Is Artificial General Intelligence Impossible?

William J. Rapaport

Department of Computer Science and Engineering,
Department of Philosophy, Department of Linguistics,
and Center for Cognitive Science
University at Buffalo, The State University of New York,
Buffalo, NY 14260-2500

rapaport@buffalo.edu
http://www.cse.buffalo.edu/~rapaport/

July 14, 2023

Abstract

Landgrebe and Smith's *Why Machines Will Never Rule the World* argues that it is impossible for artificial general intelligence to succeed, on the grounds that it is impossible to perfectly model or emulate the "complex" "human neurocognitive system". However, they only show that it is practically—not logically—impossible using current mathematical techniques. Nor do they prove that there could not be any other kinds of theories than those in current use. Even if perfect theories were impossible or unlikely, such perfection may not be needed and may even be unhelpful.

The perfect is the enemy of the good.
—Montesquieu (1726) and/or Voltaire (1770)[1]

All models are wrong, but some are useful.
—George Box (1978)[2]

When you have exhausted all possibilities, remember this . . . you haven't.
—Robert Schuller (1983)[3]

1 Introduction

Landgrebe and Smith (2023)[4] argue that it is impossible for artificial general intelligence (AGI) to succeed, on the grounds that it is impossible to perfectly model or emulate the "complex" "human neurocognitive system". However, they do not show that it is *logically* impossible; they only show that it is *practically* impossible using *current* mathematical techniques. Nor do they prove that there could not be any other kinds of theories than those in current use. Even if perfect theories were impossible or unlikely, perfection may not be needed and may even be unhelpful.

2 Their Argument

Roughly, the goal of "narrow" AI is to develop AI systems that can do "intelligent" things. Most AI research, from its beginnings to today, has been at this level. There are successful (or partially successful) systems for natural-language processing,[5] vision, planning and acting, problem solving, game playing, etc.—the sorts of systems discussed in AI textbooks and research articles. Typically, a given narrow AI system does only the one task it was designed for: AI chess programs can't see or solve algebra problems, and vice versa. (For a good survey of narrow AI successes, see Brachman and Levesque 2022, Ch. 3—highly recommended as both an antidote and a complement to L&S's book.)

Artificial *general* intelligence is the attempt to produce a *single* AI system that can do most or all of the narrow tasks in a coordinated fashion, and thus be fully as "intelligent" as a human.

[1] https://en.wikipedia.org/wiki/Perfect_is_the_enemy_of_good

[2] https://en.wikipedia.org/wiki/All_models_are_wrong

[3] https://quoteinvestigator.com/2022/10/21/exhausted/

[4] Hereafter, L&S. All page references will be to this book, unless otherwise indicated.

[5] As distinct from natural-language *understanding*! See the recent literature on the failures (and successes) of ChatGPT, e.g., Metz 2023.

If an AGI *exceeds* human-level intelligence, then it can be said to have passed a point of "Singularity". Some claim that the Singularity bodes ill for humanity; others disagree (Eden et al., 2012).

Narrow AI exists. Is A*G*I possible? If so, is the Singularity possible? L&S argue, as their book's subtitle suggests, that we can have "Artificial Intelligence without Fear" of the Singularity because AGI is, indeed, impossible. Their argument has the following structure:

1. AI is "the application of mathematics to the modeling ... of the functions of the human brain." (p. ix)
2. "The human brain ... is a complex system ...." (p. x)
3. "The only way to engineer ... technology" with human-level intelligence is via "a software emulation of the human" brain. (p. xi)
4. But it is mathematically impossible to predictively model or analyze complex systems. (pp. ix–xi, esp. premise B2, p. xi)
5. ∴ It is impossible to engineer "machines that would possess ... intelligence" greater than or equal to human intelligence. (p. x)
6. ∴ "An AGI is impossible." ("thesis" C, p. xi)
7. ∴ (*a fortiori*) An AGI with powers greater than that of a human is impossible.
8. ∴ (*a fortiori*) "The Singularity is impossible." (conclusion E, p. 3)

That is, "machines will not inherit the earth" (p. 11) and "will never rule the world" (as their title says).

3 Premise 1: The Nature of AI and AGI

3.1 What Is AI?

Characterizations of AI vary widely.[6] As a practicing AI researcher, I take AI to be *computational cognition*: the investigation of whether cognition—human or not—is computable (primarily in the sense of Turing Machine[7] computability) (Rapaport, 1998). Cognition includes natural-language *understanding*, reasoning,

[6] See http://www.cse.buffalo.edu/~rapaport/definitions.of.ai.html for a sampling. For discussion, see Rapaport 2023a, §18.2; Rapaport 2023b.

[7] Henceforth, TM. I capitalize 'Machine' in this context so as not to prejudge whether TMs are really "machines". See Rapaport 2023a, §3.4.1, 'Computation'.

perception, learning, belief, problem solving, etc.—the topics of cognitive science (Rapaport, 1998, 2023b).[8] AI's goal is not to replicate *human* cognition but to find out *how much* of cognition is computable. Thus, the methodology of AI and cognitive science is bottom up: How closely can we approach "intelligence" in our machines?

L&S look from the top down, telling us that there is a limit to how close we can get. They characterize AI as "the application of mathematics to the modelling (primarily) of the functions of the human brain" (p. ix). Curiously, given some of their later claims (see §5.2, below), note that this focuses on modeling the *functions* of the brain, not necessarily the brain itself. Taken literally, their characterization allows that certain brain *functions* (presumably the same as what I call 'cognition') might be implementable in something other than a (human) brain, e.g., produced by such non-biological means as computation.

On the notion of computation involved in AI, L&S say that

> A computer is a machine that deterministically creates a numerical output based on some numerical input using a mathematical model .... [T]he human brain and the human mind-body continuum are not machines of this kind. Indeed, they are not machines of any kind. (p. 89, fn. 12)

They cite Turing (1936) for the first sentence, which is not quite accurate, given that the inputs to a TM are "syntactic entities" ("strings" of "strokes"; Rescorla 2007), not numbers (or even necessarily numerals). But many agree with the second sentence (see, e.g., Piccinini 2015, 2018, 2020). Even if true, however, surely the goal of AGI is not to recreate the human brain, but to create something functionally equivalent to it. And the open research question historically has been whether that can be done via TM computation. Piccinini, for example, has offered a more general notion of computation according to which the brain *does* "compute" (in this more general sense) and—contrary to L&S's third sentence—*is* a machine (or, at least, a mechanism). (This depends, of course, on what's meant by 'machine'; see §7, below.)

3.2 What Is AGI?

L&S's book is an argument against AGI, not *narrow* AI. By AGI, they mean a very perfect emulation of (human) intelligence acting in the real world: "... for an AGI designed to substitute for humans in the performance of complex tasks in natural

[8] I take the goal of *AI* to be the computational study of cognition in general, and I take the goal of *cognitive science* to be the interdisciplinary (including computational) study of *human* cognition in particular (Rapaport, 2000a). "Very roughly, cognitive scientists aim to develop humanlike computer models, whereas pure AI researchers just want to get the job done" (Holyoak, 2023).

environments, inexact predictions are insufficient . . . " (p. 159). But do *humans* in such circumstances always make *exact* predictions? And must AGIs be *better* than humans? So there are two things to consider: First, is theirs a reasonable goal of AGI (is it anyone's goal?). Second, does their argument work against that goal?

My inclination is that, first, it is *not* a reasonable goal, for reasons independent of theirs, namely, perfect emulation may not be necessary in order to achieve (a reasonable, practical version of) AGI. And, second, their argument fails because they claim a mathematical impossibility without a mathematical proof, having left open the possibility that a new kind of modeling might succeed where current methods fail.

The kind of mathematical model that L&S say is needed but that cannot be had seems to be a perfect model:

> Most processes in nature . . . cannot be modelled mathematically. We cannot write down or automatically generate equations which describe, explain, or predict such processes *accurately*. (p. 119, my italics)

Their first sentence needs justification. A more cautious claim, though weaker than they need for their argument, would be that "most processes in nature" have no *known* mathematical model (but even that needs justification).[9] But is "accurate" description, explanation, and prediction needed for an AGI to behave in the world at least as well as we do? AGI practitioners themselves do not characterize their field this way.[10]

In the context of a discussion of self-driving cars, L&S say that "What we need is the sort of reliability that can be provided by mathematical proof. And where complex systems are involved this cannot be attained" (p. 179). But do we really need that "sort of reliability"? Humans are not perfect drivers, even though we are complex systems capable of dealing with other complex systems (see §4, below). Are self-driving cars supposed to be mathematically perfectly safe? Or merely as safe as, or safer than, human drivers? One measures such safety, presumably, by there being fewer or less serious accidents. But, although the best self-driving cars of the future might have fewer or less serious accidents of some of the kinds that *humans* tend to have, they might introduce newer kinds of accidents that humans *don't* have. An important statistic is whether the total number of serious accidents is fewer, not that the total is zero.

[9] Curiously, their example of a "process in nature" not modelable mathematically is "true random number sequences" (p. 119). But a sequence is not a "process". In any case, it is doubtful that there are any "true random number sequences" in nature.

[10] For how they do characterize it, see Pennachin and Goertzel 2007, p. 1; "AGI-08 Call for Participation, The First Conference on Artificial General Intelligence" (http://agi-conf.org/2008/participation.php); and the blurb for the proceedings of that conference (https://www.amazon.com/Artificial-General-Intelligence-2008-Applications/dp/1586038338/).

L&S claim that "mastery of language" is "both a necessary and a sufficient condition for AGI" (p. 217), but "the challenges humans face in understanding language are formidable because of the immense complexity of the signals we receive. How, then do we succeed in the task?" (p. 219). Their answer has three parts: First, we humans

> share linguistic capabilities and a common ground of shared knowledge. Second, language itself serves to constrain the space through which a hearer must search to determine the target intended by the speaker .... And then third, each speaker is able to actively form and interpret utterances based on his [sic] own intentions of the moment. (p. 89)[11]

And why is this something that humans can do and machines can't? Because "We can *describe* and *explain* some of what occurs in the course of such interactions; but we cannot build mathematical models that will enable us to *predict* what will occur" (p. 89). But why is prediction necessary for AGI? Wouldn't it suffice for a (partial) descriptive or explanatory model to enable an AI to converse and, more generally, to act in the real world, albeit imperfectly?[12] And why is perfection needed? Maybe imperfect action suffices.

So their argument seems to be that, because we don't understand how *we* "succeed" in the task of language use, we cannot build an *AI* that would succeed. And—crucially—the reason is that "there is no distribution from which one could sample adequate training material" (p. 233), although this focuses on just one method of model construction: statistical modeling for machine learning.

As to whether there could be some other kind of modeling (e.g., symbolic modeling of common sense, as in Brachman and Levesque 2022), they say this: Complex systems (including language) "do not meet the conditions needed for the application of any *known* type of mathematical model" (p. 235, my italics). *The missing premise here is that there are no other types of mathematical models besides those that we (now) know.* But short of showing (which they don't) that any other type would be *logically* impossible, there is no reason to believe this missing premise.

4 Premise 2: The Human Brain Is a Complex System

A "system" is "a totality of dynamically interrelated elements ... associated with some process—the system's behaviour" (p. 117). "Logic" systems can be modeled

[11] Negotiation is involved here; see Rapaport 2003.

[12] Both Lake et al. 2017 and Chomsky et al. 2023 distinguish between *explanation* via *model*-building and *prediction* via neural-network deep-learning programs, arguing that, while both explanation and prediction are necessary, explanation may be more important.

using mathematics and logic, which allows their behavior to be predictable "almost exactly" (p. 122), because they have "strict rules" (p. 152). Symbolic, GOFAI[13] systems are logic systems in this sense.

"Complex" systems, by contrast, are such that "they cannot be modelled in a way that would yield the sorts of mathematical predictions that can be reliably used in technological applications" (p. 123, citing Thurner et al. 2018, p. 5). So, almost by definition, a complex system is one that cannot be predictively modeled. (Brachman and Levesque 2022, p. 211 make a similar observation.) Interestingly, the very experts that L&S cite on complex systems seem to disagree: "There are those who say that complex systems will never be understood or that, by their very nature, they are incomprehensible. This book will demonstrate that such statements are incorrect. ... [C]omplex systems are algorithmic" (Thurner et al., 2018, pp. vi, 7).

There are actually two complex systems in play here: (1) human beings (the human brain in particular) and (2) the complex system that is the universe to be understood—and acted in—by a human or an AGI. Let's grant, for the sake of the argument, that the universe *is* a complex system. (On L&S's interpretation, this seems to mean that science will never succeed in understanding it perfectly. Yet science understands it pretty well. We'll come back to this in §§6 and 8.) L&S argue that, if the human brain is a complex system, then it can't be mathematically modeled, and so AGI fails.

One feature of complex systems that underlies their unpredictability is that they are "*non-ergodic* (they cannot be modelled by averaging over space and time without losing information) and *non-Markovian* (their behaviour depends not just on one or two immediately preceding steps)" (p. 127). And, L&S say, such systems "are out of reach of *stochastic* modeling" (p. 152, my italics), because "if we measure the behaviour of complex systems by assigning numbers to the observable events which these systems (co-)generate, we obtain data to which no predictive model can be made to fit ..." (p. 159). Hence, they cannot be modeled mathematically.

That last step goes too far. More cautious phrasings would be that stochastic modeling *alone* will not suffice (see §9, below). or that *we don't yet know how* to model them mathematically. Do we have *no* way to deal with them, even if we can't mathematically model them to perfection? As Montesquieu and Voltaire observed, we should not let the perfect stand in the way of the good. (Cf. Brachman and Levesque 2022, p. 46.) Interestingly, in L&S's introduction to the section that begins their arguments, they say that they will "review ... the models *that are available* for the mathematical representation of such systems" (p. 126, my italics),

[13] "Good Old-Fashioned AI" (Haugeland, 1985).

suggesting that there could be others. For their argument to go through, they also need to argue that there are no other possible models. (We'll consider this in §§6 and 9.)

There is a revealing lacuna in L&S's discussion of logic systems: They say that "There are three types of Turing-computable mathematical logic: non-modal and modal propositional logic, and first-order logic ..." (p. 161). Only three? Here are three more: What about modal first-order logic? Or abductive logic? (L&S suggest that there's no work on that (p. 18), but see Hobbs et al. 1993; Josephson and Josephson 1994; Douven 2021.) Or, more to the point—especially given the AI research conducted for some 40 years at Barry Smith's and my common institution—what about relevance logic?[14] This is a very curious example of how L&S have not considered all possibilities in at least this simple case, which suggests that they might not have considered all possibilities in other cases. (Cf. my opening epigraph from Schuller.)

Let's agree for now that the *brain* is a complex system. But on my view of AI, the real question is whether *cognition* is (or must be) a complex system. L&S seem to say that it is, because they seem to identify the human brain with "the human neurocognitive system" (p. xi; cf. p. x). This leads us to premise 3.

5 Premise 3: Software Emulation of the Brain

> The only way to engineer ... technology ["with an intelligence that is at least comparable to that of human beings"] is to create *a software emulation of the human neurocognitive system.* ... To create a software *emulation* of the behaviour of a system we would need to create a mathematical *model* of this system that enables *prediction* of the system's behaviour. (p. xi, my italics)

5.1 Modeling and Emulating

A model can be used to "simulate" or to "emulate" that which is being modeled. As with 'AI', these terms are vague and are used in different ways. We can try to make this a bit less vague as follows: Let the behavior of an agent (computer, human, AI, etc.) be expressed as a function f. (That is not necessary, but makes exposition easier). Then we can say that agent A1 *simulates* the f-behavior of agent A2—A2's computation of f (it need not be TM computation)—if and only if A1 also computes f, with no restrictions on *how* it does so. In particular, A1 need

[14] See, e.g., Shapiro and Wand 1976; Shapiro and Rapaport 1987, 1992, 1995; Martins and Shapiro 1988; Rapaport and Shapiro 1995.

not do it "exactly" as A2 does (whatever 'exactly' might mean). And A1 *emulates* A2's f-behavior if and only if A1 does f "exactly" as A1 does, e.g., using the same algorithm, perhaps even down to the same data structures.

These are still vague,[15] but the important point is that emulation is a very, even maximally, detailed execution of the emulated behavior (including internal states), whereas simulation need not be. There is a spectrum ranging from mere f-computation at one end to algorithmic equivalence (if not identity) at the other.

The open question at this point is: Where on this spectrum does *L&S's* "emulation" lie? Here is how they define these terms: First,

> A *simulation* is a model of a process which imitates the unfolding of the process over time in such a way that, if data about the initial state of the process are entered as input, then data about the terminal state of the process can be inferred. (p. 114)

This certainly seems to agree with the simulation end of my spectrum view, where the minimal notion is (mathematical) functional equivalence with no mention of the intervening process.

Next,

> An *emulation* is the imitation of the behaviour of an entity by means of another entity. ... The aim of an emulation is thus to **mimic** behaviour. (p. 114, original italics, my boldface)

"Mimicry" seems closer to the emulation end of my spectrum, where the maximal notion is algorithmic (procedural) equivalence. That is, in a bare-bones simulation, all that's imitated is the input-output behavior, whereas, in a full-fledged emulation, the intervening behavior is also imitated.

So I think their notions do match mine. However, despite their definitions, they say that "A 'universal Turing machine' ... is a Turing machine that *simulates* a Turing machine on arbitrary input ..." (p. 114, my italics). But even on their definition, they should have said that a Universal Turing Machine (UTM) *emulates* a TM. The instruction set of the TM to be imitated is encoded on the tape of the UTM, and the UTM fetches and executes *those* instructions, not some others that might have the same input-output behavior as the TM. (One TM might *simulate* another by having the same input-output behavior but a different instruction set.)

For AI, the important point is the distinction between doing things exactly as humans do vs. doing them in some other way. The important question with respect

[15] A lot more needs to be said. Do different sorting algorithms simulate or emulate each other? They do "the same thing", after all—sorting—albeit it very different ways. For discussion of this point and how it relates to the ontology of algorithms, see Dean 2016, esp. §2.5.1.

to L&S's argument is whether such detailed *emulation* of (some aspect of) cognition is really necessary for AGI. It may be for some AI researchers, but for many, if not most, others, *simulation* (by TM-computation) is the goal. Is complete or full modeling required? What would that even mean? Would it mean to model a specific human in full detail? Whose brain would it be?

> The mechanist claims that there can be a machine whose outputs are the same as those of a human or a group of humans. What sort of machine? What outputs? *And what sort of human?* (Stewart Shapiro 2016, §8.3, my italics)

There is no "generic brain", so it would have to be someone's in particular. But, as my colleague Stuart C. Shapiro observed,[16] there is a *range* of human-level intelligence (cognition). All that AGI has to accomplish is to get within that range. And, for that, some level of *simulation* is a more reasonable goal, presumably, *computational* simulation.

5.2 Whole Brain Emulation

> I don't care about biology. I care about intelligence.
> —John McCarthy[17]

If the goal of AI is to develop a computational theory of (human) cognition, then emulating the human brain is *not* its goal. It may be a *way* to achieve that goal—and, if L&S are right, an ultimately unsuccesful, if not impossible, way—but it is not the goal, and there may be other ways to reach the goal.

Note the change from "brain" (in premise 2, p. x) to "neurocognitive system" (as quoted in §5's epigraph, above, from p. xi). One possible reason for this switch is to emphasize that only the cognitive aspects of the brain need to be emulated, and not, for instance, the brain's management of the rest of the body (heartbeat, breathing, etc.). However, given the embodiment traditions within some branches of AI and cognitive science, as well as L&S's observations about the nature of complex systems, such *bodily* management functions might well be involved in the brain's *cognitive* functioning. Moreover, if by "software emulation of the human neurocognitive system" L&S are referring to "whole brain emulation" ('WBE', p. 199; see Mandelbaum 2022), then emulation of the *whole* brain (even if to emulate only its *cognitive* functioning) would inevitably bring the rest of the brain's functioning in its wake. Here, I see a difference in approach: L&S (and the WBE

[16] Personal communication, 12 October 2022.

[17] Paraphrase of a statement overheard by Selmer Bringsjord (personal correspondence, 10 March 2023).

community) want to emulate the whole thing in order to get a desired part, in top-down fashion. But another approach starts from trying to emulate (or simulate!) that desired part, and only expand to other functions as needed, from the bottom up, so to speak. In any case, modeling the central nervous system is not the goal of AGI. One feature of premise 3 suggests that somehow the *neural structure* of human cognition is essential. Perhaps it is, but that needs to be argued for. And it's not clear that we know enough at this still early stage of AI research to say that it is essential.

A very nice statement of L&S's basic position, showing the same weakness pointed out in §4, is this: "... it is impossible to model the central nervous system using any *existing* form of mathematics" because of "*our* inability to model complex systems mathematically" (p. 199, my italics). From "*our* inability", one cannot logically infer impossibility (unless "our inability" simply *means* impossibility). But note the hedge about "existing" forms of mathematics, which leaves open the possibility of there being some other, future form that might be used.

There are several other places where they hedge. Here's one:

> ... *we have no idea* how to build a nanomotor that could withstand the high flow-velocity and turbulence of the arterial system. Furthermore, there is no *available* method to generate sufficient amounts of energy for such small motors; indeed, *we have no idea* where their energy could come from. (p. 283, my italics)

On one reading, the italicized phrases are rhetorical euphemisms for "there is no way". But, on a literal reading, all they are saying is that we (or they!) don't *now* know how to do it, not that it can't be done *in the future*.

L&S actually have two "only way" arguments. Their principal one, to be discussed in §6, concerns the alleged impossibility of mathematically analyzing complex systems, which asserts, first, that complex systems cannot be (fully?) modeled by *current* mathematical methods, and, second, that current mathematical methods are the only way to model them.

On the other hand, premise 3's "only way" argument asserts that the only way to achieve AGI or reach the Singularity is to model the *whole* brain. Immediately after stating premise 3, they say that "Alternative strategies designed to bring about an AGI without emulating human intelligence are ... rejected" later in the book (p. xi). This seems to equate "neurocognitive system" with "intelligence". But if "brain" equals "neurocognitive system", and "neurocognitive system" equals "intelligence", then "brain" must equal "intelligence", at least on their view. But surely that's an open question. Isn't one of the goals of AI to consider whether computation is sufficient for intelligence and thus whether (human, biological) brains really are necessary for it?

6 Premise 4: Mathematical Impossibility

It is time to turn to the heart of L&S's argument: the claim that it is mathematically impossible to model complex systems in the way that would be needed for AGI.

In their Foreward, they say that they will "focus specifically on the question of whether *modelling of this sort* has limits, or whether—as proposed by the advocates of ... the 'Singularity'—AI modelling might one day lead to an ... explosion of ever more intelligent machines" (p. ix, my italics). By "modeling of this sort", they mean emulation of the human brain. And they are going to argue that there *are* limits. As I have suggested (and will discuss in §8), such emulation may be more than is needed. If so, then any limits on such modeling might not matter. I am not worried about any such limits, because I have no brief for the Singularity (Rapaport, 2012a). But they also want to argue that even reaching human-level intelligence is beyond the limits of AGI. So the question is where the limits might be.

My characterization of AI in §3.1 allows for there to be limits in the sense of aspects of cognition that are not computable—in the same sense that the Halting Problem is not computable, i.e., a *logical* limitation. However, I am not convinced by any of the several arguments in the literature—including those of L&S—to the effect that various aspects of cognition are not computable, because none of them are truly *logical* limitations (Rapaport, 2023b).

L&S's arguments for such limits waver between logical impossibility arguments and what I'll call "extreme difficulty" arguments. In a phrasing that is ambiguous between these two readings, they say that they will

> show that for mathematical reasons we cannot use ... [the] laws [of physics] to analyse the behaviours of complex systems because the complexity of such systems goes beyond our mathematical modelling abilities. (p. x)

And, as we have seen, they say that "The human brain ... is a complex system of this sort" (p. x). In what sense are complex systems such that they "cannot" be modeled? If that is a logical 'cannot', then complex systems are logically impossible to model. If it is what might be called an "epistemic" 'cannot', then their claim is that we (currently?) do not know how to model them. Surely complex systems can be *partially* modeled. The successes of science are evidence of that. And even if the functions of the complex system that is the human brain cannot be *fully* modeled by current mathematical methods, perhaps they can be modeled well enough to get within (perhaps asymptotically close to) the range of human cognitive abilities.

L&S say that conclusion 6—"An AGI is impossible"—is "analogous to the thesis that it is impossible to create a perpetual motion machine" (p. xi). Contrast

this with what two AGI researchers say:

> Work on AGI has gotten a bit of a bad reputation, *as if creating digital general intelligence were analogous to building a perpetual motion machine.* Yet, while the latter is strongly implied to be impossible by well-established physical laws, AGI appears by all known science to be quite possible. Like nanotechnology, it is "merely an engineering problem", though certainly a very difficult one. (Pennachin and Goertzel, 2007, p. 1, my italics)

The AGI and perpetual-motion-machine impossibilities are different. For P&G, a perpetual motion machine is a physical impossibility (a variety of logical impossibility in the sense of a logical inconsistency with the laws of physics). Although L&S *say* that theirs is that kind of impossibility (see below), it is really more of a "so difficult that it might as well be impossible" kind of impossibility.

We can handle this by dividing premise 4 in two:

Epistemic Cannot: There are no *currently known* mathematical methods for modeling complex systems.

Logical Cannot: There cannot be any other mathematical methods for modeling complex systems.

What is the evidence for the Logical Cannot? L&S say at one point that "Of course, mathematicians will make further discoveries of new types of models in the future" (p. 188), but then almost immediately walk this back, saying that "this would require a major revolution in mathematics of a type which has been ruled out as impossible by leaders in the field" (p. 188). Would not physicists before Newton and Leibniz have argued similarly for understanding physics? To such an objection, they reply that "this is impossible" (p. xii), because

> it would have to involve discoveries even more far-reaching than the invention by Newton and Leibniz of the differential calculus. And it would require that those who have tried in the past to model complex systems mathematically ... were wrong to draw the conclusion that such an advance will never be possible. (p. xii)

But this is just an appeal to an unargued limitation on human intellectual abilities as well as an appeal to authority. First, our human hunter-gatherer ancestors could not have imagined the invention of calculus even though their descendents eventually did (cf. Bringsjord 2022). And, second, just because two of our greatest physicists have "ruled out" "a major revolution in mathematics"—L&S cite the authority of Feynman and Heisenberg—doesn't mean that no one will be able to.

L&S explicitly endorse the Logical Cannot:

> To speak of *mathematical impossibility* ... is to assert that a solution to some mathematically specified problem ... cannot be found ... for *a priori* reasons of mathematics. This is the primary sense in which we use the term .... (p. 9)

And they give as an example the proof of the mathematical impossibility of algorithmically solving the Halting Problem. But they offer no such proof themselves!

They *assert* the Logical Cannot, but they support it by the Epistemic Cannot. A large part of their book is a detailed look at various AI research projects that, they argue, cannot succeed. For example:

> The visual system and the optic tract can ... be seen as a bridge between visual input and its conscious perception, and *both are quite well understood* .... Yet we cannot represent the events occurring during this process in a synoptic mathematical model ... that would allow us ... *to engineer the replacement* of neuronal parts of the optical tract in a way that would rectify a visual impairment. (p. 25, my italics)

If it's well understood, why can't it be mathematically modeled? And if it can't be mathematically modeled, is it really well understood? Or is it possible that something can be well understood sufficiently to reproduce it computationally even if not understood fully, completely mathematically?[18] More importantly, why would we need or want "to engineer the replacement"? That's not the goal of AGI. All that AGI needs is for us to be able to get similar results; that might just require being "well understood" (which is apparently less than being mathematically modeled).

But L&S do harp on the amount of detail that they say would be needed but cannot be had:

> Our mental ... experience and our overt physical behaviour are all emanations from the complex system which is the human mind-body continuum. But due to their nature as processes of a complex system, we are unable to model these processes mathematically, and so we cannot causally explain them at *the fine level of granularity that would be needed to answer questions such as*: which cells and molecules are involved in what way to generate a certain memory ... (p. 30, my italics)

[18] "[Presumably,] something as simple as the motion of water molecules in a glass of water is a complex system. Yet clearly there are models of how the water molecules behave. ... As an engineer, you deal with increasingly high degrees of precision and diminishingly small margins of error. You never have an absolutely perfect model of anything. However[,] that is the physical world. Remember the modeling of the fluid dynamics with a glass of water. You ask what degree of accuracy is required to get the job done and you design a system to meet that requirement." (William S. Jacobs, personal correspondence, 2 March 2023).

Even if it is impossible to *get* that "fine level of granularity", does AGI "*need* to answer" their sort of question? Here's what L&S say:

> To understand this [mind-body] relationship at the level of detail which would enable explanation or prediction ..., we would have to understand the functions of all the cells and of all the cell constituents which contribute to consciousness .... (p.35)

Really? Couldn't we make reasonably accurate predictions for the purposes of AGI without such full knowledge?

Then they say, "Moreover, ... we would have to do this at the level of instances (individual human beings) rather than at the level of general types" (p. 36). Are they saying that any individual AGI that might be constructed would have to have that fine level of detail? That's not as controversial as it might sound; after all, any constructed item of any type will differ in its fine details (even down to the subatomic level) from any other instance of that type. But I think that they are really saying that AGI must construct, not an agent that is generally intelligent within the range of human intelligence (as suggested in §3.2) but one that perfectly emulates a specific human being, and that doesn't seem within the spirit of real AGI as it is practiced.

In a discussion of "force overlay"—the problem of understanding how the four fundamental forces interact—they say that "we have no idea how this happens. ... There is no way to model mathematically what is going on here in any exact way" (p. 131). But *having no idea how it happens* is not the same as saying that *there is no way to mathematically model* (= understand?) it. A reasonable conclusion from their impossibility argument is that there can be no *scientific*—much less computational—understanding or theory of *anything* in the physical world, much less how humans behave, simply because it (and we) are too complex. And yet we understand things well enough. And science continues to progress, sometimes by inventing new methods of understanding. Moreover, our inability to mathematically analyze such systems (e.g., systems such as steam engines and computers—L&S's examples!) doesn't prevent us from constructing them. They point out that turbulence "cannot be modelled in a way that allows the computation of its flows. ... Yet turbulence is one of the very simplest sorts of complex systems" (p. 185). But I do not see how it follows that we can't discover another method that yields similar behavior. As Selmer Bringsjord pointed out to me,[19]

> A cup of coffee through time a[t] breakfast is a "complex, dynamic" system, calling for, under some constraints, analysis to model, not recursion theory—

[19] Personal email, 14 September 2022.

> but that doesn't mean one must use analysis to model this phenomen[on] & build a corresponding artifact.

(More to the point, perhaps, what does turbulence have to do with cognition?)

7 An Impossibility Argument

To prove logically that something is impossible—in particular, not computable—calls for a *reductio* argument, as in the case of the Halting Problem. But L&S's arguments are not of that kind.

Here is one of their arguments for the impossibility of fully mathematically modeling a complex system:

> [1] A machine—as this term is generally understood—is inanimate, does not consist of living cells, and therefore cannot produce energy-carrying biomolecules to survive or reproduce. [2] If it could do so, we would be able to engineer life. [3] But to do so, we would have to be able to model so well that we could re-create through engineering all those functional constituents of organisms that are essential to their survival and reproduction. [4] We would then not be emulating organisms but rather creating something like an organism *Doppelgänger*. [5a] And this we cannot do, because [5b] not only the organism as a whole but also all of these functional parts ... are complex systems, and thus we are thwarted at the very first step of any attempt to create a synoptic and adequate mathematical model. (p. 197, my interpolated numerals)

The structure of this argument seems to be this:

A Machines are not alive. [1]

B If machines were alive, then we could engineer life. [2]

C If we could engineer life, then we would be duplicating, not emulating, it. [3,4]

D Life is a complex system. [5b]

E $\therefore$ We cannot duplicate it. [5a]

Premise A is about as truthful as saying that birds fly. (It's not very truthful: Many birds don't fly.) As Turing (1950) hinted, it appeared to be a self-contradiction to say that a machine can think, because—as 'machine' was generally understood at the time—they were practically defined as non-thinking things (Wittgenstein 1934, p. 47; Mays 1952; Shanker 1987, p. 616; Sieg 2008, pp. 527, fn. 1; 574). But why

is it wrong to think of living things—even thinking things—as machines? Even Searle (1980, p. 422, col. 1,) no friend of AI, said that "our bodies with our brains are … machines".

Premise B seems plausible. After all, to the extent that machines by definition are engineered things, then, if a machine could be alive, it would be an engineered life. Surely, however, L&S don't want to deny the antecedent to conclude that we can't engineer life.

It is with premise C that I have the most problem, because emulations *can* sometimes be duplications; they *can* be the real thing. This can happen, in particular, with information-based emulations, cognition in particular.[20]

Premise D seems reasonable, but I don't see how conclusion E follows. L&S have argued that complex systems cannot be fully modeled. Suppose so. Then all that follows from that and D is that life cannot be fully modeled. Is full modeling required for engineering an emulation?

The closest they come to a logical argument against the possibility of fully modeling a complex system is in this passage:

> … even an approximation of the workings of a complex system is predestined to fail, because the results of applying it will deviate exponentially from reality …" (p. 198)

Here, they seem to be suggesting that complete models of complex systems are in NP. And that, of course, means that, even if such complete modeling is not *logically* impossible, it is still out of reach (unless P = NP, currently considered unlikely). But even if P $\neq$ NP, it might still be possible to model *some* complex systems. After all, we *can* decide of *some* programs whether they will halt or not, and we *can* solve *some* versions of the Traveling Salesman, etc. (Lipton, 2020; Fortnow, 2022).

The important point is that L&S's "impossibility" arguments are not *logical* impossibility arguments in the sense that the proof of the Halting Problem is.

8 The Partiality of Models

> To build computer applications that can function in … [" 'real-world environments' "] is the very point of AGI. The problem is that many in the AGI community do not see a big difference between 'world' and 'model of the world inside the computer'. (p. 54, fn. 44)

[20]For argumentation, see Rapaport 2012b, pp. 54–55; Rapaport 2023a, §14.2.2; Rapaport 2023b, Sidebar C.

Whether AGI researchers see this or not, it is certainly an important point that I fully agree with and that does not deserve to be buried in a footnote. Others have made it before; L&S cite Hesse (1963). I first encountered the idea in Brian Cantwell Smith's "Limits of Correctness in Computers" (1985), which is worth quoting at length:

> To build a model is to conceive of the world in a certain delimited way. . . .
>
> [E]very model deals with its subject matter *at some particular level of abstraction*, paying attention to certain details, throwing away others, grouping together similar aspects into common categories, and so forth. . . .
>
> Models have to ignore things exactly because they view the world at a level of abstraction . . . . And it is good that they do: otherwise they would drown in the infinite richness of the embedding world. . . . If you don't commit that act of violence—don't ignore some of what's going on—you would become so hypersensitive and so overcome with complexity that you would be unable to act.
>
> To capture all this in a word, we will say that models are inherently *partial*. All thinking, and all computation, are similarly partial. Furthermore—and this is the important point—thinking and computation have to be partial: that's how they are able to work. (B.C. Smith 1985, pp. 20–21)

Some AGI researchers not only see this difference; they celebrate it. Here is Pei Wang's (2019, p. 17) definition: "Intelligence is the capacity of an information-processing system to adapt to its environment while operating with *insufficient knowledge and resources*" (my italics). Partiality—Smith's "gap" between the model and the world—is unavoidable and has to be faced not only by an AGI but by us, too. This has been dubbed "efficient intelligence": "*the ability to achieve intelligence using severely limited resources*" (Pennachin and Goertzel, 2007, p. 11).

This is not merely a logical point:

> Even when your eyelids are closed, your visual system is a monumental drain on your reserves. For that reason, *no animal can sense everything well. Nor would any animal want to. It would be overwhelmed by the flood of stimuli, most of which would be irrelevant.* Evolving according to their owner's needs, the senses sort through an infinity of stimuli, filtering out what's irrelevant and capturing signals for food, shelter, threats, allies, or mates. . . . Uexküll [1934] noted that ". . . the poverty of this environment is needful for the certainty of action, and certainty is more important than riches." *Nothing can sense everything, and nothing needs to.* (Yong, 2022, p. 9, my italics)

L&S expect an AGI to need "everything" in order to deal with complexity. But everything might be too much. Partial models might be, not just logically unavoidable, but cognitively necessary.

Even if the universe is a complex system, does an AGI *also* have to be one in order to understand and act in it? The gap between an inevitably partial scientific model and the (equally inevitably) complete world that it models may be the fundamental problem that L&S are pointing to. It could be the reason that they think that complex systems cannot be fully modeled mathematically. They argue that an AGI must, but cannot, be perfect and hence that AGIs are impossible. But *not* being perfect may be what makes them possible.

L&S say that "machines fail to reach the intelligence level of higher animals. The problem is that our world is shaped by complex systems .... Each AI agent ... will therefore have to cope with a complex-system-generated environment" (p. 60). The last sentence is true, but humans have to cope with that, too. But there is also a big difference between the world and a *human's* mental model of it, and we humans seem to be able to deal with the world's complexities.

Perfection is not needed to act in the world. As B.C. Smith notes, we, too, must act on the basis of incomplete models—incomplete knowledge—of the world. And we are able to do it (albeit imperfectly!). So why might an AGI not be able to do it (albeit imperfectly)? The question is whether we can minimize the incompleteness suffiently for them to work well.

L&S argue that language is so complex that it is impossible for there to be a mathematical theory of it (Chs. 4 and 5; see, e.g., the opening paragraph of Ch. 5, p. 74). More precisely, it is "impossible to even begin to collect the gigantic amounts of data that would be needed to train a neural network that could generate responses that are appropriate to any given conversation when taken as a whole" (p. 88).[21] They conclude that AGI (and, presumably, computational linguistics, at least) will fail. But *we* do it, so why wouldn't it be possible for an AI to do it? Do we have some mysterious, vitalistic ability that AGIs cannot have? (See Rapaport 2023b, §4.4.)

L&S give their reason as follows:

> Computers, in order to compute something, require mathematical models, and because there can be no adequate mathematical model of language in general and conversation in particular, attempts must be made using inadequate models, and these lead to failure, such as issuing a routine question in response to an urgent cry for help in an emergency situation, or the refractory failure to understand a pun. (p. 243)

[21] Here, again, the distinction between natural-language *processing* and natural-language *understanding* becomes important. Arguably, ChatGPT "generate[s] responses that are appropriate to any given conversation when taken as a whole", but—more importantly, and as L&S would surely agree—-it does not *understand* what it is "saying" (Seabrook, 2019; Bender et al., 2021; Mitchell, 2021). E.g., it lacks an "internal representation" of what it's talking about (von Hippel, 2023).

But, for L&S, 'adequate' means "perfect", and so the opposite of an "adequate" model is not an "inadequate" one, but an *imperfect* one (a "wrong but useful" one, as suggested by our opening epigraph from Box). A human under the pressure of an emergency might also fail to respond appropriately. And people can fail to understand puns. Again: Must AGIs be *better* than humans? Yes—*if* the real goal is the Singularity, as suggested by the title of the book, rather than as in the title of Chapter 9: "Why There Will Be No Machine Intelligence". (Of course, an AGI that failed to respond appropriately in an emergency or to understand a pun might still "rule the world" or "inherit the earth"!)

9 Other Methods

> In order to produce a machine that thinks better than man, we don't have to understand everything about man. We still don't understand feathers, but we can fly.
> —Edward Fredkin, quoted in Shenker 1977

L&S speak as though the *only* way to achieve AGI is by the currently popular (and successful) methodology of "deep" statistical machine learning. At best, their argument shows that statistical modeling by neural networks by itself won't be able to accomplish AGI. Many working in the field today agree, programs like ChatGPT to the contrary.[22]

Deep-learning, probabilistic, large language models based on neural-network and statistical techniques are good at "simulating" natural-language understanding at a very surfacy level. But they don't really understand. They're missing a *real* ability to remember, to reason, to revise their "beliefs", to make and understand plans, to construct a model of their interlocutor's beliefs, etc. (Rapaport 2000b, §8; Rapaport 2023a, §18.10); they only *look like* they do these things. Even in cases where ChatGPT and its ilk seem to demonstrate "insight", they are only finding things that were in the raw data all along but for which computer power was needed to enable finding them. They don't *understand* that they are exhibiting insight.

But the missing things have not been shown to be uncomputable. In fact, there are very good GOFAI systems that can do some of these things, and there is work on others that are on the right track for things like common sense. What's needed is to combine these into an AGI (Brachman and Levesque 2022, pp. 48f; Marcus 2023b).

[22] See, e.g., Levesque 2017; Garnelo and Shanahan 2019; Landgrebe and Smith 2021; B.C. Smith 2019; Sablé-Meyer et al. 2021; Brachman and Levesque 2022; and the citations in fn. 21, above.

Nancy Cartwright (2022) takes a similar stand when she argues, for reasons not unlike those of L&S, that "physics can't deal with reality's complexity". But, unlike L&S, she offers an alternative:

> Instead of supposing that physics must be queen of all we survey, I recommend we construct our image of what an ultimate science might be like on the basis of what current science is like when it is most successful. ... Physics does not act as queen in these case. Rather, she does her bit as part of a motley assembly of scientific ... and engineering disciplines along with practical knowledge, all working together.

Applied to L&S's argument, her point is that there is no one, single, all-encompassing theory that will describe, explain, and predict everything. Rather, there are many theories, some perhaps not even "scientific" (e.g., "practical knowledge"), that must play a role. (Pennachin and Goertzel 2007, p. 26, make a similar point.)

There is one sticking point: Any model of reality, even one based on such a "motley", must still be partial. But an important question is how big the gap between a partial and a full model has to be. Or, rather, how small it can be before we get a system that can operate in the real world. (And if only a zero-gap will suffice for a Singularity-level AGI, then, I guess, so much the worse for such an AGI. But is a zero-gap necessary?)

Another option leverages the notion of different levels of description (cf. Dennett 1971, 1987). To use Frank Wilczek's example, a hot-air balloonist does not need to apply "the laws of mechanics to [the] atoms" of gas in the balloon:

> the atomic description contains much more information, in principle, but most of that information is worse than useless if you're interestred in flying a balloon (worse, because it adds distractions). (Wilczek, 2021, pp. 213–214)

A more speculative strategy would be to employ a different form of learning. A hint of this can be found in another of L&S's claims, that "we will never be able to engineer machines with the social and ethical capabilities of human beings" (p. 90). We develop these capabilities from growing up and living in a society. So, as Turing (1950) suggested, perhaps we would have to raise an AGI from "childhood" in a society of humans. (For a fictional treatment of this, see Chiang 2019, discussed in Brachman and Levesque 2022, pp. 185–186.) After all, why must an AGI arise fully formed from its programming? To even begin to do this successfully, we would need a new theory and implementation of this kind of learning (Bringsjord et al., 2018; Marcus, 2023a). L&S seem to agree that this is a different kind of learning when they say that "AI systems do not learn in the sense that animals and humans do. ... [T]he machine does not learn anything; it merely computes algorithms taken from the theory of optimisation ..." (p. 167). Of course, it might turn

out that learning *is* the "mere" computation of algorithms. But, more importantly, what *is* "the sense that animals and humans" learn? Why can't a computable theory of such learning be developed? (See Lake et al. 2017; Dehaene 2020 for good overviews of these issues.)

10 Conclusion 8: The Singularity

The Singularity is supposed to be the point at which an AGI becomes more intelligent than a human. But what does that mean? Intelligent along what dimension? Would it have to be more intelligent than the best musician, best politician, best judge, etc.?

L&S's ultimate conclusion—that the Singularity will not be reached—follows trivially from the conclusion that human-level AGI will not be reached. What do L&S think that we *do* have to fear, if not the Singularity or AGI? They say that "The great challenge we are facing is not the replacement of human intelligence. It is not the advent of some Singularity. ... Instead, we face the challenge of finding new occupations for those whose labour will be mechanised" (p. 301). They then go on to say that these challenges will be overcome. Most likely, they will (as the history of earlier mechanizations of labor suggests).

But even far short of true AGI, much less of the Singularity, there is much that we have to be cautious about, if not fear, with AI. Two familiar and still unsolved examples are the Black Box Problem and the Bias Problem: Most, if not all, current AI systems are based on machine-learning techniques that are "black boxes" whose internal workings are opaque and are susceptible to biases inherited from their training sets (Rapaport, 2023a, Ch. 17). And more recently, the "bloviating"[23] of programs like ChatGPT (and, to a lesser extent, DALL-E) that only *appear* to have intelligence is something that we will all have to learn to deal with.

11 Summary

L&S argue that a perfect emulation of the human brain would be necessary for AGI but that such a perfect emulation is mathematically impossible. Hence, AGI and, *a fortiori*, the Singularity are impossible and not to be feared.

[23] "What is ChatGPT doing? It is bloviating, filling the screen with text that is fluent, persuasive, and sometimes accurate—but it isn't reliable at all. ChatGPT is often wrong but never in doubt. It acts like an expert, and sometimes it can provide a convincing impersonation of one. But often it is a kind of b.s. artist, mixing truth, error, and fabrication in a way that can sound convincing unless you have some expertise yourself." (von Hippel, 2023)

I argue that they have not shown the logical impossibility of such a perfect emulation, only that at most it may not be obtainable solely by current mathematical (specifically, statistical) techniques. Hence, further research may enable progress towards the goal of AGI. Further, such perfect emulation is not only not necessary for the AGI project to succeed, but an imperfect or partial model may be both sufficient and even necessary.

L&S's argument is like saying that the only way to get to the moon is with a ladder and that no ladder can be long enough. But there are other ways to get there, such as by rocket ship. Even if a combination of, say, symbolic programming of common sense plus deep machine learning is too short a ladder, we may still need that ladder to get into our rocket ship. Moreover, although the many open issues of computationally (including mathematically) modeling cognition may provide hurdles, they should be treated as research projects for AI, not a set of barriers that cannot be overcome.

L&S fail to show that an "imperfect" AGI would not suffice for behavior in the real world, and they don't offer the necessary logical argument for the impossibility of mathematical analysis of complex systems. All they offer is pessimism.[24]

References

Bender, E. M., T. Gebru, A. McMillan-Major, and S. Shmitchell (2021). On the dangers of stochastic parrots: Can language models be too big? *CFAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://dl.acm.org/doi/10.1145/3442188.3445922.

Brachman, R. J. and H. J. Levesque (2022). *Machines Like Us: Toward AI with Common Sense*. Cambridge, MA: MIT Press.

Bringsjord, S. (2022, February). The argument for God's existence from AI. *European Journal of Science and Theology 18*(1), 77–100. http://kryten.mm.rpi.edu/bringsjord_arggodfromai_0702211715NYv4.pdf.

Bringsjord, S., N. S. Govindarajulu, S. Banerjee, and J. Hummel (2018). Do machine-learning machines learn? In V. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence 2017; PT-AI 2017*, pp. 136–157. Cham, Switzerland: Springer.

Cartwright, N. (2022, 17 October). Physics can't deal with reality's complexity. *IAI [Institute of Art and Ideas] News*. https://iai.tv/articles/nancy-cartwright-physics-cant-deal-with-reality-complexity-auid-2269.

Chiang, T. (2019). The lifecycle of software objects. In *Exhalation*, pp. 62–172. New York: Alfred A. Knopf. https://cpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/8/644/files/2017/08/Chiang-Lifecycle-of-Software-Objects-q3tsuw.pdf.

[24]Thanks to Bill Jacobs, Cliff Landesman, Steve Petersen, and an anonymous referee for very helpful comments on earlier versions of this essay.

Chomsky, N., I. Roberts, and J. Watuymull (2023, 8 March). The false promise of chatgpt. *New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

Davis, M. D. (Ed.) (1965). *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. New York: Raven Press.

Dean, W. (2016). Algorithms and the mathematical foundations of computer science. In L. Horsten and P. Welch (Eds.), *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*, pp. 19–66. Oxford: Oxford University Press.

Dehaene, S. (2020). *How We Learn: Why Brains Learn Better than any Machine . . . for Now*. New York: Penguin Books.

Dennett, D. C. (1971). Intentional systems. *Journal of Philosophy 68*, 87–106. Reprinted in Daniel C. Dennett, *Brainstorms* (Montgomery, VT: Bradford Books, 1978): 3–22.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Douven, I. (2021). Abduction. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 ed.). Metaphysics Research Lab, Stanford University.

Eden, A. H., J. H. Moor, J. H. Søraker, and E. Steinhart (Eds.) (2012). *Singularity Hypotheses: A Scientific and Philosophical Assessment*. Berlin: Springer. http://singularityhypothesis.blogspot.co.uk/.

Fortnow, L. (2022, January). Fifty years of P vs. NP and the possibility of the impossible. *Communications of the ACM 65*(1), 76–85. https://cacm.acm.org/magazines/2022/1/257448-fifty-years-of-p-vs-np-and-the-possibility-of-the-impossible/fulltext.

Garnelo, M. and M. Shanahan (2019). Reconciling deep learning with symbolic artificial intelligence: Representing objects and relations. *Current Opinion in Behavioral Sciences 29*, 17–23. https://www.sciencedirect.com/science/article/pii/S2352154618301943.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.

Hesse, M. (1963). *Models and Analogies in Science*. London: Sheed/Ward.

Hobbs, J. R., M. E. Stickel, D. E. Appelt, and P. Martin (1993, October). Interpretation as abduction. *Artificial Intelligence 63*(1–2), 69–142. Earlier version at https://aclanthology.org/P88-1012.pdf.

Holyoak, K. (2023). Can AI write authentic poetry? *The MIT Press Reader*. https://thereader.mitpress.mit.edu/can-ai-write-authentic-poetry/.

Josephson, J. and S. Josephson (1994). *Abductive Inference: Computation, Philosophy, Technology*. Cambridge, UK: Cambridge University Press.

Lake, B. M., T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences 40*. https://doi.org/10.1017/S0140525X16001837.

Landgrebe, J. and B. Smith (2021). Making AI meaningful again. *Synthese 198*, 2061–2081.

Landgrebe, J. and B. Smith (2023). *Why Machines Will Never Rule the World: Artificial Intelligence without Fear*. New York: Routledge.

Levesque, H. J. (2017). *Common Sense, the Turing Test, and the Quest for Real AI*. Cambridge, MA: MIT Press.

Lipton, R. J. (2020, 19 November). Traveling salesman problem meets complexity theory. *Gödel's Lost Letter and P=NP*. https://rjlipton.wordpress.com/2020/11/19/traveling-salesman-problem-meets-complexity-theory/.

Mandelbaum, E. (2022, August). Everything and more: The prospects of whole brain emulation. *Journal of Philosophy 119*(8), 444–459. https://philpapers.org/archive/MANEAM-4.pdf.

Marcus, G. (2023a, 22 October). 5 myths about learning and innateness. *The Road to AI We Can Trust*. https://garymarcus.substack.com/p/5-myths-about-learning-and-innateness.

Marcus, G. (2023b, 29 October). What "game over" for the latest paradigm in AI might look like. *The Road to AI We Can Trust*. https://garymarcus.substack.com/p/what-game-over-for-the-latest-paradigm.

Martins, J. P. and S. C. Shapiro (1988). A model for belief revision. *Artificial Intelligence 35*(1), 25–79. http://www.cse.buffalo.edu/~shapiro/Papers/marsha88.pdf.

Mays, W. (1952, April). Can machines think? *Philosophy 27*(101), 148–162.

Metz, C. (2023, 20 January). How smart are the robots getting? *New York Times*. https://www.nytimes.com/2023/01/20/technology/chatbots-turing-test.html.

Mitchell, M. (2021, 16 December). What does it mean for AI to understand? *Quanta Magazine*. https://www.quantamagazine.org/what-does-it-mean-for-ai-to-understand-20211216/.

Pennachin, C. and B. Goertzel (2007). Contemporary approaches to Artificial General Intelligence. In B. Goertzel and C. Pennachin (Eds.), *Artificial General Intelligence*, pp. 1–30. Berlin: Springer. https://bilder.buecher.de/zusatz/13/13732/13732357_lese_1.pdf.

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*. Oxford: Oxford University Press.

Piccinini, G. (2018, Spring). Computation and representation in cognitive neuroscience. *Minds and Machines 28*(1), 1–6. https://link.springer.com/content/pdf/10.1007%2Fs11023-018-9461-x.pdf.

Piccinini, G. (2020). *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford: Oxford University Press.

Rapaport, W. J. (1998). How minds can be computational systems. *Journal of Experimental and Theoretical Artificial Intelligence 10*, 403–419. http://www.cse.buffalo.edu/~rapaport/Papers/jetai-sspp98.pdf.

Rapaport, W. J. (2000a). Cognitive science. In A. Ralston, E. D. Reilly, and D. Hemmendinger (Eds.), *Encyclopedia of Computer Science, 4th edition*, pp. 227–233. New York: Grove's Dictionaries. http://www.cse.buffalo.edu/~rapaport/Papers/cogsci.pdf.

Rapaport, W. J. (2000b, October). How to pass a Turing test: Syntactic semantics, natural-language understanding, and first-person cognition. *Journal of Logic, Language, and Information 9*(4), 467–490. http://www.cse.buffalo.edu/~rapaport/Papers/TURING.pdf. Reprinted in James H. Moor, *The Turing Test: The Elusive Standard of Artificial Intelligence* (Dordrecht, The Netherlands: Kluwer Academic, 2003): 161–184.

Rapaport, W. J. (2003). What did you mean by that? Misunderstanding, negotiation, and syntactic semantics. *Minds and Machines 13*(3), 397–427. http://www.cse.buffalo.edu/~rapaport/Papers/negotiation-mandm.pdf.

Rapaport, W. J. (2012a). Can't we just talk? Commentary on Arel's "The threat of a reward-driven adversarial artificial general intelligence". In A. H. Eden, J. H. Moor, J. H. Søraker, and E. Steinhart (Eds.), *Singularity Hypotheses: A Scientific and Philosophical Assessment*, pp. 59–60. Berlin: Springer. https://cse.buffalo.edu/~rapaport/Papers/rapaport2012-Reply2Arel.pdf.

Rapaport, W. J. (2012b, January-June). Semiotic systems, computers, and the mind: How cognition could be computing. *International Journal of Signs and Semiotic Systems 2*(1), 32–71. http://www.cse.buffalo.edu/~rapaport/Papers/Semiotic_Systems,_Computers,_and_the_Mind.pdf. Revised version published as Rapaport 2018.

Rapaport, W. J. (2018). Syntactic semantics and the proper treatment of computationalism. In M. Danesi (Ed.), *Empirical Research on Semiotics and Visual Rhetoric*, pp. 128–176. Hershey, PA: IGI Global. References on pp. 273–307; http://www.cse.buffalo.edu/~rapaport/Papers/SynSemProperTrtmtCompnlism.pdf. Revised version of Rapaport 2012b.

Rapaport, W. J. (2023a). *Philosophy of Computer Science: An Introduction to the Issues and the Literature*. Hoboken, NJ, and Oxford: Wiley-Blackwell.

Rapaport, W. J. (2023b). Yes, AI *Can* match human intelligence. https://cse.buffalo.edu/~rapaport/Papers/aidebate.pdf. Draft of book chapter in progress.

Rapaport, W. J. and S. C. Shapiro (1995). Cognition and fiction. In J. F. Duchan, G. A. Bruder, and L. E. Hewitt (Eds.), *Deixis in Narrative: A Cognitive Science Perspective*, pp. 107–128. Hillsdale, NJ: Lawrence Erlbaum Associates. http://www.cse.buffalo.edu/~rapaport/Papers/rapaport.shapiro.95.cogandfict.pdf.

Rescorla, M. (2007). Church's thesis and the conceptual analysis of computability. *Notre Dame Journal of Formal Logic 48*(2), 253–280. http://www.philosophy.ucsb.edu/people/profiles/faculty/cvs/papers/church2.pdf.

Sablé-Meyer, M., J. Fagot, S. Caparos, T. van Kerkoerle, M. Amalric, and D. Stanislas (2021). Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. *PNAS 118*(16). https://doi.org/10.1073/pnas.2023123118.

Seabrook, J. (2019, 14 October). The next word. *The New Yorker*, 52–63. https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker. See also follow-up letters at https://www.newyorker.com/magazine/2019/11/11/letters-from-the-november-11-2019-issue.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences 3*, 417–457.

Shanker, S. (1987). Wittgenstein versus Turing on the nature of Church's thesis. *Notre Dame Journal of Formal Logic 28*(4), 615–649.

Shapiro, S. (2016). Idealization, mechanism, and knowability. In L. Horsten and P. Welch (Eds.), *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*, pp. 189–208. Oxford: Oxford University Press.

Shapiro, S. C. and W. J. Rapaport (1987). SNePS considered as a fully intensional propositional semantic network. In N. Cercone and G. McCalla (Eds.), *The Knowledge Frontier: Essays in the Representation of Knowledge*, pp. 262–315. New York: Springer-Verlag. https://www.cse.buffalo.edu/~rapaport/676/F01/shapiro.rapaport.87.pdf.

Shapiro, S. C. and W. J. Rapaport (1992). The SNePS family. *Computers and Mathematics with Applications 23*, 243–275. Reprinted in Fritz Lehmann (ed.), *Semantic Networks in Artificial Intelligence* (Oxford: Pergamon Press, 1992): 243–275; http://www.sciencedirect.com/science/article/pii/0898122192901436.

Shapiro, S. C. and W. J. Rapaport (1995). An introduction to a computational reader of narratives. In J. F. Duchan, G. A. Bruder, and L. E. Hewitt (Eds.), *Deixis in Narrative: A Cognitive Science Perspective*, pp. 79–105. Hillsdale, NJ: Lawrence Erlbaum Associates. http://www.cse.buffalo.edu/~rapaport/Papers/shapiro.rapaport.95.pdf.

Shapiro, S. C. and M. Wand (1976, November). The relevance of relevance. Technical Report 46, Indiana University Computer Science Department, Bloomington, IN. https://legacy.cs.indiana.edu/ftp/techreports/TR46.pdf.

Shenker, I. (1977, 27 August). Man and machine match minds at M.I.T. *New York Times*, 8. https://www.nytimes.com/1977/08/27/archives/man-and-machine-match-minds-at-mit-5th-conference-on-artificial.html.

Sieg, W. (2008). On computability. In A. Irvine (Ed.), *Philosophy of Mathematics*, pp. 525–621. Oxford: Elsevier. https://www.cmu.edu/dietrich/philosophy/docs/seig/On%20Computability.pdf.

Smith, B. C. (1985, January). Limits of correctness in computers. *ACM SIGCAS Computers and Society 14–15*(1–4), 18–26. Also published as *Technical Report* CSLI-85-36 (Stanford, CA: Center for the Study of Language & Information); reprinted in Charles Dunlop & Rob Kling (eds.), *Computerization and Controversy* (San Diego: Academic Press, 1991): 632–646; reprinted in Timothy R. Colburn, James H. Fetzer, & Terry L. Rankin (eds.), *Program Verification: Fundamental Issues in Computer Science* (Dordrecht, Holland: Kluwer Academic Publishers, 1993): 275–293.

Smith, B. C. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgment*. Cambridge, MA: MIT Press.

Thurner, S., R. Hanel, and P. Klimek (2018). *Introduction to the Theory of Complex Systems*. Oxford: Oxford University Press.

Turing, A. M. (1936). On computable numbers, with an application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society, Ser. 2, Vol. 42*, 230–265. Reprinted with corrections in Davis 1965, pp. 116–154; https://www.cs.virginia.edu/~robins/Turing_Paper_1936.pdf.

Turing, A. M. (1950, October). Computing machinery and intelligence. *Mind 59*(236), 433–460. https://academic.oup.com/mind/article/LIX/236/433/986238.

Uexküll, J. v. (1934). A stroll through the worlds of animals and men: A picture book of invisible worlds. In C. H. Schiller (Ed.), *Instinctive Behavior: The Development of a Modern Concept*, pp. 5–80. New York: International Universities Press, 1957. https://monoskop.org/images/1/1d/Uexkuell_Jakob_von_A_Stroll_Through_the_Worlds_of_Animals_and_Men_A_Picture_Book_of_Invisible_Worlds.pdf.

von Hippel, P. T. (2023, 4 January). ChatGPT is not ready to teach geometry (yet). *Education Next*. https://www.educationnext.org/chatgpt-is-not-ready-to-teach-geometry-yet/.

Wang, P. (2019). On defining Artificial Intelligence. *Journal of Artificial General Intelligence 10*(2), 1–37. https://content.sciendo.com/view/journals/jagi/10/2/article-p1.xml.

Wilczek, F. (2021). *Fundamentals: Ten Keys to Reality*. New York: Penguin Press.

Wittgenstein, L. (1933–1934). *The Blue and Brown Books*. Oxford: Basil Blackwell, 1964.

Yong, E. (2022). *An Immense World: How Animal Senses Reveal the Hidden Realms Around Us*. New York: Random House.