Do Machine-Learning Machines Learn?

Selmer Bringsjord[1(✉)], Naveen Sundar Govindarajulu[1], Shreya Banerjee[1], and John Hummel[2]

[1] Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA
Selmer.Bringsjord@gmail.com, Naveen.Sundar.G@gmail.com, shreyabbanerjee@gmail.com

[2] 901 West Illinois Street, Urbana, IL 61801, USA
jehummel@illinois.edu

Abstract. We answer the present paper's title in the negative. We begin by introducing and characterizing "real learning" ($\mathcal{RL}$) in the formal sciences, a phenomenon that has been firmly in place in homes and schools since at least Euclid. The defense of our negative answer pivots on an integration of *reductio* and proof by cases, and constitutes a general method for showing that any contemporary form of machine learning (ML) isn't real learning. Along the way, we canvass the many different conceptions of "learning" in not only AI, but psychology and its allied disciplines; none of these conceptions (with one exception arising from the view of cognitive development espoused by Piaget), aligns with real learning. We explain in this context by four steps how to broadly characterize and arrive at a focus on $\mathcal{RL}$.

1 Introduction

Presumably you've read the title, so: No; despite the *Zeitgeist*, according to which today's vaunted 'ML' (= "machine learning") is on the brink of disemploying most members of *H. sapiens sapiens*, no. Were the correct answer 'Yes,' a machine that machine-learns some target t would, in the determinate, non-question-begging, well-founded sense of 'learn' that has been firmly in place for millennia and which we soon define and employ,[1] learn t. But this cannot be the case.

Why? Because, as we show below, an effortless application of indirect proof with proof by cases proves the negative reply. (A formal version of the reasoning is given in the Appendix (= Sect. 8), as a general method that covers any instantiation of 'ML' in contemporary AI.)

[1] The need for the qualifications (i.e. determinate, non-question-begging) should be obvious. The answer to the present paper's title that a machine which machine-learns *by definition* learns, since 'learn' appears in 'machine-learn,' assumes at the outset that what is called 'machine learning' today *is* real learning—but that's precisely what's under question; hence the *petitio*.

V. C. Müller (Ed.): PT-AI 2017, SAPERE 44, pp. 136–157, 2018.
https://doi.org/10.1007/978-3-319-96448-5_14

2 Preliminaries

To validate the negative answer, first, without loss of generality,[2] let's regard that which is to be learned to be a unary function $f : \mathbb{N} \mapsto \mathbb{N}$. The set of all such functions is denoted by $\mathcal{F}$. We say that agent $\mathfrak{a}$ has *really learned* such a function f only if[3]

$\mathfrak{a}$ has *really learned* f

(c1) $\mathfrak{a}$ understands the formal definition D_f of f,
(c2) can[a] produce both $f(x)$ for all $x \in \mathbb{N}$, and
(c3) a proof of the correctness of what is supplied in (c2). (**Note**: (c3) is soon supplanted with (c3′).)

[a] This is the 'can' of computability theory, which assumes unlimited time, space, and energy for computation. See e.g. (Boolos et al. 2003) for explanation.

As we shall see in a moment when considering a grade-school example, real learning so defined ($= \mathcal{RL}$)[4] is intuitive, has been solidly in place for at least 2.5 millennia, and undergirds everyday education every day. Of course, we must concede immediately that the first condition, (c1), employs a notorious word: viz., 'understands.' What is understanding? Not an easy question, that; this we must also concede. Instead of laboring to give an answer, which would inevitably call up the need for a sustained defense of the view that the concept of understanding, as applied to both humans and machines that are supposedly in possession of human-level intelligence and/or consciousness, is not only sufficiently clear, but is also a property that separates real minds from mere machines, we cheerfully

[2] All mathematical models of learning relevant to the present discussion that we are aware of take learning to consist fundamentally in the learning of number-theoretic functions from $\mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N}$ to $\mathbb{N}$. Even when computational learning was firmly and exclusively rooted in classical recursion theory, and dedicated statistical formalisms were nowhere to be found, the target of learning was a function of this kind; see e.g. (Gold 1965; Putnam 1965), a modern, comprehensive version of which is given in (Jain et al. 1999). We have been surprised to hear that some in our audience aren't aware of the basic, uncontroversial fact, readily appreciated by consulting the standard textbooks we cite here and below, that machine learning in its many guises takes the target of learning to be number-theoretic functions. A "shortcut" to grasping *a priori* that all systematic, rigorously described forms of learning in matters and activities computational and mechanistic must be rooted in number-theoretic functions, is to simply note that computer science *itself* consists in the study and embodiment of number-theoretic functions, defined and ordered in hierarchies (e.g. see Davis and Weyuker 1983). We by the way focus herein on unary functions $f : \mathbb{N} \mapsto \mathbb{N}$ only for ease of exposition.

[3] A biconditional isn't needed. We use only a weaker set of necessary conditions, not a set of necessary *and* sufficient conditions.

[4] Not to be be confused with RL, reinforcement learning, in which real learning, as revealed herein, doesn't happen.

supplant the term in question with something unexceptionable.[5] Our substitute for the term is a simple and standard operationalization: instead of relying on the murky and mushy concept of understanding, we simply reply upon *testable behavior that for millennia has served as the basis for ascriptions of understanding to cognizers in the formal sciences.*[6] What behavior are we talking about? Well, the behavior of Euclid and everyone following him who has convinced the objective and the skeptical that they understand such things as mathematical (including specifically number-theoretic) functions, to wit: answers to penetrating questions, and associated proofs that those answers are correct. There literally has been no other way for a human being to provide evidence sufficiently strong to warrant an ascription, to that human being by others, of understanding in the realm of formal functions—or, since the machinery needed for careful articulation of these functions is at least something like axiomatic set theory, in the realm of mathematics itself. Here then, more explicitly, is what we replace (c1) with in order to define $\mathcal{RL}$:

(c1′) $\mathfrak{a}$ can correctly answer test questions regarding the formal definition D_f of f, where the answers in each case are accompanied by correct proofs[7] discovered, expressed, and provided by $\mathfrak{a}$.

We point out that the use of tests to sharpen what AI is, and how to judge the intelligent machines produced by AI, is a longstanding conception of AI itself, provided first by Bringsjord and Schimanski (2003), and later expanded by Bringsjord (2011).[8] It's true that philosophers may crave something more abstract and less pragmatic, but the fact of the matter is that tests are the coin of the realm in real-world AI, and also the coin of the realm in human-level learning in matters formal.[9] For economical exposition in the sequel, we continue to refer to real learning as simply '$\mathcal{RL}$.' We now turn to a simple example that shows $\mathcal{RL}$ to be, as we've said, intuitive, ancient, and operative every single day in the lives of all neurobiologically normal children with the parental or community wherewithal to be schooled:

[5] As many readers will know, Searle's (1980) Chinese Room Argument (CRA) is intended to show that computing machines can't understand anything. It's true that Bringsjord has refined, expanded, and defended CRA (e.g. see Bringsjord 1992, Bringsjord and Noel 2002; Bringsjord 2015), but bringing to bear here this argumentation in support of the present paper's main claim would instantly demand an enormous amount of additional space. And besides, as we now explain, calling upon this argumentation is unnecessary.

[6] Since at bottom, as noted (see note 2), the target of learning should be taken for generality and rigor to be a number-theoretic function, it's natural to consider learning in the realm of the formal sciences.

[7] Just as (computer) programs can be correct or incorrect, so too proofs can be correct or incorrect. For more on this, see e.g. (Arkoudas and Bringsjord 2007).

[8] If we regard Turing to have been speaking of modern AI in his famous (Turing 1950), note then too that his orientation is test-based: he gave here of course the famous 'Turing Test.'.

[9] In fact, this is why real learning for humans in mathematics is challenging; see e.g. (Moore 1994).

Example 1

You, a student, left for high school after breakfast and upon arriving were reminded in math class of the factorial function $n!$. Later in the day, when home, you inform a parent that you have learned the function in question. But you are promptly asked whether you *really* did learn it. So, you are tested by your parent, and by some homework questions that align with (c1′)–(c3):

1. The first problem relates to satisfying (c1)/(c1′): Consider the function g that maps a natural number k to the sum $k + (k + 1)$. Is it true that $\forall n \in \mathbb{N}[n! > g(n)]$? Prove it. Does this proposition hold for every natural number n after a certain size? Answer and prove it.[b]
2. A second problem asks you to ascertain whether the factorial of every natural number greater than 1 is even, and to then prove that the answer is correct.

You certainly can determine the correct answers to problems like these that probe your understanding of the factorial function, and you certainly can supply the definition in various forms and can decide whether proposed definitions are valid, and you certainly (assuming unlimited resources; see note a) can for any input n give back $n!$. Can you also prove that your outputs are correct? Yes, easily. For the fact is that you, reader, can really learn such functions.

[b] Of course an affirmative is correct, and the proof is a trivial use of mathematical induction.

Obviously, an infinite number of such examples can be effortlessly given, in order to anchor $\mathcal{RL}$. For instance, Example 2 could refer instead to the double factorial $n!!$ function, Example 3 to the Ackermann function, and so on *ad infinitum*. Without loss of generality, we rely solely on Example 1.

Now we consider two cases, each predicated on the assumption that the agent $\mathfrak{a}^\star$ we are assessing is a machine-learning one. We specifically assume that, as such, $\mathfrak{a}^\star$ is a standard artificial neural network that machine-learns by repeatedly receiving finite collections of ordered pairs (m, m') of natural numbers, some of which are from the graph of f and annotated as such, and some of which aren't from the graph of f and are annotated as such.[10] Provided that in the limit $\mathfrak{a}^\star$, upon receiving an arbitrary natural number n through time, outputs $f(n)$, save for a finite number of erroneous verdicts, $\mathfrak{a}^\star$ has machine-learned f.[11]

[10] Our assumption here thus specifically invokes *connectionist* ML. But this causes no loss of generality, as we explain by way the "tour" of ML taken in Sect. 6.1, and the fact that the proof in the Appendix, as explained there, is a general *method* that will work form any contemporary form of ML.

[11] This is a rough-and-ready extraction from (Jain et al. 1999), and must be sufficient given the space limitations of the present short paper, at least for now. Of course, there are many forms of ML/machine learning in play in AI of today. In Sect. 6.1 we consider different forms of ML in contemporary AI. In Sect. 6.2 we consider different types of "learning" in psychology and allied disciplines.

3 Case 1

Here we assume that human persons are capable of hypercomputation. Given this, humans can learn some Turing-uncomputable functions in $\mathcal{F}$. (One example might be Rado's (1963) "Busy Beaver" function Σ, which maps the size of a Turing machine measured by the number of its states to the maximum number of contiguous 1s such a TM can produce as output before halting (where the alphabet used is simply $\{0, 1\}$).) Let $h \in \mathcal{F}$ be such a function. That $\mathfrak{a}^\star$ hasn't learned h is a trivial theorem.[12]

4 Case 2

Assume now that $\mathfrak{a}^\star$ is to learn a Turing-computable unary number-theoretic function f, say one that might be seen in math classes; we here refer to Example 1 and its infinite cousins; see above. This case is likewise trivial. The models for machine learning on offer today from AI preclude even reproducing an accurate formal definition of f along with easy proofs therefrom, let alone proofs that proposed values are correct relative to such a definition; that is, conditions (c1$'$)–(c3) aren't satisfied. Since Case 1 and Case 2 are exhaustive: QED.

5 Objections; Replies

A number of objections are perfectly anticipatable. However, voicing and rebutting all of them here is beyond the reach of a reasonably sized paper. Nonetheless, perhaps substantive dialectic is possible. We start by considering a first objection (Objection 1) that we view as a family of interrelated objections.

5.1 Objection 1a: Yours is an idiosyncratic type of learning!

We imagine the objection in question expressed thus: "The definition of 'learning' employed here, i.e. what you dub 'real learning,' results in a very peculiar concept—one that captures neither machine learning nor human learning! And it certainly does not motivate why only this concept is the correct one."

This is flatly wrong. From the mathematical point of view, today's ANN-based machine learning, such as for example has been used in the construction of better-than-any-humans Go-playing systems (i.e. deep learning/DL as the specific type of ML), can be rigorously defined in only two or three ways, for the simple reason that these ways must be based directly on mathematical definitions of machine learning. We are not in the business of taking seriously modern-day

[12] Lathrop (1996) shows, it might be asserted, that uncomputable functions can be machine-learned. But in his scheme, there is only a probabilistic approximation of real learning, and—in clear tension with (c1$'$)–(c3)—no proof in support of the notion that anything has been learned. The absence of such proofs is specifically called out in the formal deduction given in the Appendix.

alchemists, let alone pointing out to them that their use of the term 'learning,' in the context of what learning has for millennia been, is outré. Some internationally famous deep-learning "engineers" have confessed to us in face-to-face conversation that what they are doing in this regard is utterly mysterious to them, mathematically speaking. We in the foregoing cite the ways that exist to understand machine learning logico-mathematically; see our References.[13] We confess that our argument, reflecting our logico-mathematical point of view, quietly but importantly includes a principle that can be summed up as follows:

(*) When investigating whether today's ML (in any of its forms) is real learning (of a number-theoretic function f), the only way to end up with an affirmative to the question is to find a mathematical account $\mathcal{A}$ of today's ML according to which in at least one of its forms its "learning" of f is real learning of f.

In a more formal version of our argument, such as what we give in the Appendix, we provide a step-by-step deductive argument for our main claim that machine-learning machines don't really learn; and this deductive argument renders the principle just given explicit and mechanical.

As to our definition of the real human learning of functions, i.e. $\mathcal{RL}$, this is extracted directly from mathematics textbooks used for many, many centuries. In fact, our triad (c1′)–(c3) can be traced clearly and unswervingly all the way back to Euclid. Real learning isn't peculiar in the least; on the contrary, it's orthodox, and the bedrock of all systematic human knowledge and technology. To validate and explicate $\mathcal{RL}$, we need nothing more than the problems, solutions, and proofs for those solutions that are part and parcel of high-school math—and in fact we only need algebra. Our triadic definition can be empirically confirmed by examining such simple textbooks; see for instance (Bellman et al. 2012). For the case of high-school calculus, see note 18. There is no small amount of irony in the fact that those touting "machine learning" in today's machines as genuine learning have invariably been required to pass the very courses, with the very textbooks, that demand $\mathcal{RL}$.

To wrap up our rebuttal, we note that $\mathcal{RL}$, far from being idiosyncratic, is directly reflective of something that most if not all ML ignores: viz., learning is what produces knowledge. An agent that has genuine knowledge of the differential-and-integral calculus is an agent whose learning has produced at least something very close to justified, true belief with respect to the relevant propositions.[14] That is, the agent believes these true propositions, and has justifications in the form of arguments that establish, or at least render highly likely, the relevant propositions.

[13] A pair of additional works help to further seal our case: (Kearns and Vazirani 1994; Shalev- Shwartz and Ben-David 2014). Study of these texts will reveal that $\mathcal{RL}$ as per (c1′)–(c3) is nowhere to be found.

[14] We of course join epistemological cognoscenti in being aware of Gettier-style cases, but they can be safely left aside here. For the record, Bringsjord claims to have a solution anyway—one that is at least generally in the spirit of Chisholm's (1966) proposed solution, which involves requiring that the justification in justified-true-belief accounts of knowledge be of a certain sort. For Gettier's landmark paper, see (Gettier 1963).

The arguments that undergird knowledge in this way (which were called out above in our example of (c1′)–(c3) in action) are nowhere to be found in contemporary ML, at least in its connectionist, probabilistic, and reinforcement forms.

5.2 Objection 1b: This isn't AI!

In a variant of Objection 1a, we imagine some saying this: "In AI, we are, as a rule, not interested in learning functions over naturals with an infinite domain, given by a graph (or table)."

This is a painfully weak objection, one that reflects, alas, the alchemic nature of much of modern AI. AIniks may not be interested in X, but mathematically they may well be doing X; and if they can't say mathematically what they're doing, then they shouldn't say anything at all in debates such as the present one. Regardless, rest assured that formally speaking, machine learning is learning such functions as we have pointed to (or alternatively learning formal grammars or idealized computing machines). We have given references that confirm this with a ring of iron.

5.3 Objection 1c: What about toads?!

"Your argument has the absurd consequence that even lower animals turn out to be classified as non-learners. Can a toad learn? Certainly. Can a toad learn a number-theoretic function in your sense of learn? Certainly not."

This objection is a kind, unwitting gift, for this is just another way to expose the absurdity of statistical ANN-based machine learning (and of—as we shall momentarily see—other forms of non-logicist machine learning[15]). Agreed: a toad can't learn a number-theoretic function, in the established triadic sense of learning such things we specified above. (We now know that *no* nonhuman animals can do anything of the sort; see e.g. (Penn et al. 2008); ergo our critic can be encouraged to substitute for 'toad' 'dog' or 'chimp,' etc.) But, by the mathematics of statistical ML/DL, a toad (or a toad-level AI produced by the likes of Deep Mind) can learn such a function. This allows us to deduce by *reductio* what the man on the street already well knows: a bunch of smarty-pants people have defined their own private, bizarre, and self-advancing sense of learning. We're now seeing the hidden underbelly of this smug operation, because adversarial tests are showing such things as that DL-based vision systems declare with 99% confidence, for example, that as a turtle is a gun.[16] Of course, we

[15] Specifically, we shall see that the formal deduction of the Appendix is actually a *method* for showing that other forms of "modern" ML, not just those that rely on ANNs, don't enable machines to really learn anything. E.g., the method can take Bayesian learning in, and yield as output that such learning isn't real learning.

[16] Shakespeare himself, or better yet even Ibsen, or better better yet Bellow, couldn't have invented a story dripping with this much irony—a story in which the machine-learning people persecuted the logicians for building "brittle" systems, and then the persecutors promptly proceeded to blithely build comically brittle systems as their trophies (given to themselves).

concede what everyone knows: connectionist ML will continue to improve, and the current brittleness of this form of learning will specifically be addressed in many applications. Yet the mere fact that there currently *is* brittleness is profoundly telling, in the context of $\mathcal{RL}$, for imagine a student Johnny who has real-learned our now-well-worn factorial function. And now imagine that to "test" Johnny, instead of presenting him with a numeral $\langle n \rangle$ where $n \in \mathbb{N}$, and a question as to what the factorial of n is, we instead present him with a picture p of a turtle, and ask him what the factorial of a turtle is. Johnny is likely to inform his parents that some teachers at this school are mentally unstable; certainly there's no chance he's going to blurt out such a response as '24.' The reason for this, speaking imprecisely (recall the earlier discussion at the outset of the paper about the concept of understanding), is that while the DL system has no real understanding of what a turtle or a gun is, Johnny, in satisfying (c1′)–(c3), does.

5.4 Objection 2: Case 1 is otiose!

"Surely your Case 1 is otiose, since—so the objection goes—finite agents, whether human or machine, as everyone concedes, don't in *any* sense learn uncomputable functions."

This is a silly objection, swept away as but dust by the relevant empirical facts; for everyone *doesn't* concede such a thing; witness (Bringsjord et al. 2006), which is in fact based on the aforementioned Σ. As is explained there, Gödel made no concession to the effect that humans don't learn uncomputable functions. For a purported proof that human persons hypercompute, see (Bringsjord and Arkoudas 2004);[17] for a book-length treatment, see (Bringsjord and Zenzen 2003).

5.5 Objection 3: Your definition of human learning is tendentious!

"Your triadic definition of learning [based on your conditions (c1′)–(c3)] conveniently stacks the deck against modern statistical machine learning (=ML in the current discussion and in the—by-your-lights fawning—media). This definition is highly unnatural, and highly demanding."

We note first in reply that convenience *per se* is of course unobjectionable. Next, telling in this dialectic is the brute fact that for well over two millennia we have known what it is for an agent to have really learned some math or formal logic, number-theoretic functions included; and what we in this regard know aligns *precisely* with the triadic account of $\mathcal{RL}$ given above. Again, empirical confirmation of this alignment can be obtained by turning to what the textbooks

[17] In which is by the way cited hypercomputational artificial neural networks.

demand in terms of proof,[18] and what the disciplines in question demand of those who wish to lay claim to having truly learned some formal logic or math.

In short, we cannot allow the field of AI, and specifically its ML subpart, now on the intellectual scene for not more than a blip of time, to trample ordinary language and ordinary meaning that has been firmly in place within the formal sciences for millennia. We are not here appealing directly to so-called "ordinary language" philosophy, and philosophers in this school (such as G.E. Moore, Austin, Norman Malcolm, and various modern defenders). As a matter of fact, the veridicality of ordinary language is something we in general find attractive, but we need only a circumspect general principle like this one:

(+) If natural-language communication has for millennia taken the *bona fide* learning of an arithmetic function f by an intelligent agent $\mathfrak{a}$ to happen only if Φ, then, absent a separate and strong argument in favor of an incompatible set Ψ of conditions that contravenes this, one is justified in applying Φ to claims that $\mathfrak{a}$ can learn/has learned some given function f'.

Perhaps the remarkable thing about (+) is that the behavior of ML practitioners *themselves* confirms its truth. The field of machine learning has both foundational theorems such as the *No Free Lunch* theorem (Wolpert 1996) and new working theorems that are constantly introduced in the scientific literature of the field, e.g. Theorem 2.1 in (Achab et al. 2017). Leaving aside theorems and other formal knowledge produced by ML practitioners, consider the case of ANN-based ML, for instance today's DL. DL experts examine some given data, and through domain expertise built up in the past (via a process much mediated by natural communication, written and oral), devise a target set of functions (denoting the architecture of the neural network in question):

$$\{f_{\mathbf{w}} \mid \mathbf{w} \text{ is in some large space}\}$$

The machine then simply tunes the weights $\mathbf{w}$. Specifically, in convolutional (artificial) neural networks, the form of the function best suited for image processing was conceptualized by humans and justified, by not just performance measures, but by an *argument* in good old-fashioned English for the conclusion that this form of neural networks might be good for image processing. See (LeCun et al. 1998) and Chap. 9 in (Goodfellow et al. 2016) for examples of this process. Note that even if a machine selects the architecture, that selection is happening from a

[18] E.g. even beginning textbooks introducing single-variable differential/integral calculus ask for verification of human learning by asking for proofs. The cornerstone and early-on-introduced concept of a *limit* is accordingly accompanied by requests to students that they supply proofs in order to confirm that they understand this concept. Thus we e.g. have on p. 67 of (Stewart 2016) a request that our reader prove that

$$\lim_{x \to 3} g(x) = (4x - 5) = 7$$

. What machine-learning machine that has learned the function g here can do *that*?

class of architectures designed by none other than the guiding humans, and there is no justification from the machine beyond performance measures. A relatively different form of machine learning, *inductive programming* (Kitzelmann 2009), seeks to learn functions like addition by looking at a very small set of sample inputs and outputs. But even this is shallow when stacked against real learning: In $\mathcal{RL}$, humans can not only look at inputs and outputs, but also descriptions of the properties of the function written in English (and other natural languages, as the case may be) that go well beyond the examples.

5.6 Objection 4: Do flying machines (really) fly?

"Unfortunately, you are dancing around an unanswerable quagmire that has been with us for rather a long time, one summed up by the seemingly innocent question: Do flying machines (really) fly?"

Suppose there is an embodied AI $\mathfrak{a}'$ with all sorts of relevant sensors and effectors in the form of an autonomous drone that can take off by itself and travel great distances adroitly, land, and so on—all without any human intervention in it or its supporting systems during some flight from time t to t'. Did $\mathfrak{a}'$ really fly during this interval? Of course it did. Do eagles really fly over intervals of time? Of course they do. There is no objection to our argument to be found in the vicinity of these (nonetheless interesting) questions. In the case of $\mathcal{RL}$, there are no machines on the planet, and indeed no machines in the remotely foreseeable future of our solar system, that have the attributes constitutive of this learning.

5.7 Objection 5: You concede that your case is limited to the formal sciences!

"You have conceded, perhaps even stipulated, that real learning in your argument is restricted to the realm of the formal sciences. Hence, if your case is victorious, its reach is rather limited, no?"

Quite the contrary, actually. We have indeed restricted real learning to the formal sciences. However, we had assumed that it would be clear to all readers that adaptation and expansion of (c1$'$)–(c3) to *non*-formal domains would if anything bolster our case, if not immediately render it transparently victorious. Apparently our critic in the case of the current objection needs to be enlightened. Consider creative writing. What does it take to learn the "functions" at the heart of creative writing, so that eventually one can take as input the premise for a story and yield as output a good story?[19] We can safely say that any agent capable of doing this must be able to read not formal-scientist Euclid, but, say, Aristophanes, and a line of creative writers who have been excelling since the ancient Greeks; and learn from such exemplars how such a "function" can be computed. But reading and understanding literary prose, and learning thereby,

[19] This is essentially the Short Short Story Game of (Bringsjord 1998), much harder than such Turing-computable games as Checkers, Chess, and Go, which are all at the same easy level of difficulty (EXPTIME).

is patently outside the purview of current and foreseeable AI. And it gets worse for anyone who thinks that today's machine-learning machines learn in such domains: In order to learn to be a creative writer one must *generate* stories, over and over, and learn from the reaction and analysis thereof, and then generate again, and iterate the process. Such learning, which is real learning in creative writing, isn't only not happening in ML today; it's also hard to *imagine* it happening in even ML of tomorrow.

6 Real Learning in Context

The dialectic in the previous section makes it abundantly clear that 'learning' is polysemous: it means many different things to many different people. Given this fact, we think it's worthwhile to a bit more systematically place real learning within the context of different senses of learning in play in contemporary AI and cognitive science/psychology. We thus briefly review the prominent senses of learning in AI (Sect. 6.1), and then in cognitive science/psychology (Sect. 6.2); and then, this two-part review complete, we proceed (Sect. 6.3) to quickly explain in broad strokes how by a series of four steps real learning can be isolated within the broader context afforded by the review.

6.1 Learning in AI

Everyone must admit that there are many different extant ways to map the geography of what is called "learning" in the field of AI. This is easily confirmed by the existence of modern, credible overviews of learning in AI, in textbooks (each of which, of course, has been fully professionally vetted): the geographies offered in each pair of these books is different between the two. Given this divergence, we can't possibly give here a single, definitive, received breakdown of learning in its various forms within contemporary AI. On the other hand, it's nonetheless clear that *any* orthodox breakdown of the types of learning in the field, in any textbook, will immediately reveal that no type matches real learning = $\mathcal{RL}$.[20] We here follow Luger (2008), whom we find particularly perspicuous, and quickly point out, as we move through his geography, that real learning is nowhere to be found. Nonetheless, it will be seen that Luger (2008), to his credit, opens a

[20] Outside of the present paper, we have carried out a second analysis that confirms this, by examining learning in AI as characterized in (Russell and Norvig 2009), and invite skeptical readers to carry out their own analysis for this textbook, and indeed for any comprehensive, mainstream textbook. The upshot will be the stark fact that $\mathcal{RL}$, firmly in place since Euclid as what learning in the formal sciences is, will be utterly absent.

door to a path that could conceivably lead to real learning, at some point in AI's future.[21]

Luger (2008) devotes Part IV of his book to "Machine Learning;" four chapters, 10–13, compose this part, and each is devoted to a different form of machine learning:

- Chap. 10: "Symbol-Based"
- Chap. 11: "Connectionist"
- Chap. 12: "Genetic and Emergent"
- Chap. 13: "Probabilistic"

As one would expect, connectionist learning covers machine learning that is rooted in ANNs. For reasons given in the present paper, there isn't a scintilla of overlap between what is covered in Chap. 11 and $\mathcal{RL}$. This is true for starters because the familiar, immemorial declarative information that has defined such things as the factorial function are nowhere to be found within an any artificial neural network whatsoever. The same applies, *mutatis mutandis*, to the genetic-and-emergent type of learning covered in Luger's (2008) Chap. 12, as should be obvious to all readers. (Genetic algorithms, for example, make no use of the sort of declarative content that defines number-theoretic functions.) We are thus left to consider whether $\mathcal{RL}$ appears in symbol-based learning presented in Chap. 10, or in probabilistic learning covered in Chap. 13. In point of fact, which energetic readers can confirm when reading for themselves, real learning doesn't appear in either of these places.

Now, we said above that Luger (2008) opens the door to a future in which AI includes real learning. We end the present section by explaining what we mean.

In the final part of (Luger 2008), V, entitled "Advanced Topics for AI Problem Solving," two topics are covered, each of which is given its own chapter: "Automated Reasoning," covered in Chap. 14; and "Understanding Natural Language," presented in Chap. 15. Luger's (2008) core idea is that for truly powerful forms of problem-solving in a future AI, that remarkable machine will need at least two key things: it will need to be able to reason automatically and autonomously in deep ways, starting with deductive reasoning; and second, this AI will need to be able to really and truly understand natural language,

[21] Instead of looking to published attempts to systematically present AI (such as the textbooks upon which we rely herein), one could survey practitioners in AI, and see if their views harmonize with the publications explicitly designed to present all of AI (from a high-altitude perspective). E.g., one could turn to such reports as (Müller and Bostrom 2016), in which the authors report on a specific question, given at a conference that celebrated AI's "turning 50" (*AI@50*), which asked for an opinion as to the earliest date (computing) machines would be able to simulate human-level learning. It's rather interesting that 41% of respondents said this would *never* happen. It would be interesting to know if, in the context of the attention ML receives these days, the number of these pessimists would be markedly smaller. If so, that may well be because, intuitively, plenty of people harbor suspicions that ML in point of fact hasn't achieved any human-level real learning.

including complete sentences, following one upon another.[22] It will not escape the alert reader's notice that the capability constituted by this pair is at the heart of what it takes to be a "real learner," that is, to be an agent that really learns as per (c1′)–(c3) in the formal sciences. Unless an AI can *itself* prove such things as—to repeat a part of the Example we began with — that the factorial function's range consists of the even natural numbers, and receives and understands/processes challenges to prove such things, where these challenges come in the form of arbitrary, full sentences like "Show that the factorial of every number is even," it won't be an AI that really learns. Unfortunately, while Luger (2008) points the way toward aspects of two key capabilities needed for real learning, he does only that, by his own admission: point. So, while the door is open, our claim, that machine-learning machines of *today* don't really learn, is unscathed.

6.2 Learning in Psychology and Allied Disciplines

Recall the factorial-function example we gave at the outset. When, upon returning home after school, you are asked by a parent, "So, what did you learn today in math?" it's rather doubtful that if you answered earnestly and sincerely, and if your time in class was a pedagogical success, you replied in accordance with anything violently outside the bounds of $\mathcal{RL}$. Nonetheless, psychology and its allied disciplines (= psychology$^+$) have (perhaps inadvertently) erected an ontology of forms of learning that at least in principle offer viable alternatives to $\mathcal{RL}$, or even perhaps forms of learning that match, overlap, or conceivably subsume $\mathcal{RL}$. Put intuitively, the question before us in the present section is this one: Could you reasonably have conversed with your parent on the basis of any of the types of learning in psychology$^+$'s ontology thereof? As we now reveal, the answer is No.[23] We begin with the authoritative (Domjan 2015), which is based on this operationally inclined definition:

---

[22] Luger's book revolves around a fundamental distinction between what he calls *weak problem-solving* versus *strong problem-solving*.

[23] There are a few exceptions. Hummel (2010) has explained that sophisticated and powerful forms of symbolic learning, ones aligned with second-order logic, are superior to associative forms of learning. Additionally, there's one clear historical exception, but it's now merely a sliver in psychology (specifically, in psychology *of reasoning*), and hence presently has insufficient adherents to merit inclusion in the ontology we now proceed to canvass. We refer here to the type of learning over the years of human development and formal education posited by Piaget; e.g. see (Inhelder and Piaget 1958). Piaget's view, in a barbaric nutshell, is that, given solid academic education, nutrition, and parenting, humans develop the capacity to reason with and even eventually *over* first-order and modal logic—which means that such humans would develop the capacity to learn in $\mathcal{RL}$ fashion, in school. Since attacks on Piaget's view, starting originally with those of Wason and Johnson-Laird (e.g. see Wason and Johnson-Laird 1972), many psychologists have rejected Piaget's position. For what it's worth, Bringsjord has defended Piaget; see e.g. (Bringsjord et al. 1998).

> Learning is an enduring change in the mechanisms of behavior involving specific stimuli and/or responses that results from prior experience with those or similar stimuli or responses. (Domjan 2015, p. 14)

That learning is here attributed to a change in the 'mechanisms of behavior' would seem to draw a hard line between learning and performance. Performance can after all be the effect of multiple factors besides learning, and hence is not a sole determinant of the latter. At any rate, in our study of types of learning in psychology$^+$, we found the following six forms of learning. As we progress through the enumeration of these forms, we offer in turn a rather harshly economical summary of each, and render a verdict as to why each is separate from and irrelevant to real learning (with the possible exception, as we note, of the last). Here goes:

1. *Associative Learning: Classical and Instrumental Conditioning.* The theory of classical conditioning originates from the (Pavlovian) finding that if two stimuli, one unconditional (US), such as food, and the other neutral (CS), come in close temporal contiguity, and if US elicited some response naturally (say salivation), then CS too eventually elicits that response. While here the change in behavior is attributed to some contingency between CS and US (also called *reinforcer*), in instrumental conditioning this change results from some contingency between that behavior and the reinforcer (Mackintosh 1983). Obviously, if this strengthening or reinforcement of the new pattern in behavior is no more than a new stimulus-response connection, real learning is nowhere to be found.[24]
2. *Representational Learning.* The representational theory of learning (Gallistel 2008) views the brain as a functional model capable of computing a representation of the experienced world; and that representation in turn informs the agent's behavior. While learning here is taken to be a process of acquiring knowledge from experience, 'knowledge' here means nothing like the knowledge that is front and center in Example 1 of $\mathcal{RL}$.
3. *Observational Learning.* Here, a new behavior is learned simply by observing someone else. Mostly associated with the social learning theory of psychologist Albert Bandura (1977), his Bobo-doll experiment (Bandura et al. 1961) is an interesting study of how children learn social behavior such as aggression through the process of observational learning. This type of learning in psychology$^+$ is learning by straight imitation, and as such is obviously not $\mathcal{RL}$. Put simply and baldly, the decision problems we presented in our starting example (e.g., is $n!$ invariably even?), and the confirmatory proofs for each answer, are not supplied by shallow imitation of the likes of inflatable Bobo dolls.
4. *Statistical Learning.* Extraction of recurring patterns in the sensory input generated from the environment over time is the core essence of this type

[24] We are happy to concede that years of laborious (and tedious?) study of conditioning using appetitive and aversive reinforcement (and such phenomena as inhibitory conditioning, conditioned suppression, higher-order conditioning, conditioned reinforcement, and blocking) has revealed that conditioning can't be literally reduced to new reflexes, but there is no denying that in conditioning, any new knowledge and representation that takes form falls light years short of $\mathcal{RL}$.

of learning (Schapiro and Turk-Browne 2015). Taking a cue from associative learning in nonhuman primates, past studies showed a possibility of sensitivity of certain parts of the brain when exposed to temporally structured information. Detection of conditional probability patterns in sound streams as a precursor to language parsing, leading to predictions of some sounds given other sounds, would be a good example. Schapiro and Turk-Browne (2015) give a nice overview of various studies related to auditory and visual statistical learning in humans, including neural investigations towards the role of different regions of brain in diverse forms of such learning. Though statistical learning is suggested as a pervasive element of cognition, it is yet early to state this as a form of real learning.

Marblestone et al. (2016) draw a parallel between human brain functioning and the activity of ANNs in connectionist ML. They specifically claim that the neural structure of the brain coincides with various methods of weight assignments to multiple hidden layers of ANNs when machine learning takes place. We gladly concede for the sake of argument that this direction holds promise for the neurological "decoding" of the human brain, since the core idea is that there's a match between brain activity and ANNs through time in ML. But since this activity cannot in any way be interpreted to constitute embodiments of the three clauses that define $\mathcal{RL}$, we once again see here an entirely irrelevant form of learning.

5. *Neurocentric Learning*. Titley et al. (2017) propose a non-exclusive, neurocentric type of learning. For ease of exposition, let's label this type of learning simply 'L_{ne}.' L_{ne} marks a move away from a synaptocentric neurobiological form of learning: in L_{ne}, both synaptic plasticity and intrinsic plasticity play a role in learning and memory. More specifically, synaptic plasticity assigns connectivity maps, while intrinsic plasticity drives engram integration. While L_{ne} is certainly interesting, and while it may well hold much promise, it's undeniable that learning in this sense is clearly not relevant to our conception of $\mathcal{RL}$. Confirmation of this comes from the brute fact that no account based on the building-blocks of L_{ne} can be used to express even the tiniest part of $\mathcal{RL}$. Colloquially put, no agent who learns, say, the Ackermann function in a given recursion-theory class, and is proud that she has, can report this happy event by expressing her enlightenment in terms of the proofs demanded by the clauses that define $\mathcal{RL}$.
6. *Instructional Learning*. Instructional learning is in play when an individual learns from instruction (for example, a teacher's verbal commands in a classroom) and responds with corresponding action/s. While we of course agree that instruction acts as a purposeful direction of the learning process (Huitt 2003), this learning fails to qualify as $\mathcal{RL}$ because action alone doesn't define learning. Of course, in theory, the actions of student learners could be fleshed out to correspond to $\mathcal{RL}$'s three clauses. Were this carried out, it would merely show that instructional learning, at least of a particular type (e.g., instructional learning in the formal sciences), corresponds to $\mathcal{RL}$—but this we've known from, and has indeed been plain to readers since, the outset of the present paper.

6.3 The Four-Step Road to Real Learning

Having completed our rapid tour of ML in contemporary AI, and learning in psychology$^+$, we now provide a general characterization of what real learning is, within this context. Saying what real learning is in the broader context constituted by the previous two subsections can be achieved by first by throwing aside irrelevant, lesser forms of cognition; this will be the first of four general steps taken to arrive at $\mathcal{RL}$:

Step 1: We begin by observing that the cognitive powers of creatures on Earth are discontinuous, because human persons have reasoning and communication powers of a wholly different nature than those possessed by nonhuman animals. A non-technical version of this observation is provided by Penn et al. (2008). A more specific, technical analysis, undertaken from a logico-mathematical standpoint, allows us to simply observe that only members of *H. sapiens sapiens* are capable of such things as[25]

- understanding and employing indubitable abstract inference schemas that are independent of physical stuff (e.g. *modus tollens*; see Ross 1992);
- understanding and employing arbitrary, layered quantification (such as that 'Everyone likes anyone who likes someone' along with 'Alvin likes Bobby' allows us to prove that 'Everyone likes Bobby');
- recursion (e.g. as routinely introduced in coverage of the recursive functions in an intermediate formal-logic course, which might wisely use (Boolos et al. 2003));
- infinite structures and infinitary reasoning (a modern example being the proof that the Goodstein sequence goes to zero; see (Goodstein 1944));
- etc.

Step 2: We next exclude forms of "learning" made possible via exclusive use of reasoning and communication powers in nonhuman animals, and set a focus on learning enabled by human-level-and-above (HLAB) reasoning and communication powers. (Given the previous two subsections, this step makes perfect sense. Recall our discussion, for example, regarding Luger's layout for learning in modern AI, all of which, save for what might be possible in the future, made no use whatsoever of the human capacity to read.)

Step 3: Within the focus arising from Step 2, we next avail ourselves of basic facts of cognitive development in order to narrow the focus to HLAB reasoning and communication sufficiently mature to perceive, and be successfully applied to, both (i) cohesive, abstract bodies of declarative content, and to (ii) sophisticated natural-language content. A paradigmatic case of such content would be axiom systems, such as those for geometry routinely introduced in high school. Another such case would be elementary number theory, also introduced routinely in high school; such coverage includes

[25] Note that all occurrences of 'understanding' in the itemized list that follows, in keeping with the psychometric operationalization introduced at the outset in order not to rely on the murky concept of understanding, could be invoked here; but doing so would take much space and time, and be quite inelegant.

the example of the factorial function, with which we started the present paper.[26] Let's denote such reasoning and communication by 'RC$^{h^\star}$.'

Step 4: Finally, we proceed to define real learning = $\mathcal{RL}$ as the acquisition of new knowledge by using RC$^{h^\star}$. For example, forms of reasoning that use sophisticated analogical reasoning, or deduction applied to the axiom system *PA* (see note 26), can be used to allow an agent to *really* learn new things in the formal domain. Of course, the specific account of real learning will always boil down to specifics such as those given in (c1′)–(c3), but we have sought here to put real learning in a broader context, via our tour and, following on that, the four steps now concluded.

7 Final Remarks

We have heard echoes of an objection not explicitly presented and rebutted above; viz., "Perhaps you should do some soul-searching. For does it not simply boggle the mind that, if you're right, real learning hasn't even been seriously targeted by AI, despite all the praise that it receives for machines that 'learn'?!" Well, it *does* boggle the mind. All of us, the authors and all our readers, know quite well what real learning is, and how it came to be that on its shoulders we all arrived at a place that allows us to study and do AI: we got here by learning in precisely the fashion that $\mathcal{RL}$, in its three conditions, prescribes. We thus take ourselves to have simply revealed in the present paper what everyone in their heart of hearts knows: the exuberant claims of today that machine-learning machines learn are, when stacked against how we all learn enough to put ourselves in position to study and do AI, are simply silly. Accordingly, since AI in the new millennium increasingly penetrates the popular consciousness, we recommend that those working to advance non-real forms of ML extend to the public the courtesy of issuing a disclaimer that the type of learning to which they are devoted isn't real learning. This is a public, of course, that thinks of learning in connection not with artificial agents, but with schoolchildren, with high-schoolers, with undergraduates, with those in job-training programs, etc., all these groups being, of course, natural agents in the business of real learning.

Finally, we admit that the case we have delivered herein isn't yet complete, for there is an approach to computation, and an approach to the study of intelligence, neither of which we have discussed in connection with our core claim that contemporary ML isn't real learning. The approach to computation can be called *natural computation*, and the core idea is that nature itself computes (and perhaps *is* computation) (an excellent introduction is provided in Dodig-Crnkovic and Giovagnoli 2013); the approach to intelligence that we have left aside puts a premium on bodies and their interconnection with the physical environment (see e.g. Barrett 2015). In subsequent work, we plan to consider the relationship between $\mathcal{RL}$ and forms of learning based on these two intertwined approaches.

[26] Peano Arithmetic (*PA*) is rarely introduced by name in K–12 education, but all the axioms of it, save perhaps for the Induction Schema, are introduced and taught there.

Even now, though, it's safe to say that because $\mathcal{RL}$ takes little to no account of the physical (it's after all based in the formal sciences), and because it's conception of an agent is of a disembodied one,[27] it's highly unlikely that forms of physical-and-embodied learning not considered above will overlap real learning.

8 Appendix: The Formal Method

The following deduction uses fonts in an obvious and standard way to sort between functions ($\mathfrak{f}$), agents ($\mathfrak{a}$), and computing machines ($\mathfrak{m}$) in the Arithmetical Hierarchy. Ordinary italicized Roman is used for particulars under these sorts (e.g. f is a particular function). In addition, '$\mathcal{C}$' denotes any collection of conditions constituting jointly necessary-and-sufficient conditions for a form of current ML, which can come from relevant textbooks (e.g. Luger 2008; Russell and Norvig 2009) or papers; we leave this quite up to the reader, as no effect upon the validity of the deductive inference chain will be produced by the preferred instantiation of '$\mathcal{C}$.' It will perhaps be helpful to the reader to point out that the deduction eventuates in the proposition that no machine in the ML fold that in this style learns a relevant function $\mathfrak{f}$ thereby also real-learns $\mathfrak{f}$. We encode this target as follows:

$$(\star)\ \neg\exists\mathfrak{m}\ \exists\mathfrak{f}\ [\phi \coloneqq\ MLlearns(\mathfrak{m},\mathfrak{f}) \wedge \psi \coloneqq\ RLlearns(\mathfrak{m},\mathfrak{f}) \wedge \mathcal{C}_\phi(\mathfrak{m},\mathfrak{f}) \vdash^*\ (\mathrm{c1}') - (\mathrm{c3})_\psi(\mathfrak{m},\mathfrak{f})]$$

Note that $(\star)$ employs meta-logical machinery to refer to particular instantiations of $\mathcal{C}$ for a particular, arbitrary case of ML (ϕ is the atomic sub-formula that can be instantiated to make the particular case), and particular instantiations of the triad (ci$'$)–(ciii) for a particular, arbitrary case of $\mathcal{RL}$ (ψ is the atomic sub-formula that can be instantiated to make the particular case). Meta-logical machinery also allows us to use a provability predicate to formalize the notion that real learning is produced by the relevant instance of ML. If we "pop" ϕ/ψ to yield ϕ'/ψ' we are dealing with the particular instantiation of the atomic sub-formula.

The deduction, as noted earlier when the informal argument was given, is indirect proof by cases; accordingly, we first assume $\neg(\star)$, and then proceed as follows under this supposition.

[27] This conception matches that of an agent in orthodox AI: see the textbooks, e.g. (Luger 2008; Russell and Norvig 2009).

	(1)	$\forall\, \mathfrak{f}, \mathfrak{a}\ [\mathfrak{f} : \mathbb{N} \mapsto \mathbb{N} \rightarrow (\mathit{RLlearns}(\mathfrak{a}, \mathfrak{f}) \rightarrow \text{(c1}'\text{)–(c3)})]$	Def of Real Learning
	(2)	$\mathit{MLlearns}(m, f) \wedge \mathit{RLlearns}(m, f) \wedge f : \mathbb{N} \mapsto \mathbb{N}$	supp (for $\exists$ elim on $\neg(\star)$)
	(3)	$\forall\, \mathfrak{m}, \mathfrak{f}\ [\mathfrak{f} : \mathbb{N} \mapsto \mathbb{N} \rightarrow (\mathit{MLlearns}(\mathfrak{m}, \mathfrak{f}) \leftrightarrow \mathcal{C}(\mathfrak{m}, \mathfrak{f}))]$	Def of ML
	(4)	$\forall\, \mathfrak{f}\ [\mathfrak{f} : \mathbb{N} \mapsto \mathbb{N} \rightarrow (\mathit{TurComp}(\mathfrak{f}) \vee \mathit{TurUncomp}(\mathfrak{f}))]$	theorem
	(5)	$\mathit{TurUncomp}(f)$	supp; Case 1
	(6)	$\neg\exists\ \mathfrak{m}\ \exists\ \mathfrak{f}\ [(\mathfrak{f} : \mathbb{N} \mapsto \mathbb{N} \wedge \mathit{TurUncomp}(\mathfrak{f}) \wedge \mathcal{C}(\mathfrak{m}, \mathfrak{f})]$	theorem
$\therefore$	(7)	$\neg\exists\ \mathfrak{m}\ \mathit{MLlearns}(\mathfrak{m}, f)$	(6), (3)
$\therefore$	(8)	$\bot$	(7), (2)
	(9)	$\mathit{TurComp}(f)$	supp; Case 2
$\therefore$	(10)	$\mathcal{C}_{\phi'}(m, f)$	(2), (3)
$\therefore$	(11)	$\text{(c1}'\text{)–(c3)}_{\psi'}(m, f)$	from supp for $\exists$ elim on $\neg(\star)$ and provability
$\therefore$	(12)	$\neg\text{(c1}'\text{)–(c3)}_{\psi'}(m, f)$	inspection: proofs wholly absent from $\mathcal{C}$
$\therefore$	(13)	$\bot$	(11), (12)
$\therefore$	(14)	$\bot$	*reductio*; proof by cases

A final remark to end the present Appendix: Note that the explicit deductive argument given immediately above conveys a general **method**, m, for showing that real learning = $\mathcal{RL}$ can't be achieved by other forms of limited learning. (Methods, or proof methods, are generalized proof "recipes" that can be composed and built up like computer programs. Proof methods were first introduced in (Arkoudas 2000), and extensive usage of proof methods can be found in (Arkoudas and Musser 2017).) This method m, given suitable input, produces a valid formal proof. All that needs to be done in order to follow the method is to shift out the set $\mathcal{C}$ of conditions to some other set $\mathcal{C}'$ that captures some alternative kind of ML, i.e. some alternative kind of limited learning *Xlearning*. For instance, Bayesian learning (*Blearning*) can by this method be proved to fail to yield real learning in a machine (or agent) that employs *Blearning*.

Acknowledgement. We are deeply appreciative of feedback received at PT-AI 2017, the majority of which is addressed herein. The first author is also specifically indebted to John Hummel for catalyzing, in vibrant discussions at MAICS 2017, the search for formal arguments and/or theorems establishing the proposition Hummel and Bringsjord co-affirm: viz., statistical machine learning simply doesn't enable machines to actually learn, period. Bringsjord is also thankful to Sergei Nirenburg for valuable conversations. Many readers of previous drafts have been seduced by it's-not-really-learning forms of learning (including worse-off-than artificial neural network (ANN) based deep learning (DL) folks: Bayesians), and have offered spirited objections, all of which are refuted herein; yet we are grateful for the valiant tries. Bertram Malle stimulated and guided our sustained study of types of learning in play in psychology^{+}, and we are thankful. Jim Hendler graciously read an early draft; his resistance has been helpful (though perhaps now he's a convert). The authors are also grateful for five anonymous reviews, some portions of which reflected at least partial and passable understanding of our logico-mathematical perspective, from which informal notions of "learning" are inadmissible in such debates as the present one. Two perspicacious comments and observations from two particular PT-AI 2017 participants, subsequent to the conference, proved productive to deeply ponder. We acknowledge the invaluable support of "Advanced Logicist Machine Learning" from ONR, and of "Great Computational

Intelligence" from AFOSR. Finally, without the wisdom, guidance, leadership, and raw energy of Vincent Müller, PT-AI 2017, and any ideas of ours that have any merit at all, and that were expressed there and/or herein, would not have formed.

References

Achab, M., Bacry, E., Gaïffas, S., Mastromatteo, I., Muzy, J.F.: Uncovering causality from multivariate Hawkes integrated cumulants. In: Precup, D., Teh, Y.W. (eds) Proceedings of the 34th International Conference on Machine Learning, PMLR, International Convention Centre, Sydney, Australia. Proceedings of Machine Learning Research, vol. 70, pp. 1–10 (2017). http://proceedings.mlr.press/v70/achab17a.html

Arkoudas, K.: Denotational proof languages. Ph.D. thesis, MIT (2000)

Arkoudas, K., Bringsjord, S.: Computers, justification, and mathematical knowledge. Minds Mach. **17**(2), 185–202 (2007)

Arkoudas, K., Musser, D.: Fundamental Proof Methods in Computer Science: A Computer-Based Approach. MIT Press, Cambridge (2017)

Bandura, A., Walters, R.H.: Social Learning Theory, vol. 1. Prentice-Hall, Englewood Cliffs (1977)

Bandura, A., Ross, D., Ross, S.A.: Transmission of aggression through imitation of aggressive models. J. Abnorm. Soc. Psychol. **63**(3), 575 (1961)

Barrett, L.: Beyond the Brain: How Body and Environment Shape Animal and Human Minds. Princeton University Press, Princeton (2015)

Bellman, A., Bragg, S., Handlin, W.: Algebra 2: Common Core. Pearson, Upper Saddle River (2012). Series Authors: Charles, R., Kennedy, D., Hall, B., Consulting Authors: Murphy, S.G

Boolos, G.S., Burgess, J.P., Jeffrey, R.C.: Computability and Logic, 4th edn. Cambridge University Press, Cambridge (2003)

Bringsjord, S.: What Robots Can and Can't Be. Kluwer, Dordrecht (1992)

Bringsjord, S.: Chess is too easy. Technol. Rev. **101**(2), 23–28 (1998). http://kryten.mm.rpi.edu/SELPAP/CHESSEASY/chessistooeasy.pdf

Bringsjord, S.: Psychometric artificial intelligence. J. Exp. Theor. Artif. Intell. **23**(3), 271–277 (2011)

Bringsjord, S.: The symbol grounding problem-remains unsolved. J. Exp. Theor. Artif. Intell. **27**(1), 63–72 (2015). https://doi.org/10.1080/0952813X.2014.940139

Bringsjord, S., Arkoudas, K.: The modal argument for hypercomputing minds. Theor. Comput. Sci. **317**, 167–190 (2004)

Bringsjord, S., Noel, R.: Real robots and the missing thought experiment in the Chinese room dialectic. In: Preston, J., Bishop, M. (eds.) Views into the Chinese Room: New Essays on Searle and Artificial Intelligence, pp. 144–166. Oxford University Press, Oxford (2002)

Bringsjord, S., Schimanski, B.: What is artificial intelligence? psychometric AI as an answer. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003), pp. 887–893. Morgan Kaufmann, San Francisco (2003). http://kryten.mm.rpi.edu/scb.bs.pai.ijcai03.pdf

Bringsjord, S., Zenzen, M.: Superminds: People Harness Hypercomputation, and More. Kluwer Academic Publishers, Dordrecht (2003)

Bringsjord, S., Bringsjord, E., Noel, R.: In defense of logical minds. In: Proceedings of the 20th Annual Conference of the Cognitive Science Society, pp. 173–178. Lawrence Erlbaum, Mahwah (1998)

Bringsjord, S., Kellett, O., Shilliday, A., Taylor, J., van Heuveln, B., Yang, Y., Baumes, J., Ross, K.: A new Gödelian argument for hypercomputing minds based on the busy beaver problem. Appl. Math. Comput. **176**, 516–530 (2006)

Chisholm, R.: Theory of Knowledge. Prentice-Hall, Englewood Cliffs (1966)

Davis, M., Weyuker, E.: Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science, 1st edn. Academic Press, New York (1983)

Dodig-Crnkovic, G., Giovagnoli, R. (eds.): Computing Nature: Turing Centenary Perspective. Springer, Berlin (2013). https://www.springer.com/us/book/9783642372247

Domjan, M.: The Principles of Learning and Behavior, 7th edn. Cengage Learning, Stamford (2015)

Gallistel, C.R.: Learning and representation. In: Learning and Memory: A Comprehensive Reference, vol. 1. Elsevier (2008) https://doi.org/10.1016/j.neuron.2017.05.021

Gettier, E.: Is justified true belief knowledge? Analysis **23**, 121–123 (1963). http://www.ditext.com/gettier/gettier.html

Gold, M.: Limiting recursion. J. Symb. Logic **30**(1), 28–47 (1965)

Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016). http://www.deeplearningbook.org

Goodstein, R.: On the restricted ordinal theorem. J. Symb. Logic **9**(31), 33–41 (1944)

Huitt, W.: Classroom Instruction. Educational Psychology Interactive (2003)

Hummel, J.: Symbolic versus associative learning. Cogn. Sci. **34**(6), 958–965 (2010)

Inhelder, B., Piaget, J.: The Growth of Logical Thinking from Childhood to Adolescence. Basic Books, New York (1958)

Jain, S., Osherson, D., Royer, J., Sharma, A.: Systems That Learn: An Introduction to Learning Theory, 2nd edn. MIT Press, Cambridge (1999)

Kearns, M., Vazirani, U.: An Introduction to Computational Learning Theory. MIT Press, Cambridge (1994)

Kitzelmann, E.: Inductive programming: a survey of program synthesis techniques. In: International Workshop on Approaches and Applications of Inductive Programming, pp 50–73. Springer (2009)

Lathrop, R.: On the learnability of the uncomputable. In: Saitta, L. (ed.) Proceedings of the 13th International Conference on Machine Learning, The conference was held in Italy, 3–6 July 1996, pp 302–309. Morgan Kaufman, San Francisco (1996). https://pdfs.semanticscholar.org/6919/b6ad91d9c3aa47243c3f641ffd30e0918a46.pdf

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)

Luger, G.: Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 6th edn. Pearson, London (2008)

Mackintosh, N.J.: Conditioning and Associative Learning. Calendron Press, Oxford (1983)

Marblestone, A.H., Wayne, G., Kording, K.P.: Toward an integration of deep learning and neuroscience. Front. Comput. Neurosci. **10**(94) (2016). https://doi.org/10.3389/fncom.2016.00094

Moore, R.C.: Making the transition to formal proof. Educ. Stud. Math. **27**(3), 249–266 (1994)

Müller, V., Bostrom, N.: Future progress in artificial intelligence: a survey of expert opinion. In: Müller, V. (ed.) Fundamental Issues of Artificial Intelligence (Synthese Library), pp. 553–571. Springer, Berlin (2016)

Penn, D., Holyoak, K., Povinelli, D.: Darwin's mistake: explaining the discontinuity between human and nonhuman minds. Behav. Brain Sci. **31**, 109–178 (2008)

Putnam, H.: Trial and error predicates and a solution to a problem of Mostowski. J. Symbolic Logic **30**(1), 49–57 (1965)

Rado, T.: On non-computable functions. Bell Syst. Tech. J. **41**, 877–884 (1963)

Ross, J.: Immaterial aspects of thought. J. Philos. **89**(3), 136–150 (1992)

Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall, Upper Saddle River (2009)

Schapiro, A., Turk-Browne, N.: Statistical learning. Brain Mapp. Encyclopedic Ref. **3**, 501–506 (2015)

Searle, J.: Minds, brains and programs. Behav. Brain Sci. **3**, 417–424 (1980)

Shalev-Shwartz, S., Ben-David, S.: Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, Cambridge (2014)

Stewart, J.: Calculus. 8th edn. Cengage Learning, Boston (2016), We refer here to an electronic version of the print textbook. The "Student Edition" of the hard-copy textbook has an ISBN 978-1-305-27176-0

Titley, H.K., Brunel, N., Hansel, C.: Toward a neurocentric view of learning. Neuron **95**(1), 19–32 (2017)

Turing, A.: Computing machinery and intelligence. Mind LIX **59**(236), 433–460 (1950)

Wason, P., Johnson-Laird, P.: Psychology of Reasoning: Structure and Content. Harvard University Press, Cambridge (1972)

Wolpert, D.H.: The lack of a priori distinctions between learning algorithms. Neural Comput. **8**(7), 1341–1390 (1996)