The Future of AI: Gödel's Either/Or (*not* Kierkegaard's); What about consciousness? (Westworld); Machine Learning!

Selmer Bringsjord

Are Humans Rational?

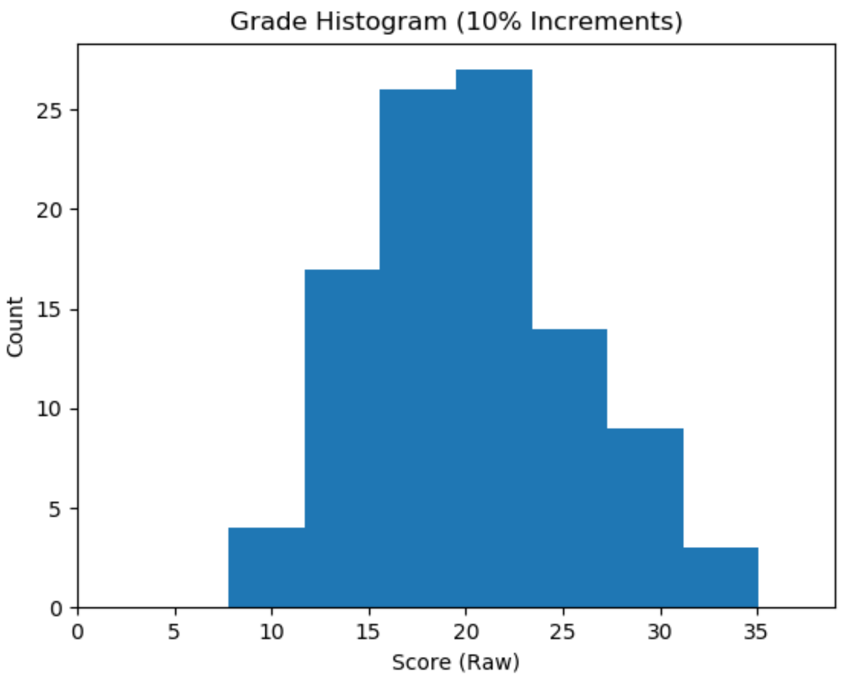
10/7/19

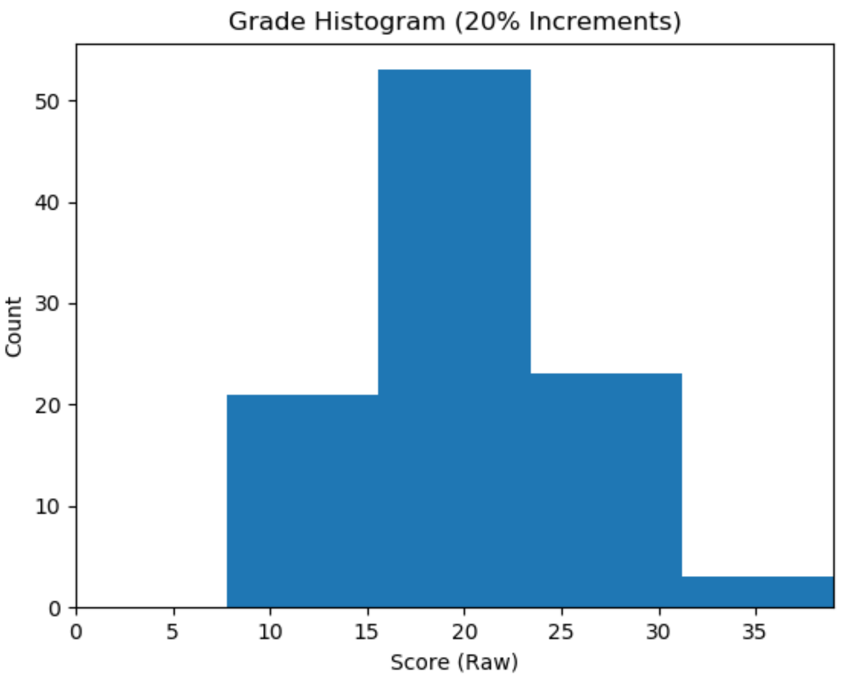
Selmer.Bringsjord@gmail.com

RAIR

Rensselaer AI and Reasoning Lab

How to assign letter grades? — rationally





Gödel's Either/Or …

The Question

Q* Is the human mind more powerful than the class of standard computing machines?

Gödel's Either/Or

"[E]ither … the human mind (even within the realm of pure mathematics) infinitely surpasses the power of any finite machine, or else there exist absolutely unsolvable diophantine problems."
— Gödel, 1951

Background

problem?[7] In his lecture, Gödel precisely defines diophantine problems, but we don't need to bother with all of the details here; we only need to appreciate the general structure of such a problem, and that can be achieved quickly as follows, given what was introduced in Chapter 2.

Each diophantine problem has at its core a polynomial $\mathcal{P}$ whose variables are comprised by two lists, $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$; all variables must be integers, and the same for subscripts n and m. To represent a polynomial in a manner that announces its variables, we can write

$$\mathcal{P}(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_j).$$

But Gödel was specifically interested in whether, for all integers that can be set to the variables x_i, there are integers that can be set to the y_j, such that the polynomial equals 0. To make this clearer, first, here are two particular, simple equations that employ polynomials that are both instances of the needed form:

E1 $\quad 3x - 2y = 0$

E2 $\quad 2x^2 - y = 0$

All we need to do now is prefix these equations with quantifiers in the pattern Gödel gave. This pattern is quite simple: universally quantify over each x_i variable (using the now-familiar $\forall$), after which we existentially quantify over each y_i variable (using the also-now-familiar $\exists$). Thus, here are the two diophantine problems that correspond to the pair E1 and E2 from just above:

P1 $\quad$ Is it true that $\forall x \exists y (3x - 2y = 0)$?

P2 $\quad$ Is it true that $\forall x \exists y 2x^2 - y = 0$?

The Crux

$\exists \mathcal{P}$ s.t. no human mind could ever decide $\forall x_1 \forall x_2 \cdots \forall x_k \exists y_1 \exists y_2 \cdots \exists x_j (\mathcal{P}(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_j)$?

The Crux

$\exists \mathcal{P}$ s.t. no human mind could ever decide $\forall x_1 \forall x_2 \cdots \forall x_k \exists y_1 \exists y_2 \cdots \exists x_j (\mathcal{P}(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_j)$?

Yes.

The Crux

$\exists \mathcal{P}$ s.t. no human mind could ever decide $\forall x_1 \forall x_2 \cdots \forall x_k \exists y_1 \exists y_2 \cdots \exists x_j (\mathcal{P}(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_j)$?

Yes.

The Crux

$\exists \mathcal{P}$ s.t. no human mind could ever decide $\forall x_1 \forall x_2 \cdots \forall x_k \exists y_1 \exists y_2 \cdots \exists x_j (\mathcal{P}(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_j)$?

Yes.

The human mind is *not* infinitely more powerful than any standard computing machine.

The Crux

$\exists \mathcal{P}$ s.t. no human mind could ever decide $\forall x_1 \forall x_2 \cdots \forall x_k \exists y_1 \exists y_2 \cdots \exists x_j (\mathcal{P}(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_j)$?

Yes. No.

The human mind is *not* infinitely more powerful than any standard computing machine.

The Crux

$\exists \mathcal{P}$ s.t. no human mind could ever decide $\forall x_1 \forall x_2 \cdots \forall x_k \exists y_1 \exists y_2 \cdots \exists x_j (\mathcal{P}(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_j)$?

Yes.

No.

The human mind is *not* infinitely more powerful than any standard computing machine.

The Crux

$\exists \mathcal{P}$ s.t. no human mind could ever decide $\forall x_1 \forall x_2 \cdots \forall x_k \exists y_1 \exists y_2 \cdots \exists x_j (\mathcal{P}(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_j)$?

Yes.

No.

The human mind *is* infinitely more powerful than any standard computing machine.

The human mind is *not* infinitely more powerful than any standard computing machine.

AI & Consciousness …


WESTWORLD
OCTOBER 2 9PM
HBO

First-rate sci fi?

Bona fide art?

NEaF Artifacts

Profound Art

Mimetic Art

“Art of” Art

Dabbling

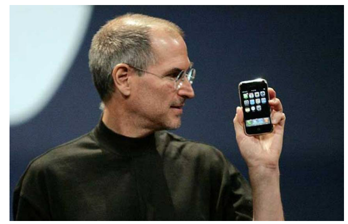
Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

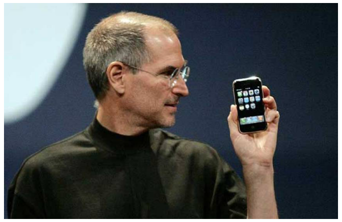
Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art



Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art



Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

Dabbling	Mimetic Art	NEaF Artifacts	"Art of" Art	Profound Art

Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

Dabbling	Mimetic Art	NEaF Artifacts	“Art of” Art	Profound Art

These violent delights have violent ends
And in their triumph die, like fire and powder,
Which as they kiss consume: the sweetest honey
Is loathsome in his own deliciousness
And in the taste confounds the appetite:
Therefore love moderately; long love doth so;
Too swift arrives as tardy as too slow

Dabbling	Mimetic Art	NEaF Artifacts	"Art of" Art	Profound Art

These violent delights have violent ends
And in their triumph die, like fire and powder,
Which as they kiss consume: the sweetest honey
Is loathsome in his own deliciousness
And in the taste confounds the appetite:
Therefore love moderately; long love doth so;
Too swift arrives as tardy as too slow

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Visitor-22 believes that Dolores is devoted to her father.

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Visitor-22 believes that Dolores is devoted to her father.

$$\mathfrak{E}(\mathbf{S})$$

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Visitor-22 believes that Dolores is devoted to her father.

$$\mathfrak{E}(\mathbf{S})\ \checkmark$$

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Visitor-22 believes that Dolores is devoted to her father.

$$\mathfrak{E}(\mathbf{S})\ \checkmark$$

$$\Phi_k + c \in \mathcal{C} \text{ desires, plans, acts } \Rightarrow \text{ consequences } \Rightarrow \Phi'_{k+1}$$

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Visitor-22 believes that Dolores is devoted to her father.

$$\mathfrak{E}(\mathbf{S})\ \checkmark$$

$$\Phi_k + c \in \mathcal{C} \text{ desires, plans, acts } \Rightarrow \text{ consequences } \Rightarrow \Phi'_{k+1}$$

$$\mathbf{S}' = \Phi'_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi'_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi'_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}] \; \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}] \; \cdots \; \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Visitor-22 believes that Dolores is devoted to her father.

$$\mathfrak{E}(\mathbf{S}) \quad \surd$$

$$\Phi_k + c \in \mathcal{C} \text{ desires, plans, acts } \Rightarrow \text{ consequences } \Rightarrow \Phi'_{k+1}$$

$$\mathbf{S}' = \Phi'_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}] \; \Phi'_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}] \; \cdots \; \Phi'_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

$$\vdash \mathfrak{E}(\mathbf{S}')?$$

Westworld's Improv

$$\mathbf{S} = \Phi_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

Visitor-22 believes that Dolores is devoted to her father.

$$\mathfrak{E}(\mathbf{S}) \ \checkmark$$

$$\Phi_k + c \in \mathcal{C} \text{ desires, plans, acts } \Rightarrow \text{ consequences } \Rightarrow \Phi'_{k+1}$$

$$\mathbf{S}' = \Phi'_1[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \Phi'_2[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]\ \cdots\ \Phi'_n[\mathcal{C}, \mathcal{A}, \mathcal{G}, \mathcal{M}]$$

$$\boxed{\vdash \mathfrak{E}(\mathbf{S}')?}$$

Harder than the *Entscheidungsproblem*!

“Computer, is this a theorem??”
i.e., the *Entscheidungsproblem*

"Computer, is this a theorem??"

i.e., the *Entscheidungsproblem*

“Computer, is this a theorem??”
i.e., the *Entscheidungsproblem*

“Computer, is this a theorem??”
i.e., the *Entscheidungsproblem*

“Computer, is this a theorem??”
i.e., the *Entscheidungsproblem*

$$\vdash \mathfrak{E}(\mathbf{S}')$$

$\longrightarrow$ *input*

$\longrightarrow$ *output*

What about consciousness …

What about consciousness …

What about consciousness …

search WWW using: `consciousness AI jaynes westworld`

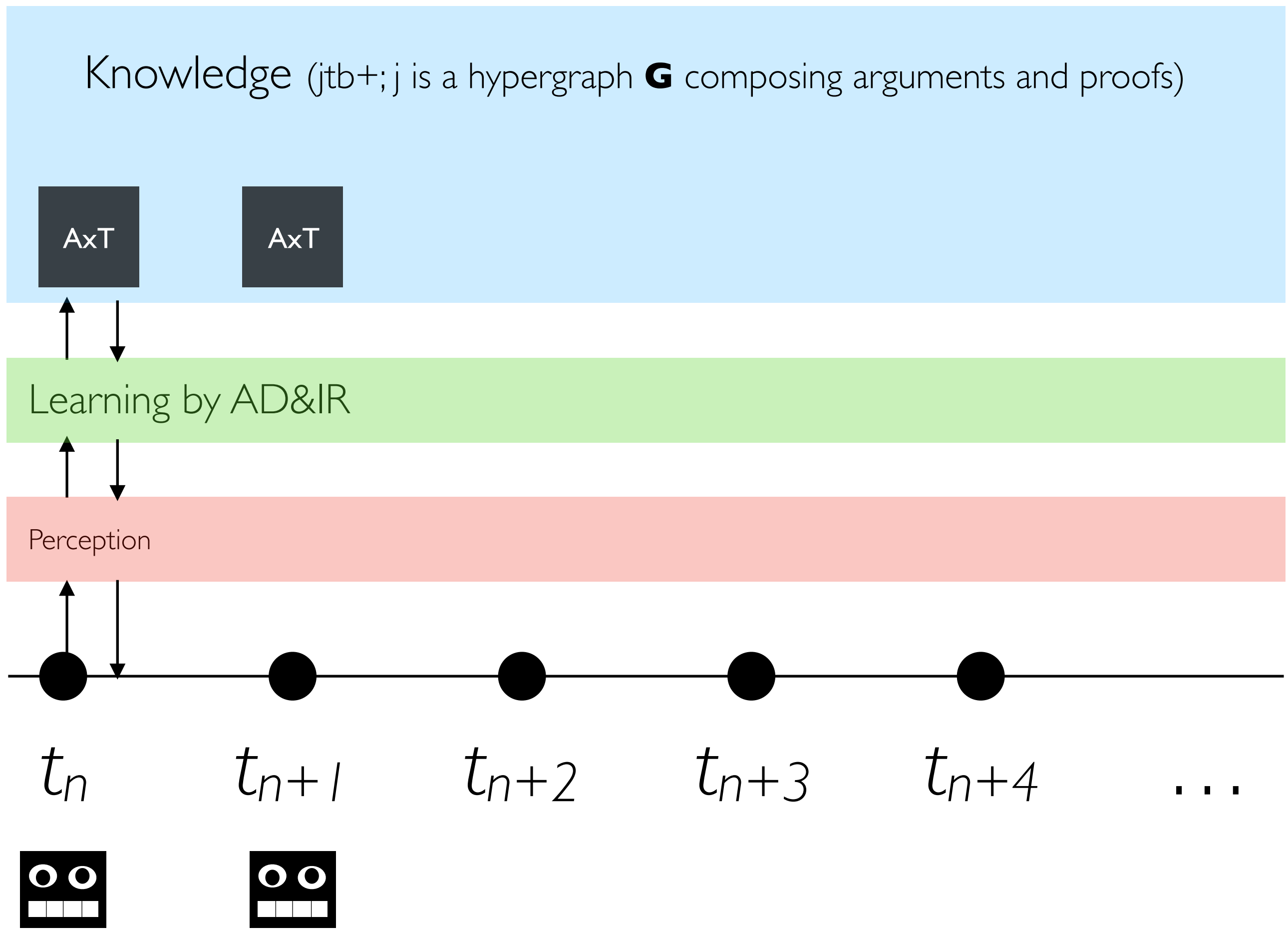
Machine Learning …

Career Advice

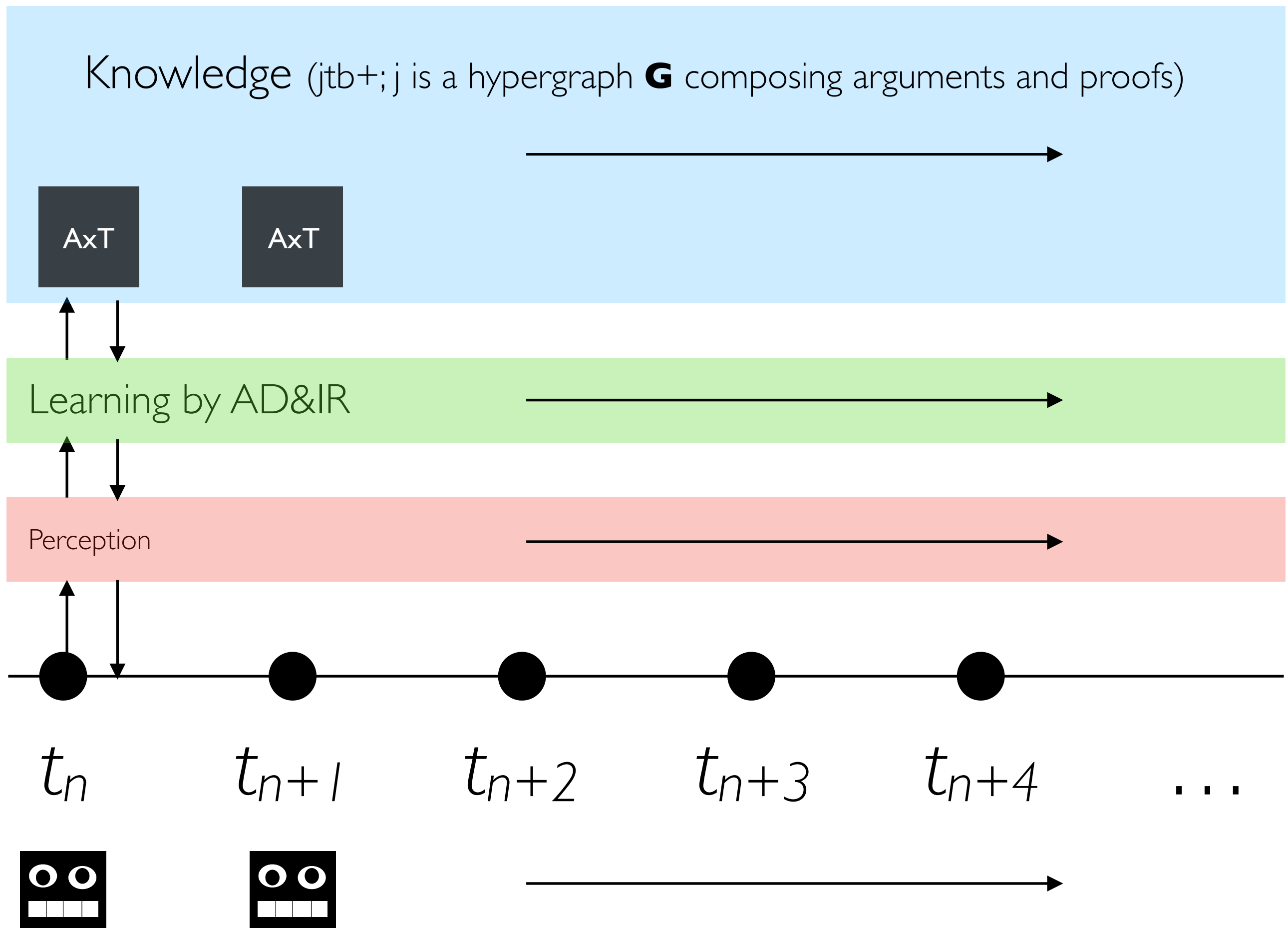
Career Advice

Master the current math for ML, & ponder the beyond (it).

Advanced Logicist (Real) Machine Learning



Advanced Logicist (Real) Machine Learning



Do Machine-Learning Machines Learn? No.

Do Machine-Learning Machines Learn? No.

Do Machine-Learning Machines Learn?

Selmer Bringsjord and Naveen Sundar Govindarajulu and Shreya Banerjee and John Hummel

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 $\mathcal{RL}$.

Selmer Bringsjord
Rensselaer Polytechnic Institute, 110 8th Street Troy, NY USA 12180, e-mail: selmerbringsjord@gmail.com

Naveen Sundar Govindarajulu
Rensselaer Polytechnic Institute, 110 8th Street Troy, NY USA 12180, e-mail: Naveen.Sundar.G@gmail.com

Shreya Banerjee
Rensselaer Polytechnic Institute, 110 8th Street Troy, NY USA 12180, e-mail: shreyabbanerjee@gmail.com

John Hummel
901 West Illinois Street, Urbana, IL 61801, e-mail: jehummel@illinois.edu

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:

$$(\star)\quad \neg\exists\mathfrak{m}\,\exists\mathfrak{f}\,[\phi := \mathit{MLlearns}(\mathfrak{m},\mathfrak{f}) \wedge \psi := \mathit{RLlearns}(\mathfrak{m},\mathfrak{f}) \wedge \mathcal{C}_\phi(\mathfrak{m},\mathfrak{f}) \vdash^* (\mathrm{ci}')\text{–}(\mathrm{ciii})_\psi(\mathfrak{m},\mathfrak{f})]$$

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 in 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.

	(1)	$\forall\,\mathfrak{f},\mathfrak{a}\;[\mathfrak{f}:\mathbb{N}\mapsto\mathbb{N} \to (\mathit{RLlearns}(\mathfrak{a},\mathfrak{f}) \to \text{(i)–(iii)})]$	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 $(\star)$)
	(3)	$\forall\,\mathfrak{m},\mathfrak{f}\;[\mathfrak{f}:\mathbb{N}\mapsto\mathbb{N} \to (\mathit{MLlearns}(\mathfrak{m},\mathfrak{f}) \leftrightarrow \mathcal{C}(\mathfrak{m},\mathfrak{f}))]$	Def of ML
	(4)	$\forall\,\mathfrak{f}\;[\mathfrak{f}:\mathbb{N}\mapsto\mathbb{N} \to (\mathit{TurComp}(\mathfrak{f}) \vee \mathit{TurUncomp}(\mathfrak{f}))]$	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)	$(\mathrm{ci}')\text{–}(\mathrm{ciii})_{\psi'}(m,f)$	from supp for $\exists$ elim on $(\star)$ and provability
$\therefore$	(12)	$\neg(\mathrm{ci}')\text{–}(\mathrm{ciii})_{\psi'}(m,f)$	inspection: proofs wholly absent from $\mathcal{C}$
$\therefore$	(13)	$\bot$	(11), (12)
$\therefore$	(14)	$\bot$	*reductio*; proof by cases

What is Real Learning (RL)?

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.

What is Real Learning (RL)?

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.

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

The Four-Step Road to Real Learning

The Four-Step Road to Real Learning

- Step 1: Observe the acute discontinuity of human vs. nonhuman cognition. (Only humans understand and employ e.g. abstract reasoning schemas unaffected by the physical; layered quantification; recursion; and infinite structures/infinitary reasoning.)

The Four-Step Road to Real Learning

- Step 1: Observe the acute discontinuity of human vs. nonhuman cognition. (Only humans understand and employ e.g. abstract reasoning schemas unaffected by the physical; layered quantification; recursion; and infinite structures/infinitary reasoning.)

- Step 2: Exclude forms of "learning" made possible via exclusive use of reasoning and communication capacities in nonhuman animals (i.e. exclude forms of "learning" that don't eventuate in bona fide *knowledge*).

The Four-Step Road to Real Learning

- Step 1: Observe the acute discontinuity of human vs. nonhuman cognition. (Only humans understand and employ e.g. abstract reasoning schemas unaffected by the physical; layered quantification; recursion; and infinite structures/infinitary reasoning.)

- Step 2: Exclude forms of "learning" made possible via exclusive use of reasoning and communication capacities in nonhuman animals (i.e. exclude forms of "learning" that don't eventuate in bona fide *knowledge*).

- Step 3: Within the focus arising from Step 2, further narrow the focus to HLAB reasoning and communication sufficiently powerful to perceive, and be successfully applied to, both (i) cohesive bodies of declarative content, and (ii) sophisticated natural-language content. Dub this **RC**.

The Four-Step Road to Real Learning

- Step 1: Observe the acute discontinuity of human vs. nonhuman cognition. (Only humans understand and employ e.g. abstract reasoning schemas unaffected by the physical; layered quantification; recursion; and infinite structures/infinitary reasoning.)

- Step 2: Exclude forms of "learning" made possible via exclusive use of reasoning and communication capacities in nonhuman animals (i.e. exclude forms of "learning" that don't eventuate in bona fide *knowledge*).

- Step 3: Within the focus arising from Step 2, further narrow the focus to HLAB reasoning and communication sufficiently powerful to perceive, and be successfully applied to, both (i) cohesive bodies of declarative content, and (ii) sophisticated natural-language content. Dub this **RC**.

- Step 4: Real Learning ($\mathscr{RL}$) is the acquisition of genuine knowledge via **RC**.

But how is this mechanizable? How about a new form of machine learning? (by reasoning)

Learning *Ex Nihilo*

(or Learning *Ex Minima*)

Learning *Ex Nihilo*

Selmer Bringsjord[1], **Naveen Sundar Govindarajulu**[2], **John Licato**[3]

[1,2]Rensselaer AI & Reasoning (RAIR) Lab
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA

[3]Advancing Machine and Human Reasoning (AMHR) Lab
University of South Florida
Tampa FL 33620 USA

selmer.bringsjord@gmail.com, naveensundarg@gmail.com, john.licato@gmail.com

Abstract

This paper introduces, philosophically and to a degree formally, the novel concept of learning *ex nihilo*, intended (obviously) to be analogous to the concept of creation *ex nihilo*. Learning *ex nihilo* is an agent's learning "from nothing," by the suitable employment of schemata for deductive and inductive reasoning. This reasoning must be in machine-verifiable accord with a formal proof/argument theory in a *cognitive calculus* (i.e., roughly, an intensional higher-order multi-operator quantified logic), and this reasoning is applied to percepts received by the agent, in the context of both some prior knowledge, and some prior and current interests. Learning *ex nihilo* is a challenge to contemporary forms of ML, indeed a severe one, but the challenge is here offered in the spirt of seeking to stimulate attempts, on the part of non-logicist ML researchers and engineers, to collaborate with those in possession of learning-*ex nihilo* frameworks, and eventually attempts to integrate directly with such frameworks at the implementation level. Such integration will require, among other things, the symbiotic interoperation of state-of-the-art automated reasoners and high-expressivity planners, with statistical/connectionist ML technology.

1 Introduction

This paper introduces, philosophically and to a degree logico-mathematically, the novel concept of learning *ex nihilo*, intended (obviously) to be analogous to the concept of creation *ex nihilo*.[1] Learning *ex nihilo* is an agent's learning "from nothing," by the suitable employment of schemata for deductive and inductive[2] (e.g., analogical, enumerative-inductive, abductive, etc.) reasoning. This reasoning must be in machine-verifiable accord with a formal proof/argument theory in a *cognitive calculus*, and this reasoning is applied to percepts received by the agent, in the context of both some prior knowledge, and some prior and current interests. Roughly, cognitive calculi include inferential components of intensional higher-order multi-operator quantified logics, in which expressivity far outstrips off-the-shelf modal logics and possible-worlds semantics, and a number of such calculi have been introduced as bases for AI that is unrelated to learning; e.g. see (Govindarajulu & Bringsjord 2017*a*), where the application area is AI ethics. The very first cognitive calculus, a purely deductive one, replete with a corresponding implementation in ML, was introduced in (Arkoudas & Bringsjord 2009).

Learning *ex nihilo* is a challenge to contemporary forms of ML, indeed a severe one, but the challenge is offered in the spirt of seeking to stimulate attempts, on the part of non-logicist ML researchers and engineers, to collaborate with those in possession of learning *ex nihilo* frameworks, and eventually attempts to integrate directly with such frameworks at the implementation level. Such integration will require, among other things, the symbiotic use of state-of-the-art automated reasoners (such as ShadowReasoner, the particular reasoner that for us powers learning *ex nihilo*) with statistical/connectionist ML technology.

The sequeal unfolds as follows. §2 offers a starting para-

[1]No such assumption as that creation *ex nihilo* is real or even formally respectable is made or needed in the present paper. The concept of creation *ex nihilo* is simply an intellectual inspiration — but as a matter of fact, the literature on it in analytic philosophy does provide some surprisingly rigorous accounts. In the present draft of the present paper, we don't seek to mine these accounts.

[2]Not to be confused with inductive logic programming (about more will be said later), or inductive deductive techniques and schemas (e.g. matheatical induction, the induction schema in Peano Arithmetic, etc.). As we explain later, learning *ex nihilo* is powered by non-deductive inference schemata seeen in inductive logic. An introuctory overview of inductive logic is provided in (Johnson 2016).

Sherlock Explains How He Learned *Ex Nihilo* Quite a Bit About Watson

Sherlock Explains How He Learned *Ex Nihilo* Quite a Bit About Watson

Holmes Now Knows Watson Now Believes He (Watson) Isn't Dealing With an Amateur

```
:assumptions {Premise1 (Knows! holmes
                               (if (Believes! watson
                                              (and (not (Knows! holmes t1 (PersonalFact (inMilitary watson))))
                                                   (Knows! holmes t2 (PersonalFact (inMilitary watson)))))
                                 (Believes! watson (not (Amateur holmes)))))

              Premise2 (Knows! holmes (Knows! watson (not (Knows! holmes t1 (PersonalFact (inMilitary watson))))))


              Premise3 (Knows! holmes
                               (Believes! watson
                                          (Knows! holmes t2
                                                  (if (and (tan watson) (wounded watson)) (PersonalFact (inMilitary watson))))))
              Premise4 (Knows! holmes
                               (Believes! watson
                                          (Knows! holmes t2 (and (tan watson) (wounded watson)))))}

:goal         (Knows! holmes (Believes! watson (not (Amateur holmes))))}
```

Implemented in ShadowProver

```
/Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java ...
objc[25214]: Class JavaLaunchHelper is implemented in both /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java (0x101f0e4c0) and /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/jre/lib/libinstrument.dylib (0x101f9a4e0)
; Running SNARK from /Users/naveensundar/projects/prover/snark-20120808r02/snark-system.lisp in Armed Bear Common Lisp 1.0.1-svn-13751 on Naveens-MacBook-Pro-2.local at 2019-05-28T01:54:51
```

Implemented in ShadowProver

```
/Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java ...
objc[25214]: Class JavaLaunchHelper is implemented in both /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java (0x101f0e4c0) and /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/jre/lib/libinstrument.dylib (0x101f9a4e0)
; Running SNARK from /Users/naveensundar/projects/prover/snark-20120808r02/snark-system.lisp in Armed Bear Common Lisp 1.0.1-svn-13751 on Naveens-MacBook-Pro-2.local at 2019-05-28T01:54:51
```