledge. This will no doubt change with later models but for now there is no systematic and robust way of doing this.

Large language models are doing just a fraction of what intelligence involves. There is still much that is already well understood in human psychology, physiology, and neurology that has not yet found its way into the development of artificial intelligence.

You may note that most of the references in this paper are from the psychology literature from about half a century ago. Also, I have hardly touched on references from the mainstream AI literature. Many of the techniques that underlie large language models have been around for a long time, and we have still not exploited much of what is known about human intelligence. Concepts like prototypes, stereotypes, episodic and semantic memory, spreading activation, priming, executive function and metacognition could easily be more systematically recruited into AI approaches promising more generalised and efficient intelligence. And that's before we get onto the psychology of belief systems and linking thought to action with application to agentic AI. I hope that while reading this you see that AI has a long way to go before it matches the flexibility of human intelligence. AI, advanced and capable as it seems, has only just started.

Finally, I can admit that my son was right that my subconscious mission in doing my video was to lay claim to my rightful heritage in the history of AI. The reality, however, is that although I may have been early off the mark in understanding one,

albeit important, aspect of intelligence (namely, mechanisms of System 1 inferencing), despite my attempt to weave my own history into that of AI, I have no substantive claim to being the grandfather of AI. My son though, whatever he cares to pursue, could potentially claim to be the grandson of something (whatever you might infer that to mean)!

References

Anderson, A. (1985) 'Becoming an expert', *New Scientist*, 14 November, p. 61.

Bundy, A. and Wallen, L. (eds.) (1984), *Catalogue of Artificial Intelligence Tools*. Berlin: Springer-Verlag, p. 53.

Chase, W.G. and Simon, H.A. (1973) 'Perception in chess', *Cognitive Psychology*, 4(1), pp. 55–81.

Collins, A.M. and Loftus, E.F. (1975) 'A spreading-activation theory of semantic processing', *Psychological Review*, 82(6), pp. 407–428.

Damasio, A.R. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Grosset/Putnam.

Davis, M.H. & Johnsrude, I.S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, 229(1–2), 132–147.

de Groot, A.D. (1965) *Thought and Choice in Chess*. The Hague: Mouton Publishers.

Fang, M., Deng, S., Zhang, Y., Shi, Z., Chen, L., Pechenizkiy, M. & Wang, J. (2024). Large language models are neurosymbolic reasoners. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(16), 17985–17993. https://doi.org/10.1609/aaai.v38i16.29754 (Accessed: 20 March 2026).

Fransella, F. and Bannister, D. (1977) *A Manual for Repertory Grid Technique*. London: Academic Press.

Hitzler, P., Eberhart, A., Ebrahimi, M., Sarker, M.K. & Zhou, L. (2022). Neurosymbolic approaches in artificial intelligence. *National Science Review*, 9(6), nwac035. https://doi.org/10.1093/nsr/nwac035 (Accessed: 20 March 2026).

Kahneman, D. (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., Slovic, P. and Tversky, A. (eds.) (1982) *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kelly, G.A. (1955) *The Psychology of Personal Constructs: A Theory of Personality*. New York: Norton.

Landauer, T.K. & Meyer, D.E. (1972). Category size and semantic memory retrieval. *Journal of Verbal Learning and Verbal Behavior*, 11, 539–549.

McNamee, S. (2015) *Radical Presence: Practices for Be-ing in Relationship*. 2nd edn. Chagrin Falls, OH: Taos Institute Publications.

Meyer, D.E. & Schvaneveldt, R.W. (1971) *Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations*. Journal of Experimental Psychology, 90(2), 227–234.

Rivers R. (1984), Expert 4, Full Details, https://robotethics.co.uk/x4/ (Accessed: 20 March 2026).

Rivers R. (2026), Expert 4, Full Details, https://robotethics.co.uk/x4/ (Accessed: 20 March 2026).

Rosch, E. (1973) 'Natural categories', *Cognitive Psychology*, 4(3), pp. 328–350.

Rosch, E. (1975) 'Cognitive representations of semantic categories', *Journal of Experimental Psychology: General*, 104(3), pp. 192–233.

Rosch, E. and Mervis, C.B. (1975) 'Family resemblances: Studies in the internal structure of categories', *Cognitive Psychology*, 7(4), pp. 573–605.

Slater, P. (1977) *The Measurement of Intrapersonal Space by Grid Technique, Vol. 2: Dimensions of Intrapersonal Space*. London: John Wiley & Sons.

Smith, E.E. and Medin, D.L. (1981) *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Spiehler, V., Spiehler, E. and Osselton, M.D. (1988) 'Application of expert systems analysis to interpretation of fatal cases involving amitriptyline', *Journal of Analytical Toxicology*, 12(4), pp. 216–224. Available at: https://doi.org/10.1093/jat/12.4.216 (Accessed: 19 March 2026).

Suchman, L.A. (1987) *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press.

Tulving, E. (1972) ‘Episodic and semantic memory’, in Tulving, E. and Donaldson, W. (eds.) *Organization of Memory*. New York: Academic Press, pp. 381–403.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 6000–6010.

Wittgenstein, L. (1953) *Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford: Blackwell. [Published posthumously.]

Zubek, J.P. (ed.) (1969) *Sensory Deprivation: Fifteen Years of Research*. New York: Appleton-Century-Crofts.