on horizon for GCI

potential GCI branch

spinoff

synergy

Great Computational Intelligence (GCI), Mature & Further Applied

(FA9550-17-1-0191)

PI: Selmer Bringsjord (RPI)

Co-PI: John Hummel (UIUC)

Co-PI: John Licato (USF)

Key Senior Collaborator:

Dr Naveen Sundar G (RPI)

Student Researchers:

#4 Mike Giancola (RPI)

Mark Bodger (USF)

Rachel Heaton (UIUC)

AFOSR Program Review (Dr James Lawton):
Computational Cognition & Machine Intelligence (CCMI)
(Oct 6–Oct 8 2020 • "Arlington VA" via Zoom)

Invariant Slides …

GCI, Mature & Further Applied (Bringsjord et al.)

Objective

The driving goal of early GCI was the production of computational systems able to prove Gödel's two incompleteness theorems (& other formal results) from first principles with an unprecedentedly high degree of autonomy, in significant part on the strength of integrating deductive reasoning with analogical reasoning. Now, with this integration retained, novel forms of analogical reasoning that are completely non-deductive are being specified, and a number of objectives are targeted, including — on the mathematical-logic front — Gödel's Speedup Theorem from first principles, likewise the Gödel-Rosser Theorem, and an array of (seven) additional ambitious objectives each of which has a detailed SOW.

Technical Approach

This research draws from the state-of-the-art in automated theorem proving/automated reasoning in contemporary AI, from unprecedentedly expressive formal, computational languages for modeling and replicating the best of human reasoning in the formal science, and from computational architectures and corresponding models of human analogical reasoning. Overall, the approach is to inform and guide the best logicist AI available today with up-to-date computational modeling of analogical reasoning at the human level.

Key Scientific Contributions

A new form of machine-discovery and machine-assisted discovery (analogico-deductive reasoning, ADR) invented & validated in domains of mathematical logic, mathematical physics, cyber/nuclear strategy, theory of just war, etc. Now, ADR is extended to automated inductive reasoning (AIR), and applied to a series of challenges that demand AI of unprecedented power.

DoD Benefit

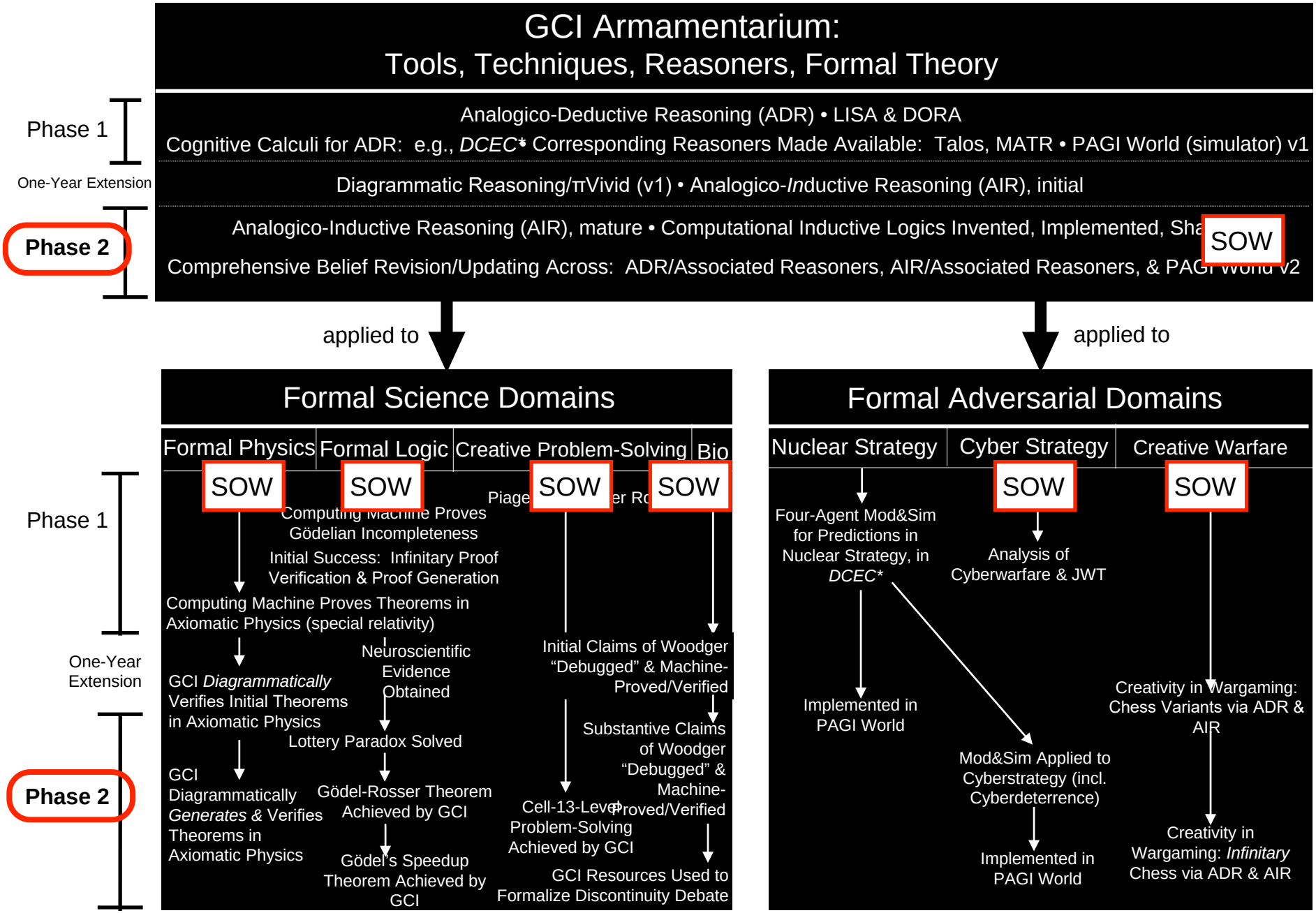
New, powerful class of automated and semi-automated reasoning systems/representational frameworks capable of answering queries over, and making discoveries relative to, complex and intricate information.

Shorter-term benefits will accrue from the deployment of the newly invented systems to nuclear strategy and — coming — creativity in warfare.

List of 8 Project Goals

1. Invent, specify, and implement systems of automated & semi-automated inductive reasoning, to include, first and foremost, automated analogical inductive reasoning (AIR), and on those systems these AIR systems are in turn based on (e.g. ATPs).
2. Invent, specify, and implement systems of automated & semi-automated belief (cognitive-operator) revision through time, based on systems in Goal 1, and on those systems these AIR systems are based on (e.g. ATPs).
3. Using the systems obtained by reaching Goals 1 & 2, enable AI to establish Gödel's Speedup Theorem from first principles, as well as, in like manner, the Gödel-Rosser Theorem.
4. Using the systems obtained by reaching Goals 1 & 2, enable AI to obtain (additional) substantive theorems in formal physics.
5. Using the systems obtained by reaching Goals 1 & 2, enable AI to obtain (additional) theory-of-mind-level problem-solving capacity.
6. Using the systems obtained by reaching Goals 1 & 2, adjudicate the discontinuity debate in biology and comparative psychology.
7. Using the systems obtained by reaching Goals 1 & 2, make further progress in the modeling and simulation of cyber/nuclear strategy.
8. Using the systems obtained by reaching Goals 1 & 2, make further

Great Computational Intelligence (GCI), Mature & Further Applied

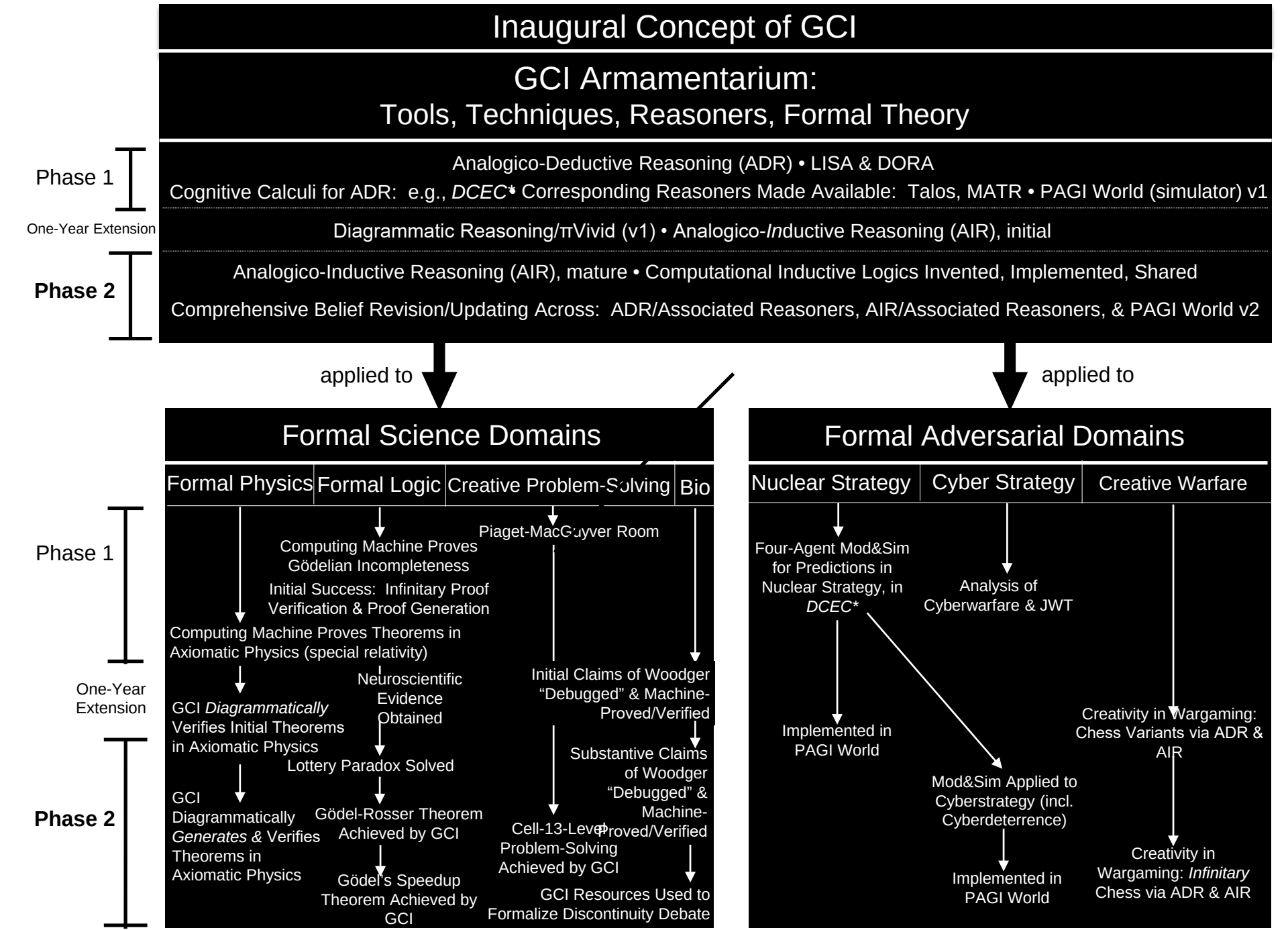


Today's Sequence

1. Review for Context: Concept of GCI, &
    1. Reflecting on Some Hierarchies
    2. Λ (*not* Φ) & our JAIC Paper
    3. Can the formalisms & content be learned?
2. GCI. Formal Phvsics. Gödelian "Insoluble" Time-Travel Paradox Solved
3. Automated Analogical Reasoning (under Automated Inductive Reasoning, *not* Automated Deductive Reasoning): Miracle on the Hudson
4. On Progress Toward Automated Reasoning to Reach/Exploit Gö  s Speedup Theorem
5. On Progress Toward Formalized & Thereby Resolving the "Discontinuity Debate": FBT^{∞}
6. Re Some Pubs
7. For the (precious) few who want more Selmer on AI: https://mindmatters.ai/podcast/ep101

1. Refresher on, & the Nature of, GCI …

Great Computational Intelligence (GCI), Mature & Further Applied



Core Ideas @ Inception:
Great Computational Intelligence ...

In general, for a computational artifact $\mathcal{C}$ to have GCI, we hold that it must produce a result ρ that is,

Significant by at least near-consensus among relevant humans, intrinsically significant;

Independent generated by a problem-solving run carried out to a high degree by $\mathcal{C}$ independent of human insight and assistance; and

Innovative where this problem-solving run begins from a starting point ι that is a "long distance" from ρ.

We shall assume that λ applied to a pair (ι, ρ) yields a distance δ; we therefore write

$$\lambda(\iota, \rho) = \delta.$$

To say that $\mathcal{C}$ produces ρ having started with ι, we write

$$\mathcal{C} : \iota \longrightarrow \rho.$$

We shall further assume that the general space of inputs is $\iota^\star$, and the general space of results $\rho^\star$. Under this notation, it can be informatively said that a good indicator of whether a result is significant is that the function f from $\iota^\star$ to $\rho^\star$ is Turing-unsolvable. Were this indicator promoted to an absolute requirement, which is quite tempting, the first property of GCI could plausibly be formalized via something like the following equation as a necessary condition for this property (significance) to be possessed.[7]

$$\mathcal{C} : \iota \longrightarrow \rho \text{ where the function } f : \iota^\star \longrightarrow \rho^\star \text{ is Turing-unsolvable.} \tag{2}$$

[7] One must be careful here. Let h be a binary halting function taking as input the Gödel number n^M of a Turing machine M along with input m to that Turing machine. As is well-known, h is Turing-uncomputable. Yet there are individual Turing machines, accompanied by inputs to them, which can be instantly declared and proved to be either

In general, for a computational artifact $\mathcal{C}$ to have GCI, we hold that it must produce a result ρ that is,

Significant by at least near-consensus among relevant humans, intrinsically significant;

Independent generated by a problem-solving run carried out to a high degree by $\mathcal{C}$ independent of human insight and assistance; and

Innovative where this problem-solving run begins from a starting point ι that is a "long distance" from ρ.

We shall assume that λ applied to a pair (ι, ρ) yields a distance δ; we therefore write

$$\lambda(\iota, \rho) = \delta.$$

To say that $\mathcal{C}$ produces ρ having started with ι, we write

$$\mathcal{C} : \iota \longrightarrow \rho.$$

We shall further assume that the general space of inputs is $\iota^\star$, and the general space of results $\rho^\star$. Under this notation, it can be informatively said that a good indicator of whether a result is significant is that the function f from $\iota^\star$ to $\rho^\star$ is Turing-unsolvable. Were this indicator promoted to an absolute requirement, which is quite tempting, the first property of GCI could plausibly be formalized via something like the following equation as a necessary condition for this property (significance) to be possessed.[7]

$$\mathcal{C} : \iota \longrightarrow \rho \text{ where the function } f : \iota^\star \longrightarrow \rho^\star \text{ is Turing-unsolvable.} \tag{2}$$

[7] One must be careful here. Let h be a binary halting function taking as input the Gödel number n^M of a Turing machine M along with input m to that Turing machine. As is well-known, h is Turing-uncomputable. Yet there are individual Turing machines, accompanied by inputs to them, which can be instantly declared and proved to be either

We'd said: "Even famous AI systems strike out." — pre AlphaG

In general, for a computational artifact $\mathcal{C}$ to have GCI, we hold that it must produce a result ρ that is,

Significant by at least near-consensus among relevant humans, intrinsically significant;

Independent generated by a problem-solving run carried out to a high degree by $\mathcal{C}$ independent of human insight and assistance; and

Innovative where this problem-solving run begins from a starting point ι that is a "long distance" from ρ.

We shall assume that λ applied to a pair (ι, ρ) yields a distance δ; we therefore write

$$\lambda(\iota, \rho) = \delta.$$

To say that $\mathcal{C}$ produces ρ having started with ι, we write

$$\mathcal{C} : \iota \longrightarrow \rho.$$

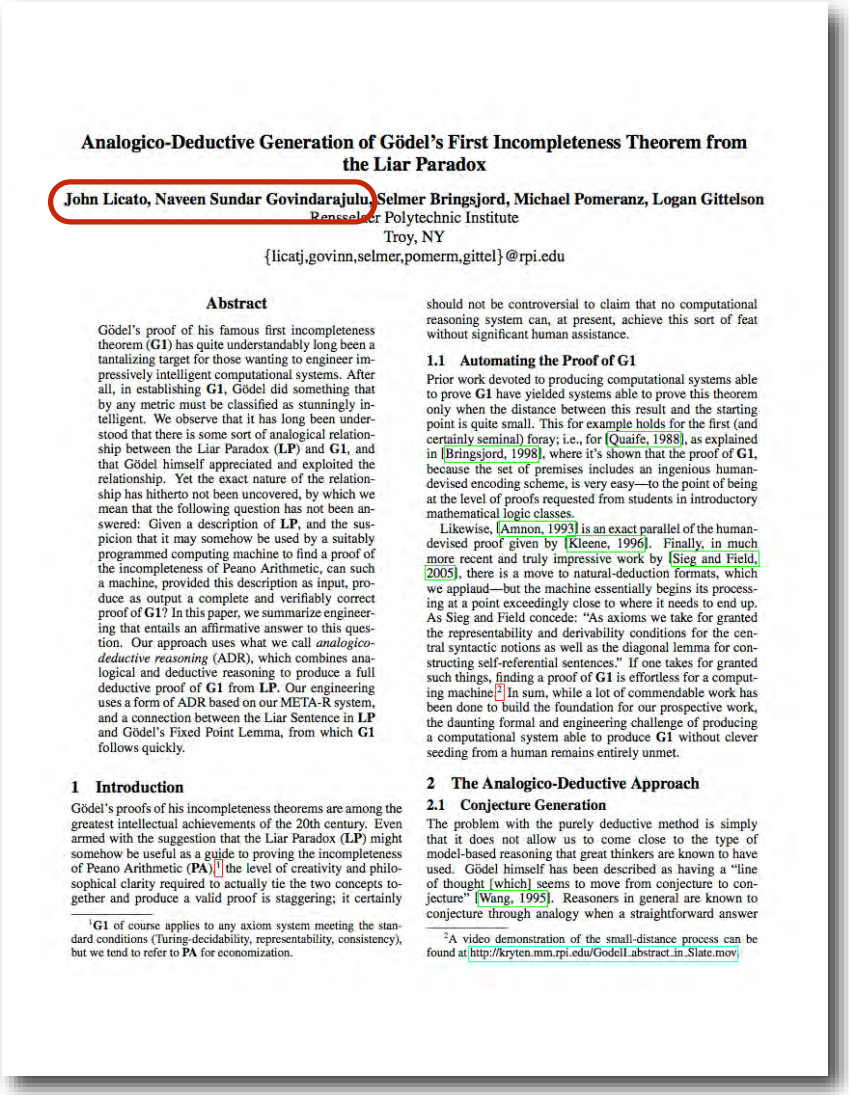
We shall further assume that the general space of inputs is $\iota^\star$, and the general space of results $\rho^\star$. Under this notation, it can be informatively said that a good indicator of whether a result is significant is that the function f from $\iota^\star$ to $\rho^\star$ is Turing-unsolvable. Were this indicator promoted to an absolute requirement, which is quite tempting, the first property of GCI could plausibly be formalized via something like the following equation as a necessary condition for this property (significance) to be possessed.[7]

$$\mathcal{C} : \iota \longrightarrow \rho \text{ where the function } f : \iota^\star \longrightarrow \rho^\star \text{ is Turing-unsolvable.} \tag{2}$$

[7] One must be careful here. Let h be a binary halting function taking as input the Gödel number n^M of a Turing machine M along with input m to that Turing machine. As is well-known, h is Turing-uncomputable. Yet there are individual Turing machines, accompanied by inputs to them, which can be instantly declared and proved to be either

We'd said: "The result must be provably correct (even when won on the strength of *in*ductive reasoning)."

“Classic” ADR Result



Analogico-Deductive Generation of Gödel’s First Incompleteness Theorem from the Liar Paradox

John Licato, Naveen Sundar Govindarajulu, Selmer Bringsjord, Michael Pomeranz, Logan Gittelson
Rensselaer Polytechnic Institute
Troy, NY
{licatj,govinn,selmer,pomerm,gittel}@rpi.edu

Abstract

Gödel’s proof of his famous first incompleteness theorem (**G1**) has quite understandably long been a tantalizing target for those wanting to engineer impressively intelligent computational systems. After all, in establishing **G1**, Gödel did something that by any metric must be classified as stunningly intelligent. We observe that it has long been understood that there is some sort of analogical relationship between the Liar Paradox (**LP**) and **G1**, and that Gödel himself appreciated and exploited the relationship. Yet the exact nature of the relationship has hitherto not been uncovered, by which we mean that the following question has not been answered: Given a description of **LP**, and the suspicion that it may somehow be used by a suitably programmed computing machine to find a proof of the incompleteness of Peano Arithmetic, can such a machine, provided this description as input, produce as output a complete and verifiably correct proof of **G1**? In this paper, we summarize engineering that entails an affirmative answer to this question. Our approach uses what we call *analogico-deductive reasoning* (ADR), which combines analogical and deductive reasoning to produce a full deductive proof of **G1** from **LP**. Our engineering uses a form of ADR based on our META-R system, and a connection between the Liar Sentence in **LP** and Gödel’s Fixed Point Lemma, from which **G1** follows quickly.

1 Introduction

Gödel’s proofs of his incompleteness theorems are among the greatest intellectual achievements of the 20th century. Even armed with the suggestion that the Liar Paradox (**LP**) might somehow be useful as a guide to proving the incompleteness of Peano Arithmetic (**PA**),[1] the level of creativity and philosophical clarity required to actually tie the two concepts together and produce a valid proof is staggering; it certainly should not be controversial to claim that no computational reasoning system can, at present, achieve this sort of feat without significant human assistance.

1.1 Automating the Proof of G1

Prior work devoted to producing computational systems able to prove **G1** have yielded systems able to prove this theorem only when the distance between this result and the starting point is quite small. This for example holds for the first (and certainly seminal) foray; i.e., for [Quaife, 1988], as explained in [Bringsjord, 1998], where it’s shown that the proof of **G1**, because the set of premises includes an ingenious human-devised encoding scheme, is very easy—to the point of being at the level of proofs requested from students in introductory mathematical logic classes.

Likewise, [Arnon, 1993] is an exact parallel of the human-devised proof given by [Kleene, 1996]. Finally, in much more recent and truly impressive work by [Sieg and Field, 2005], there is a move to natural-deduction formats, which we applaud—but the machine essentially begins its processing at a point exceedingly close to where it needs to end up. As Sieg and Field concede: “As axioms we take for granted the representability and derivability conditions for the central syntactic notions as well as the diagonal lemma for constructing self-referential sentences.” If one takes for granted such things, finding a proof of **G1** is effortless for a computing machine.[2] In sum, while a lot of commendable work has been done to build the foundation for our prospective work, the daunting formal and engineering challenge of producing a computational system able to produce **G1** without clever seeding from a human remains entirely unmet.

2 The Analogico-Deductive Approach

2.1 Conjecture Generation

The problem with the purely deductive method is simply that it does not allow us to come close to the type of model-based reasoning that great thinkers are known to have used. Gödel himself has been described as having a “line of thought [which] seems to move from conjecture to conjecture” [Wang, 1995]. Reasoners in general are known to conjecture through analogy when a straightforward answer

[1] **G1** of course applies to any axiom system meeting the standard conditions (Turing-decidability, representability, consistency), but we tend to refer to **PA** for economization.

[2] A video demonstration of the small-distance process can be found at http://kryten.mm.rpi.edu/GodelI_abstract_in_Slate.mov

Licato, J.; Govindarajulu, N.; Bringsjord, S.; Pomeranz, M.; Gittelson, L. 2013. Analogico-Deductive Generation of Godel’s First Incompleteness Theorem from the Liar Paradox. In Proceedings of IJCAI 2013. Pdf

1.1. Great Computational Intelligence (GCI) & Hierarchies …

Familiar Hierarchies (as context for characterization of GCI Itself)

Analytical Hierarchy

Arithmetical Hierarchy

$\vdots$
Π_2
Σ_2
Π_1
Σ_1
Σ_0

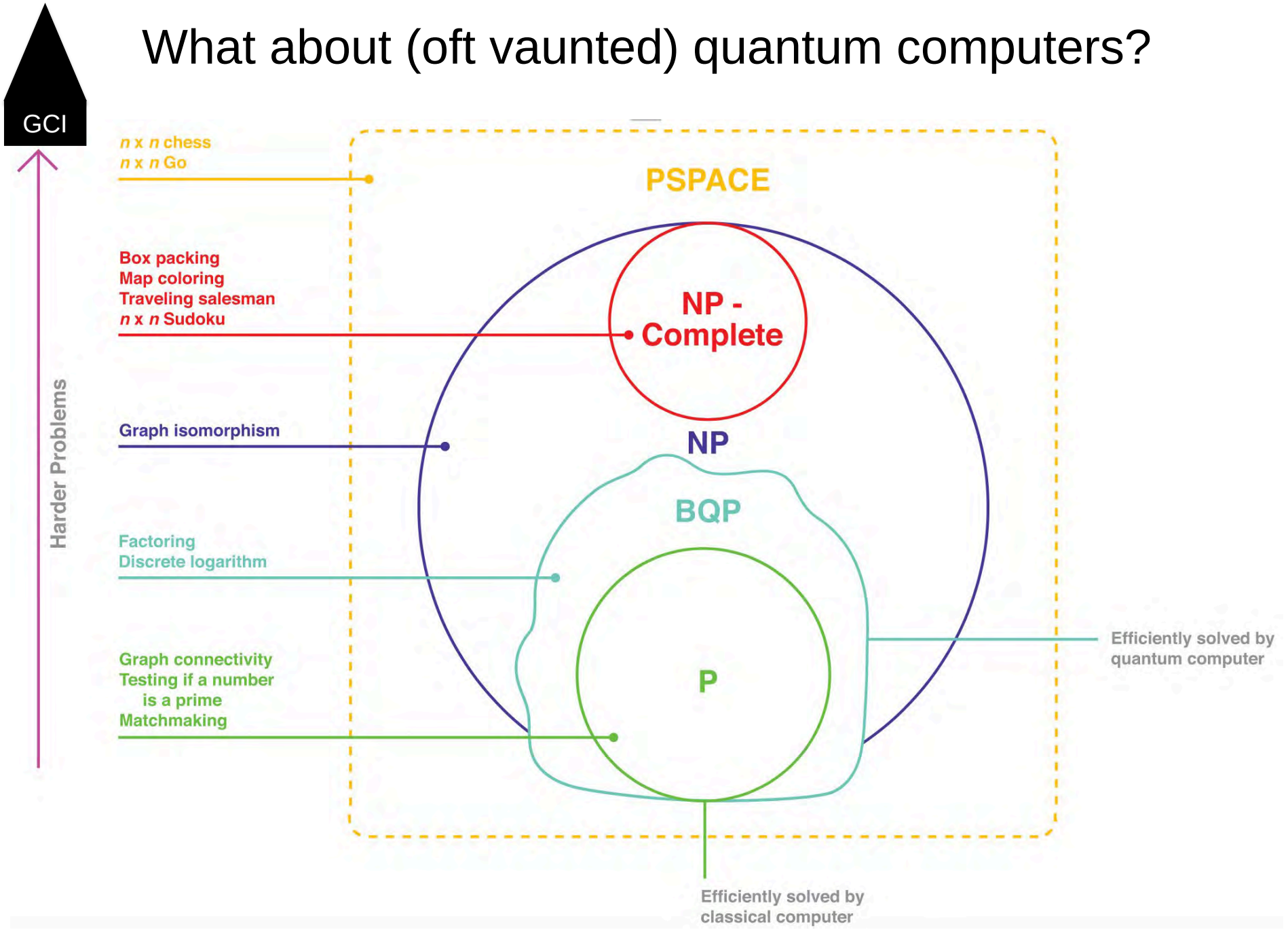
"We need to surmount **G1**!"

Polynomial Hierarchy

(If "An Argument for **P**=**NP**" in (Bringsjord 2017) is sound, PH completely collapses.)

$$\mathbf{P} \subseteq \mathbf{NP} \subseteq \mathbf{PSPACE} = \mathbf{NPSPACE} \subseteq \mathbf{EXPTIME} \subseteq \mathbf{NEXPTIME} \subseteq \mathbf{EXPSPACE}$$

What about (oft vaunted) quantum computers?

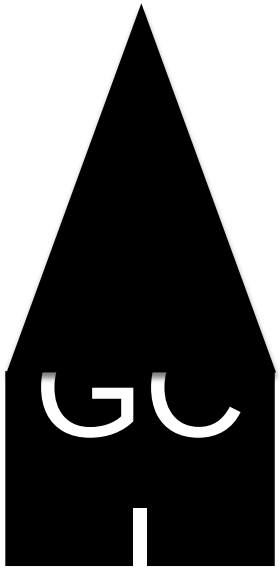


LOGIC MACHINES AND DIAGRAMS

Martin Gardner

McGRAW-HILL BOOK COMPANY, INC.
New York Toronto London 1958

Logic-Machines Hierarchy



$$\vdots$$

$$\mathbb{P} \quad \mathfrak{q} \quad \mathscr{L}_2 \quad \mathbb{R} \quad \mathbb{C}$$

Turing Machines $\mathbb{P} \quad \mathfrak{q} \quad \mathscr{L}_1 \quad \mathbb{R} \quad \mathbb{C}$

PH $\mathbb{P} \quad \mathfrak{q} \quad \mathscr{L}_0 \quad \mathbb{R} \quad \mathbb{C}$

$$\mathbb{P} \quad \mathfrak{q} \quad \mathscr{L}_{\mathrm{PC}} \quad \mathbb{R} \quad \mathbb{C}$$

program

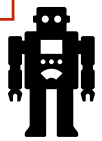
query

logical system
(prop calc here)

automated reasoner

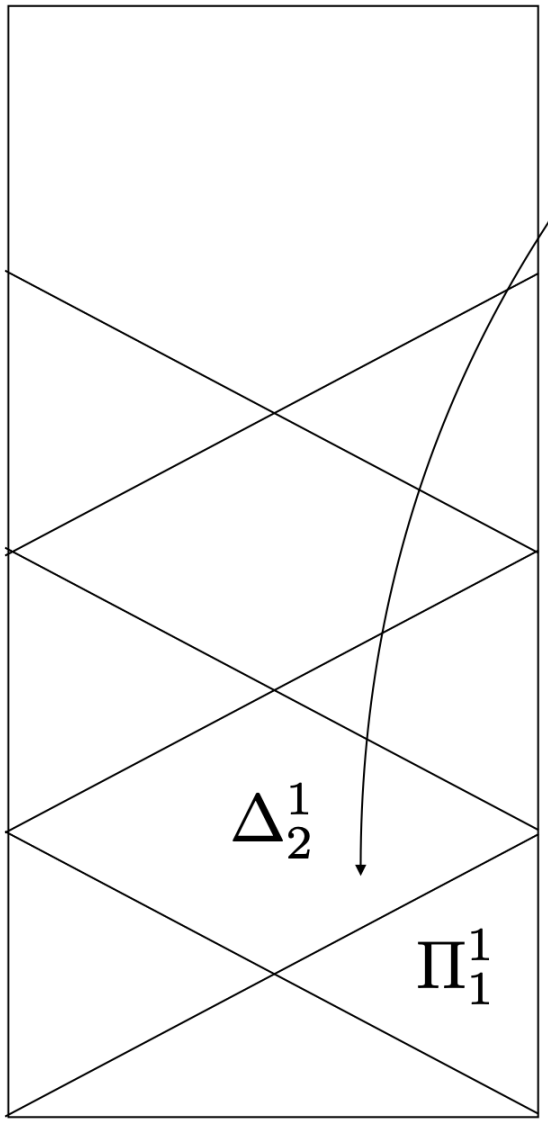
proof/argument checker

AI, as a science, needs to say more about where AI falls/can fall in the landscape.



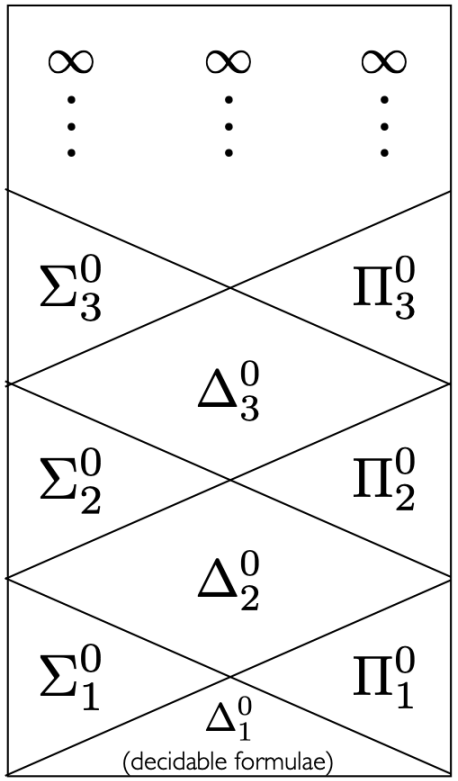
$\mathfrak{LM}$

$\mathcal{AH}$ (Analytic Hierarchy)



Infinite Time Turing Machines (ITTMs)

$\mathcal{A}^r\mathcal{H}$ (Arithmetic Hierarchy)



Human Persons (according to Bringsjord)

Human Brains (according to Granger)

$\mathcal{CH}$ (Chomsky Hierarchy)

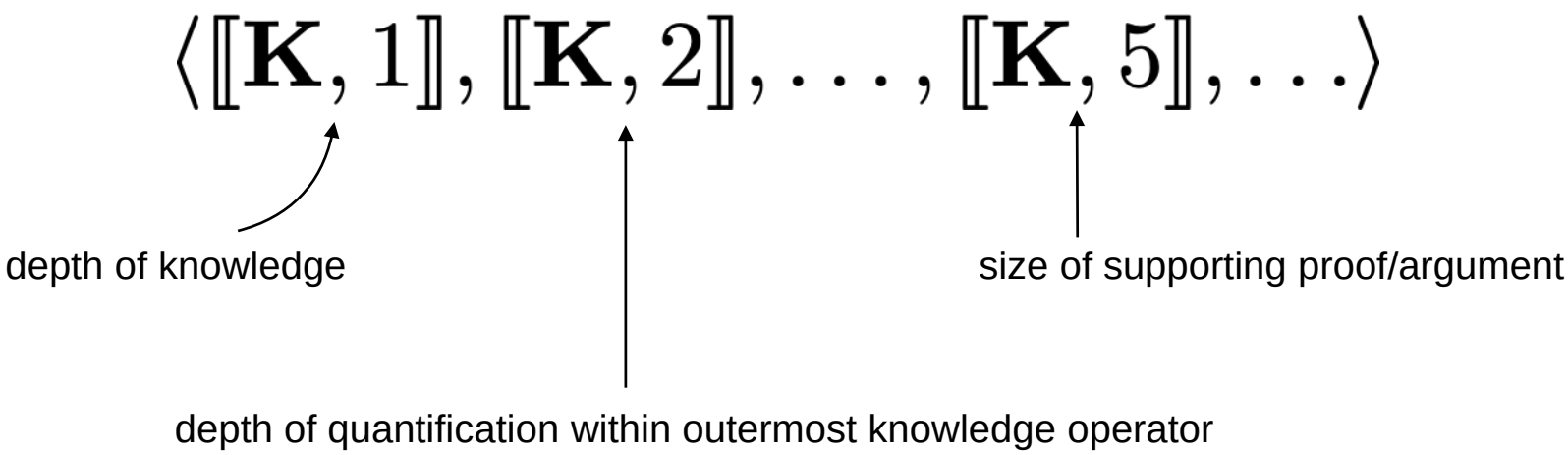
Turing Machines (TMs)
Linear Bounded Automata (LBAs)
Push Down Automata (PDAs)
Finite State Automata (FSAs)

1.2. Measuring Great Computational Intelligence (GCI): & “Cognitive Consciousness” …

Λ vs Φ

Basic Idea, Intuitively Put

The level of (cognitive) intelligence/consciousness of an AI at a time is a list of tuples (= matrix) giving eg the size of logical depth of (at least) five measures for each cognitive operator (i.e. for **K**, **B**, **P**, …).



Journal of Artificial Intelligence and Consciousness | Vol. 07, No. 02, pp. 155-181 (2020)

The Theory of Cognitive Consciousness, and Λ (Lambda)

Extending Measures from $\mathcal{L}^0$ to $\mathcal{L}$

$$\mu_\omega(\phi) = \begin{cases} \mu(\phi) & \text{if } \phi \in \mathcal{L}^0 \\ \max_\psi \mu_\omega(\psi) + 1 & \text{if } \phi \equiv \omega_i[a_1, t_1, \ldots \psi \ldots] \end{cases}$$

For example, let μ count the number of predicate symbols in a formula.

Example

$$\mu(Happy(john)) = 1$$
$$\mu_\omega(Happy(john)) = 1$$
$$\mu_\omega\Big(\mathbf{B}(mary, t_2, Happy(john))\Big) = 2$$

For any agent a, we want to look at the new complexity the agent introduces that is above any input complexity. For this, we introduce $\Delta : 2^{\mathcal{L}} \times 2^{\mathcal{L}} \to 2^{\mathcal{L}}$ operator that computes differences between two sets of formulae. This can be simply the set-difference operator. For convenience, let $\omega_j[\Gamma]$ denote the subset of formulae with operators ω_j in Γ:

$$\omega_j[\Gamma] = \{\phi \mid \phi \equiv \omega_j[\ldots] \text{ and } \phi \in \Gamma \text{ or } \phi \text{ a subformula } \in \Gamma\}$$

Given a set of measures $\{\mu^0, \ldots, \mu^N\}$ and a set of modal (or cognitive) operators $\{\omega_0, \ldots, \omega_M\}$, we define Λ as a function mapping an agent at a time point to a matrix $\mathbb{N}^{M \times N}$:

$$\Lambda : A \times T \to \mathbb{N}^{M \times N}$$

Definition of Λ

$$\Lambda(a,t)_{i,j} = \max_\phi \left\{ \mu^i(\phi) \mid \phi \in \Delta\Big(\omega_j[o(a,t)], \omega_j[i(a,t)]\Big) \right\}$$

Example 2

Let us consider two modal operators $\{\mathbf{B}, \mathbf{D}\}$ and the following base measures μ^0 which measures quantificational complexity via Σ or Π measures, μ^1 which counts the total number of predicate symbols (not a count of unique predicate symbols), and μ^2 which counts the number of distinct time expressions. This gives $\Lambda : A \times T \to \mathbb{N}^{2\times 3}$. At some timepoint t, let an agent a have the following $\Delta(o(a,t), i(a,t)) = \{\mathbf{B}(\phi_1), \mathbf{D}(\phi_2)\}$

$\phi_1 \equiv \neg\forall a : Happy(a,t)$; $\phi_2 \equiv \forall b : \neg Hungry(b,t) \to Happy(b,t)$

Applying the measures:

$$\mu^u(\phi_1) = 1, \mu^1(\phi_1) = 1; \mu^2(\phi_1) = 1$$
$$\mu^u(\phi_2) = 1; \mu^1(\phi_2) = 2; \mu^2(\phi_2) = 1$$

Giving us:

$$\Lambda(a,t) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

6.1. *Some Distinctive Properties of* Λ *(vs. Φ)*

Here are some properties of the Λ framework of potential interest to our readers:

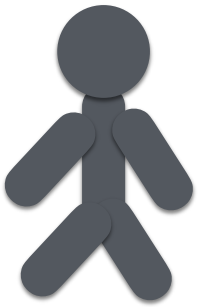
Non-Binary Whereas Φ is such that an agent either is or is not (P-) conscious, cognitive consciousness as measured by Λ admits of a fine-grained range of the *degree* of cognitive consciousness.

Zero Λ for Some Animals and Machines Animals such as insects, and computing machines that are end-to-end statistical/connectionist "ML," have zero Λ, and hence cannot be cognitively conscious. In contrast, as emphasized to Bringsjord in personal conversation,[6] Φ says that even lower animals are conscious.

Human-Nonhuman Discontinuity Explained by Λ From the computational/AI point of view, cognitive scientists have taken note of a severe discontinuity between *H. sapiens sapiens* and other biological creatures on Earth [Penn *et al.*, 2008], and the sudden and large jump in level of Λ from (say) chimpanzees and dolphins to humans is in line with this observation. It's for instance doubtful that any nonhuman animals are capable of reaching third-order belief; hence $\Lambda[\mathbf{B}, 0] = n$, where $n \geq 3$, for any nonhuman animal, is impossible. In stark contrast, each of us believes that you, the reader, believe that we believe that San Francisco is located in California.

Human-Human Discontinuity Explained by Λ A given neurobiologically normal human, over the course of his or her lifetime, has very different cognitive capacity. E.g., it's well-known that such a human, before the age of four or five, is highly unlikely to be able to solve what has become known as the *false-belief task* (or sometimes the *sally-anne task*), which we denote by 'FBT.' From the point of view of Λ, the explanation is simply that an agent with insufficiently high cognitive consciousness is incapable of solving such a task; specifically, solving FBT requires an agent to have

[6] With Tononi and C. Koch, SRI T&C Series.



Formal Conditions for $\mathcal{DDE}$

$\mathbf{F_1}$ α carried out at t is not forbidden. That is:

$$\Gamma \not\vdash \neg \mathbf{O}\Big(a, t, \sigma, \neg happens\big(action(a,\alpha),t\big)\Big)$$

$\mathbf{F_2}$ The net utility is greater than a given positive real γ:

$$\Gamma \vdash \sum_{y=t+1}^{H} \left(\sum_{f \in \alpha_I^{a,t}} \mu(f,y) - \sum_{f \in \alpha_T^{a,t}} \mu(f,y) \right) > \gamma$$

$\mathbf{F_{3a}}$ The agent a intends at least one good effect. ($\mathbf{F_2}$ should still hold after removing all other good effects.) There is at least one fluent f_g in $\alpha_I^{a,t}$ with $\mu(f_g,y) > 0$, or f_b in $\alpha_T^{a,t}$ with $\mu(f_b,y) < 0$, and some y with $t < y \leq H$ such that the following holds:

$$\Gamma \vdash \begin{pmatrix} \exists f_g \in \alpha_I^{a,t}\ \mathbf{I}\Big(a,t,Holds\big(f_g,y\big)\Big) \\ \vee \\ \exists f_b \in \alpha_T^{a,t}\ \mathbf{I}\Big(a,t,\neg Holds\big(f_b,y\big)\Big) \end{pmatrix}$$

$\mathbf{F_{3b}}$ The agent a does not intend any bad effect. For all fluents f_b in $\alpha_I^{a,t}$ with $\mu(f_b,y) < 0$, or f_g in $\alpha_T^{a,t}$ with $\mu(f_g,y) > 0$, and for all y such that $t < y \leq H$ the following holds:

$$\Gamma \not\vdash \mathbf{I}\Big(a,t,Holds\big(f_b,y\big)\Big) \text{ and}$$
$$\Gamma \not\vdash \mathbf{I}\Big(a,t,\neg Holds\big(f_g,y\big)\Big)$$

$\mathbf{F_4}$ The harmful effects don't cause the good effects. Four permutations, paralleling the definition of $\rhd$ above, hold here. One such permutation is shown below. For any bad fluent f_b holding at t_1, and any good fluent f_g holding at some t_2, such that $t < t_1, t_2 \leq H$, the following holds:

$$\Gamma \vdash \neg \rhd \Big(Holds\big(f_b,t_1\big), Holds\big(f_g,t_2\big)\Big)$$

Example from Sim in IJCAI Paper

looking at one single chunk

$$\left\{\begin{array}{l}\mathbf{K}\left(I, now, \sigma_{trolley}\right), \\ \mathbf{B}\left(I, now, \mathbf{O}\left(\begin{array}{l} I, now, \sigma_{trolley}, \\ \left[\begin{array}{c}\neg\exists t : \text{Moment}\, Holds\left(dead\left(P_1, t\right)\right) \\ \wedge \\ \neg\exists t : \text{Moment}\, Holds\left(dead\left(P_2, t\right)\right)\end{array}\right]\end{array}\right)\right), \\ \mathbf{O}\left(I, now, \sigma_{trolley}, \left[\begin{array}{l}\neg\exists t : \text{Moment}\, Holds\left(dead\left(P_1, t\right)\right) \wedge \\ \neg\exists t : \text{Moment}\, Holds\left(dead\left(P_2, t\right)\right)\end{array}\right]\right)\end{array}\right\}$$

$$\vdash \mathbf{I}\left(I, now, \left[\begin{array}{l}\neg\exists t : \text{Moment}\, Holds\left(dead\left(P_1, t\right)\right) \wedge \\ \neg\exists t : \text{Moment}\, Holds\left(dead\left(P_2, t\right)\right)\end{array}\right]\right)$$

$\Lambda[\mathbf{B}, 1] = 2$

$\Lambda[\mathbf{B}, 2] = 1$

$\Lambda[\mathbf{K}, 1] = 1$

$\Lambda[\mathbf{O}, 1] = 1$

$\Lambda[\mathbf{O}, 1] = 1$

$\Lambda[\mathbf{I}, 1] = 1$

$\Lambda[\mathbf{I}, 2] = 1$

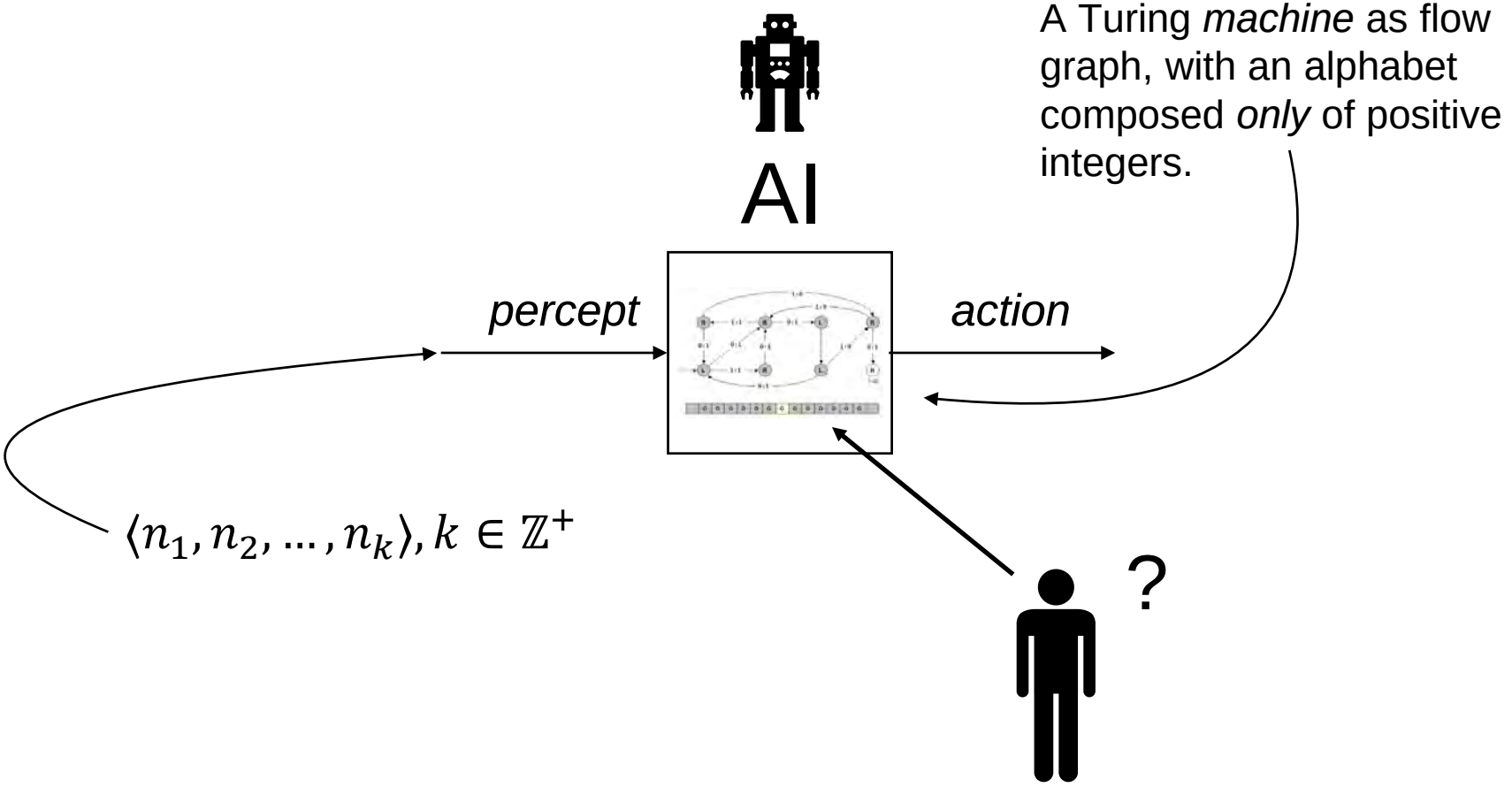
$\Lambda[\mathbf{B}, 3] = 1$

$\Lambda[\mathbf{B}, 4] = \infty$

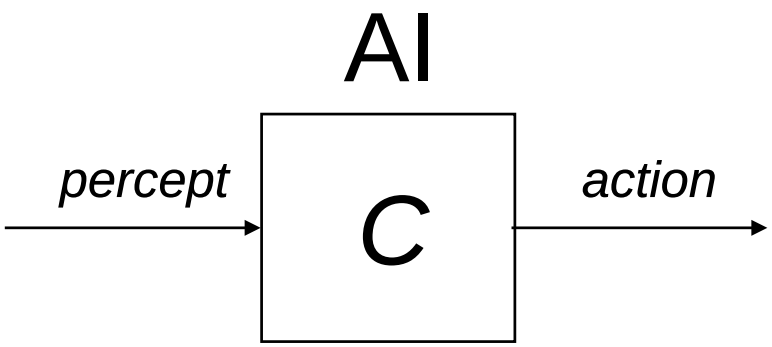
$\vdots$

The application of Λ to eg “Deep Learning” machines implies that they have zero cognitive intelligence/cognitive consciousness.

AI:MLn


AI
A Turing *machine* as flow graph, with an alphabet composed *only* of positive integers.
percept
action
$\langle n_1, n_2, \ldots, n_k \rangle, k \in \mathbb{Z}^+$
?

We will be able to measure the intelligence of *any* AI, not with *g*-loaded tests of intelligence, but with Λ-loaded tests of machine intelligence, in keeping with Psychometric AI.



CA: 11 Axioms (Initially)

Plan

P2B

$\mu\mathcal{DCEC}^*_3$ K2B $\forall a[\mathbf{K}_a\phi \rightarrow (\mathbf{B}_a\phi \wedge \mathbf{B}_a\exists\Phi\exists\alpha(\Phi \rightsquigarrow_{\alpha/\pi} \phi)]$

Intro

Incorr $\forall a\forall t\forall F[(F is\ contingent\ \wedge F \in C'') \rightarrow (\Box\mathbf{B}(a,t,Fa) \rightarrow Fa)]$

Ess

¬CompE

Irr

Free

$[A_1]$ $\mathbf{C}(\forall\, f,t\ .\ initially(f) \wedge \neg clipped(0,f,t) \Rightarrow holds(f,t))$

$[A_2]$ $\mathbf{C}(\forall\, e,f,t_1,t_2\ .\ happens(e,t_1) \wedge initiates(e,f,t_1) \wedge t_1 < t_2 \wedge \neg clipped(t_1,f,t_2) \Rightarrow holds(f,t_2))$

$[A_3]$ $\mathbf{C}(\forall\, t_1,f,t_2\ .\ clipped(t_1,f,t_2) \Leftrightarrow [\exists\, e,t\ .\ happens(e,t) \wedge t_1 < t < t_2 \wedge terminates(e,f,t)])$

$[A_4]$ $\mathbf{C}(\forall\, a,d,t\ .\ happens(action(a,d),t) \Rightarrow \mathbf{K}(a,happens(action(a,d),t)))$

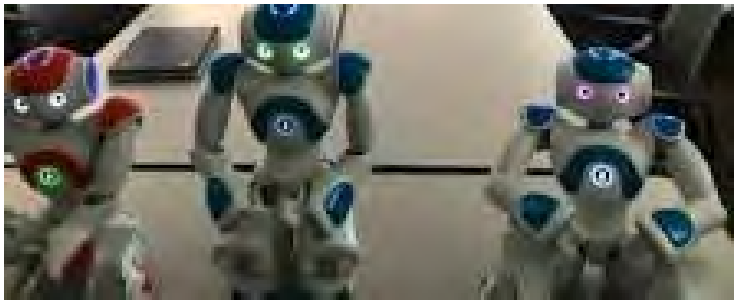
$[A_5]$ $\mathbf{C}(\forall\, a,f,t,t'\ .\ \mathbf{B}(a,holds(f,t)) \wedge \mathbf{B}(a,t < t') \wedge \neg\mathbf{B}(a,clipped(t,f,t')) \Rightarrow \mathbf{B}(a,holds(f,t')))$

$\mathbf{C}$ SpecRel

CCaus $\mathbf{C}$ $\mathcal{EC}$

Thel

What is the level of consciousness ($= \Lambda$ value) enjoyed by this self-conscious robot?



https://motherboard.vice.com/en_us/article/mgbyvb/watch-these-cute-robots-struggle-to-become-self-aware

1.3. Can GCI Formalisms be (Machine) Learned?

...

Inductive Calculus

Requirements

- **Requirement 1**: Learn reasoning schemes from a minimal number of examples?

Number of examples required should be bounded by the complexity of the inference scheme.

$$N := \text{number of examples}$$
$$I := \text{inference scheme}$$
$$s := \text{complexity of the inference scheme}$$
$$f := \text{some non-decreasing function from}$$

$$N(I) \leq f(S(I))$$

Inductive Calculus

Requirements

- **Requirement 2**: Learnt information should be inspectable and modifiable manually.

Ideally, learnt information should be encoded as formulae in some logic.

Inductive Calculus

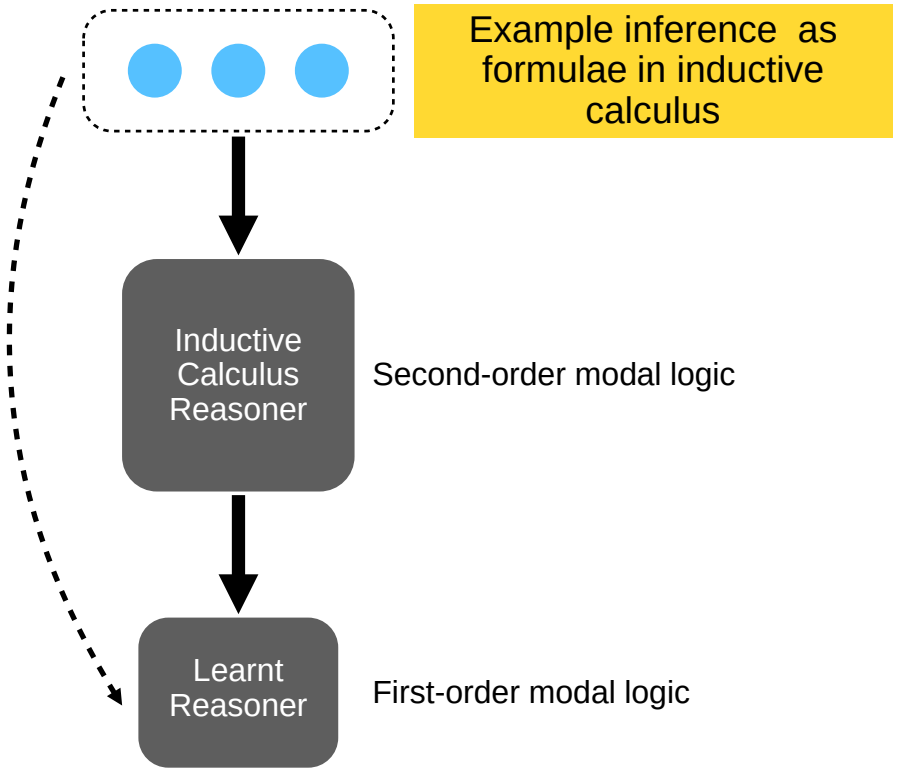
Solution

- Use shadowing and higher-order logic to solve these challenges in the **inductive calculus**

Govindarajulu, Naveen & Bringsjord, Selmer & Peveler, Matthew. (2019). “On Quantified Modal Theorem Proving for Modeling Ethics.” *Electronic Proceedings in Theoretical Computer Science*. 311. 43-49. 10.4204/EPTCS.311.7.

Inductive Calculus

Overview



"Inductive Calculus: A Higher-Order Quantified Modal Logic For Learning Inference Schemata"

DOI: 10.5281/zenodo.3903803

Inductive Calculus

Syntax

$$\phi ::= \begin{cases} P \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \forall x : \phi(x) \\ \mathbf{E}(\phi, \psi) \end{cases}$$

Inductive Calculus

Syntax

$$\frac{\phi_1, \ldots, \phi_n}{\psi}$$

$$\mathbf{E}(\phi_1 \wedge \ldots \wedge \phi_n, \psi)$$

Inductive Calculus

Inference

$$\frac{\mathbf{E}(\phi_1, \ldots, \phi_n, \psi)}{\forall^2 \vec{\nu} : g(\phi_1, \vec{\nu}) \wedge \ldots \wedge g(\phi_n, \vec{\nu}) \rightarrow g(\psi, \vec{\nu})}$$

$g(\phi, \vec{\nu})$ Generalization operator

Generalization operator from:
Govindarajulu, N.S., Bringsjord, S., Ghosh, R. & Sarathy, V. (2019) “Toward the Engineering of Virtuous Machines." In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 29–35.

Inductive Calculus

Example 1: Learning Infinitary Common Knowledge

$$\frac{\mathbf{C}(\phi)}{\mathbf{K}(a_1, \mathbf{K}(a_2, \ldots, \mathbf{K}(a_n, \phi)))}$$

```
{:name        "IC4.3"
 :description
 "Learning infinitary common knowledge.
  R3 (without restriction on iteration) in http://kryten.mm.rpi.edu/CognitiveCalculus092808.pdf"

 :assumptions {:example1 (e=> (Common! t1 P)
                              (and (Knows! a t1 P)
                                   (Common! t1 (Knows! a t1 P))))

               :input (Common! now Q)
               }
 :al          (Knows! a now (Knows! b now (Knows! c now Q)))}
```

Example inference

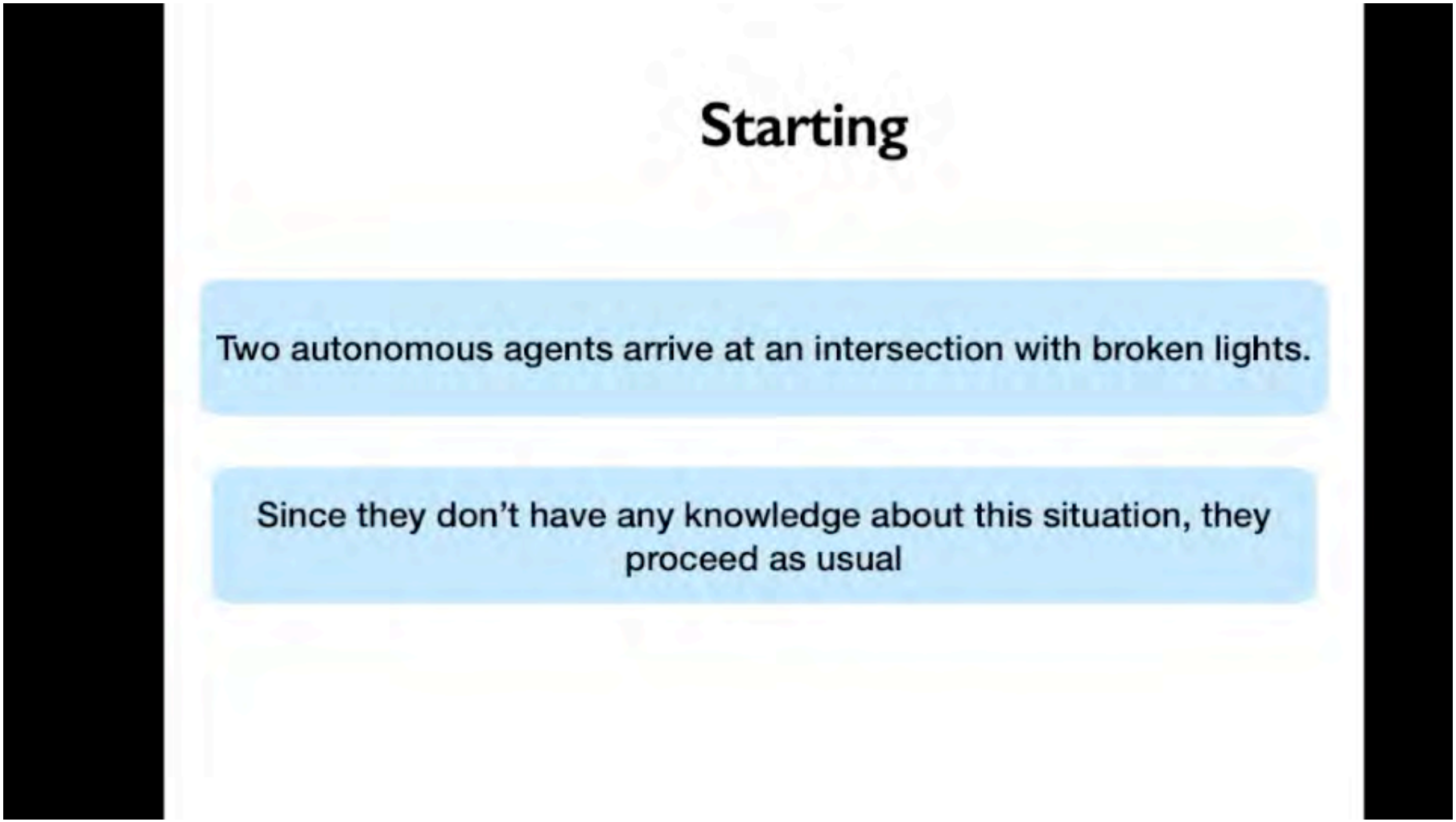
Input

Output

https://github.com/naveensundarg/prover

Inductive Calculus

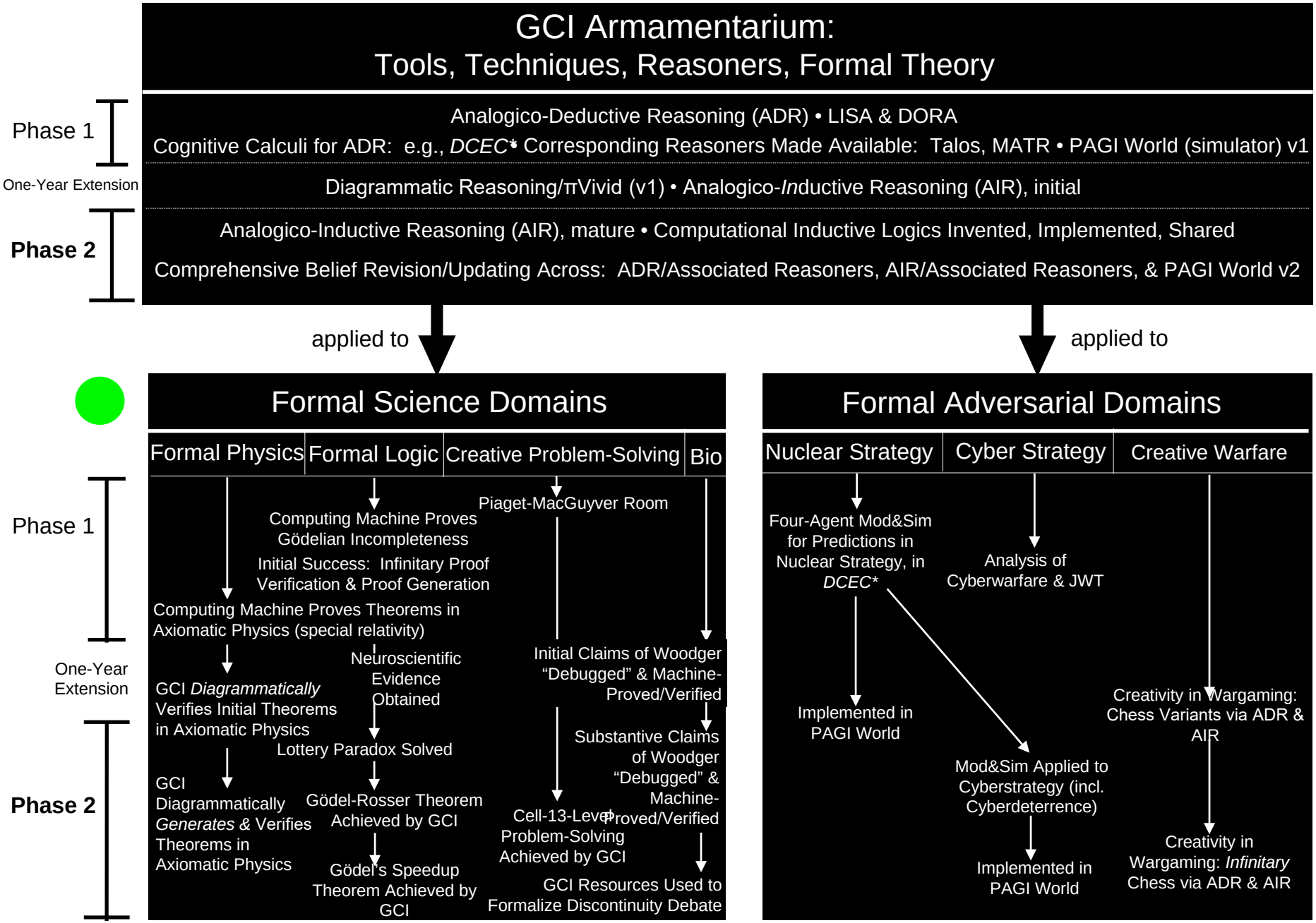
Demo



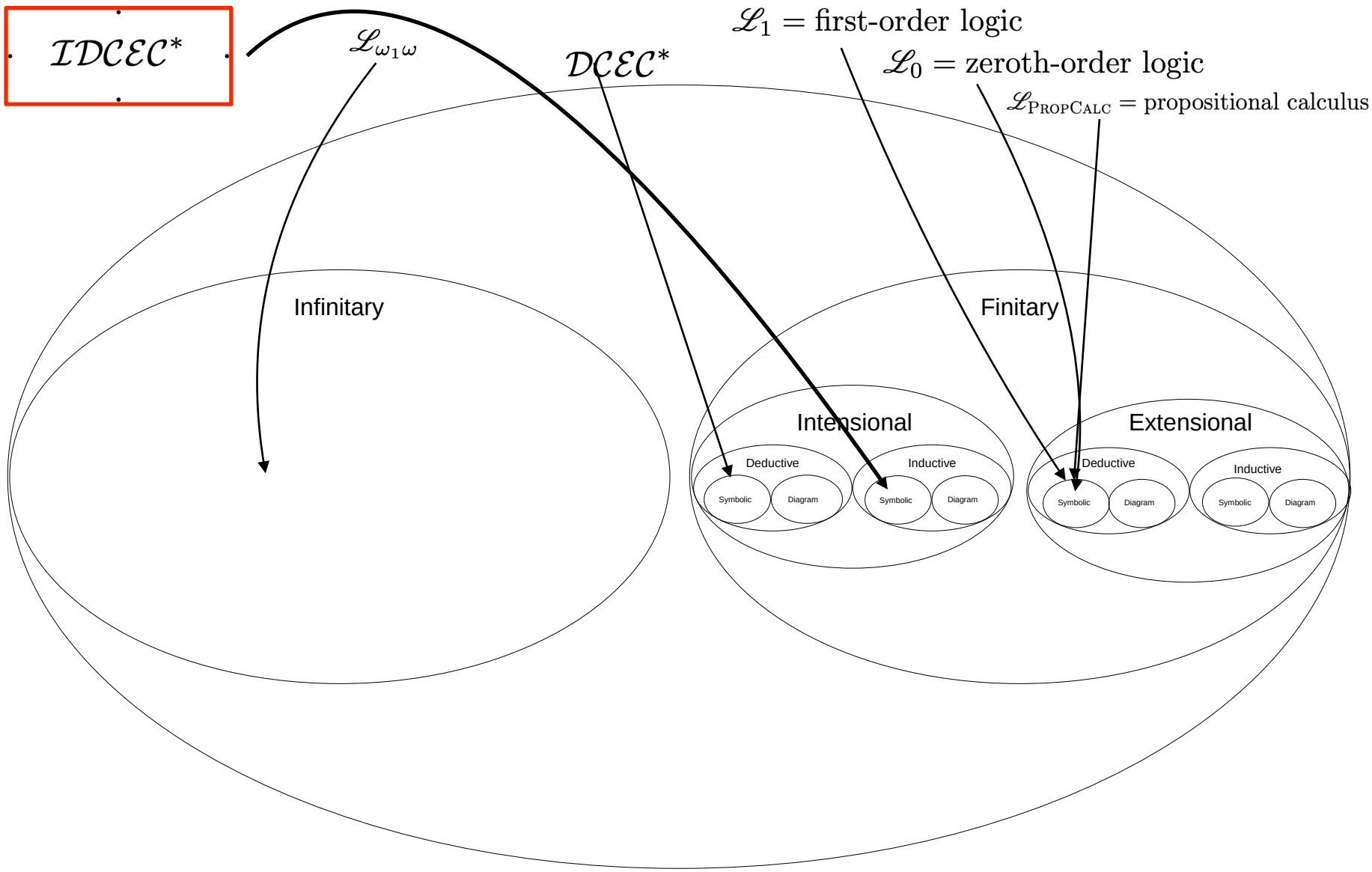
2. Gödelian “Insoluble” Time-Travel Paradox Solved …

3. Transition from ADR to Automated *Inductive* Reasoning

Great Computational Intelligence (GCI), Mature & Further Applied



The Universe of Logics



Licatoan Inductive Inference Schemata (e.g.)

Inference Schemata for $\mathcal{S}$

$$\frac{\mathbf{P}(a, t_1, \phi_1), \;\; \Gamma \vdash t_1 < t_2}{\mathbf{B}^5(a, t_2, \phi)} \; [R^s_{\mathbf{P}}]$$

$$\frac{\mathbf{B}^{s_1}(a, t_1, \phi_1), \ldots, \mathbf{B}^{s_m}(a, t_m, \phi_m), \{\phi_1, \ldots, \phi_m\} \vdash \phi, \Gamma \vdash t_i < t}{\mathbf{B}^{min(s_1, \ldots, s_m)}(a, t, \phi)} \; [R^s_{\mathbf{B}}]$$

with $max(\{s_1, \ldots, s_m\}) - min(\{s_1, \ldots, s_m\}) \leq u$

Licato, Boger & Zhang (2018)
Developed method for identifying informal arguments (of ad hominem form) from Reddit comments, to address inter-annotator agreement problem

SOURCE		TARGET
$\mathcal{P}$	$\longrightarrow$	$\mathcal{P}^*$
$\mathcal{A}$	$\not\longrightarrow$	$\neg\mathcal{A}^*$
$\neg\mathcal{B}$	$\not\longrightarrow$	$\mathcal{B}^*$
proposals		*proposals*
χ		
		χ^*

$$\frac{\mathbf{B}^x(\varphi_S), \mathbf{B}^y(M_{ST}), \mathbf{B}^w((M_{ST} \wedge \varphi_S) \rightarrow \varphi_T)}{\mathbf{B}^{\min(w,x,y)}(\varphi_T)} f$$

$$\frac{\mathbf{B}^x(\varphi_S), \mathbf{B}^y(\neg\varphi_T), \mathbf{B}^w((M_{ST} \wedge \varphi_S) \rightarrow \varphi_T)}{\mathbf{B}^{\min(w,x,y)}(\neg M_{ST})} b1$$

- Data on schema use
- Argument schemas that are automated-reasoner-friendly
- Automated reasoning tool capable of working with formal schemas, unstructured text, and everything in between

“MATR 2.0” - Some details were discussed in Licato’s presentation last year, & are at any rate available from him.

For Miracle on Hudson:

$\mathcal{IDCEC}^*$

A.2 Implementation

The scenario we presented was simulated in Shadow-Prover [11], ShadowProver is a quantified multi-modal prover capable of handling reasoning in cognitive calculi [12]. The inputs to ShadowProver are shown in Figure 4. Time taken to simulate the adjudicator's reasoning about other agents' beliefs is shown in Table 4.

Table 4. Time Taken For Reasoning

Agent	Time taken (s)
hdrone	3.52
Idrone$_1$	3.82
radar	2.88
Idrone$_2$	2.87

(a) *The adjudicator reasoning about hdrone.*

(b) *The adjudicator reasoning about the faulty Idrone.*

(c) *The adjudicator reasoning about information from the radar.*

(d) *The adjudicator reasoning about the fixed Idrone.*

Fig. 4. Inputs to ShadowProver to Model the Scenario

A.3 Definitions for $\mathcal{DCEC}$ and $\mathcal{IDCEC}$

Below is the signature of the standard $\mathcal{DCEC}$. It contains the sorts, function signatures, and grammar of this cognitive calculus.

$\mathcal{DCEC}$ Signature

$$
\begin{aligned}
&S \supseteq \mathsf{Agent} \mid \mathsf{ActionType} \mid \mathsf{Action} \subseteq \mathsf{Event} \mid \mathsf{Moment} \mid \mathsf{Fluent}\\
&f \coloneqq \begin{cases}
action : \mathsf{Agent} \times \mathsf{ActionType} \to \mathsf{Action}\\
initially : \mathsf{Fluent} \to \mathsf{Formula}\\
holds : \mathsf{Fluent} \times \mathsf{Moment} \to \mathsf{Formula}\\
happens : \mathsf{Event} \times \mathsf{Moment} \to \mathsf{Formula}\\
clipped : \mathsf{Moment} \times \mathsf{Fluent} \times \mathsf{Moment} \to \mathsf{Formula}\\
initiates : \mathsf{Event} \times \mathsf{Fluent} \times \mathsf{Moment} \to \mathsf{Formula}\\
terminates : \mathsf{Event} \times \mathsf{Fluent} \times \mathsf{Moment} \to \mathsf{Formula}\\
prior \mid \mathsf{Moment} \times \mathsf{Moment} \to \mathsf{Formula}
\end{cases}\\
&t \coloneqq x : S \mid c : S \mid f(t_1, \ldots, t_n)\\
&\phi \coloneqq \begin{cases}
q : \mathsf{Formula} \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \forall x : \phi(x) \mid \exists x : \phi(x)\\
\mathbf{P}(a,t,\phi) \mid \mathbf{K}(a,t,\phi) \mid \mathbf{S}(a,b,t,\phi) \mid \mathbf{S}(a,t,\phi)\\
\mathbf{C}(t,\phi) \mid \mathbf{B}(a,t,\phi) \mid \mathbf{D}(a,t,\phi) \mid \mathbf{I}(a,t,\phi)\\
\mathbf{O}(a,t,\phi,(\neg)happens(action(a^*,\alpha),t'))
\end{cases}
\end{aligned}
$$

Perceives, Knows, Says, Common-knowledge
Believes, Desires, Intends, Ought-to

Next is the standard set of inference schemata for $\mathcal{DCEC}$.

$\mathcal{DCEC}$ Inference Schemata

$$
\frac{\mathbf{K}(a,t_1,\Gamma),\ \Gamma \vdash \phi,\ \ t_1 \geq t_2}{\mathbf{K}(a,t_2,\phi)}\ [I_K] \qquad \frac{\mathbf{B}(a,t_1,\Gamma),\ \Gamma \vdash \phi,\ \ t_1 \leq t_2}{\mathbf{B}(a,t_2,\phi)}\ [I_B]
$$

$$
\frac{}{\mathbf{C}(t,\mathbf{P}(a,\tau,\phi) \rightarrow \mathbf{K}(a,t,\phi))}\ [I_1] \qquad \frac{}{\mathbf{C}(t,\mathbf{K}(a,t,\phi) \rightarrow \mathbf{B}(a,t,\phi))}\ [I_2]
$$

$$
\frac{\mathbf{C}(t,\phi),\ t \leq t_1,\ \ldots,\ t \leq t_n}{\mathbf{K}(a_1,t_1,\ldots,\mathbf{K}(a_n,t_n,\phi)\ldots)}\ [I_3] \qquad \frac{\mathbf{K}(a,t,\phi)}{\phi}\ [I_4]
$$

$$
\frac{t_1 \leq t_2 \leq t_3}{\mathbf{C}(t,\mathbf{K}(a,t_1,\phi_1 \rightarrow \phi_2)) \rightarrow \mathbf{K}(a,t_2,\phi_1) \rightarrow \mathbf{K}(a,t_3,\phi_2)}\ [I_5]
$$

$$
\frac{t_1 \leq t_2 \geq t_3}{\mathbf{C}(t,\mathbf{B}(a,t_1,\phi_1 \rightarrow \phi_2)) \rightarrow \mathbf{B}(a,t_2,\phi_1) \rightarrow \mathbf{B}(a,t_3,\phi_2)}\ [I_7]
$$

$$
\frac{t_1 \not\geq t_2 \leq t_3}{\mathbf{C}(t,\mathbf{C}(t_1,\phi_1 \rightarrow \phi_2)) \rightarrow \mathbf{C}(t_2,\phi_1) \rightarrow \mathbf{C}(t_3,\phi_2)}\ [I_7]
$$

$$
\frac{}{\mathbf{C}(t,\forall x\colon \beta \rightarrow \phi[x \mapsto t])}\ [I_8] \qquad \frac{}{\mathbf{C}(t,\phi_1 \leftrightarrow \phi_2 \rightarrow \neg\phi_1 \leftrightarrow \neg\phi_2)}\ [I_9]
$$

$$
\frac{}{\mathbf{C}(t,[\phi_1 \wedge \ldots \wedge \phi_n \rightarrow \phi] \rightarrow [\phi_1 \rightarrow \ldots \rightarrow \phi_n \rightarrow \phi])}\ [I_{10}]
$$

$$
\frac{\mathbf{D}(a,t,\psi)\ \ \mathbf{B}(a,t,\phi \rightarrow \psi)}{\mathbf{B}(a,t,\phi)}\ [I_{11}] \qquad \frac{\mathbf{B}(a,t,\phi)\ \ \mathbf{B}(a,t,\psi)}{\mathbf{B}(a,t,\phi \wedge \psi)}\ [I_{12}]
$$

$$
\frac{\mathbf{S}(a,b,t,\phi)}{\mathbf{B}(b,t,\mathbf{B}(a,t,\phi))}\ [I_{13}] \qquad \frac{\mathbf{I}(a,t,happens(action(a^*,\alpha),t'))}{\mathbf{P}(a,t,happens(action(a^*,\alpha),t'))}\ [I_{14}]
$$

$$
\frac{\mathbf{B}(a,t,\phi)\ \ \mathbf{B}(a,t,\mathbf{O}(a,t,\phi,\chi))\ \ \mathbf{O}(a,t,\phi,\chi)}{\mathbf{K}(a,t,\mathbf{I}(a,t,\chi))}\ [I_{15}]
$$

Finally, the following two boxes specify the signature and inference schemata for $\mathcal{IDCEC}$, respectively. These specifications enable reasoning about uncertain belief and knowledge. (For additional information beyond what is provided in these specifications regarding strength factors, in connection with the uncertainty system that underlies the system $\mathcal{S}$ used above, see [35].)

Additional Syntax for $\mathcal{IDCEC}$

$$
\begin{aligned}
&\phi \coloneqq \left(\mathbf{B}^{\sigma}(a,t,\phi) \mid \mathbf{K}^{\sigma}(a,t,\phi)\right.\\
&\text{where } \sigma \in [-6,-5,\ldots,5,6]
\end{aligned}
$$

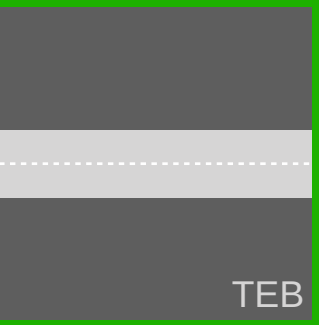
Additional Inference Schemata for $\mathcal{IDCEC}$

$$
\frac{\mathbf{P}(a,t_1,\phi_1),\ \ \Gamma \vdash t_1 < t_2}{\mathbf{B}^1(a,t_2,\phi)}\ [I^1_{\mathbf{P}}]
$$

$$
\frac{\begin{array}{c}\mathbf{B}^{n_1}(a,t_1,\phi_1),\ldots,\mathbf{B}^{n_m}(a,t_m,\phi_m),\\ \{\phi_2,\ldots,\phi_m\} \vdash \phi,\ \{\phi_1,\ldots,\phi_m\} \not\vdash \zeta,\\ \Gamma \vdash t_i < t\end{array}}{\mathbf{B}^{min(n_1,\ldots,n_m)}(a,t,\phi)}\ [I^{\mathbf{B}}_{\mathbf{B}}]
$$

$$
\text{where } \sigma \in [0,1,\ldots,5,6]
$$

$$
\frac{}{\mathbf{C}(t,\mathbf{B}^{-\sigma}(a,t,\phi) \leftrightarrow \mathbf{B}^{\sigma}(a,t,\neg\phi))}\ [I^{-}_{\sigma}]
$$

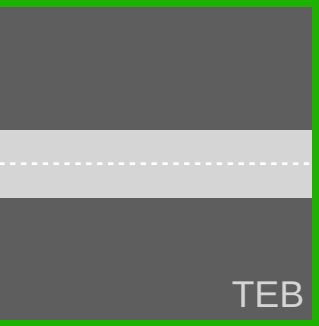

TEB

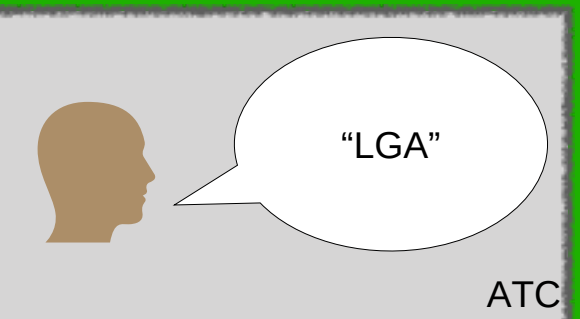

LGA

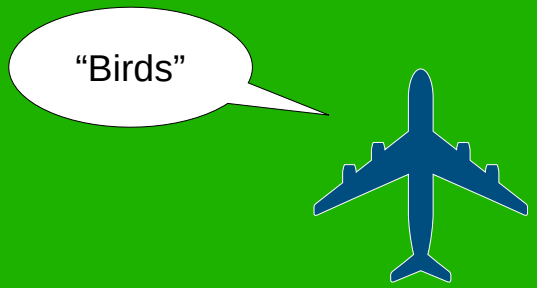
$\mathbf{S}(atc, capt, t_1, \mathrm{Land}(capt, t_1, lga_{13}))$

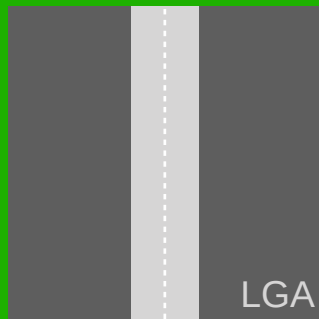
$\therefore \mathbf{B}\big(capt, t_1, \mathbf{B}(atc, t_1, \mathrm{Land}(capt, t_1, lga_{13}))\big)$ $[I_{12}]$ ✓

$\therefore \mathbf{B}^1(capt, t_1, \mathrm{Land}(capt, t_1, lga_{13}))$ $[\mathbf{B}^1\text{-def}]$ ✓









$\therefore \mathrm{Land}(capt, t_2, teb) \succ_{t_2}^{capt} \mathrm{Land}(capt, t_2, lga_{13})$ [$\succ_t^a$ -def] ✓

$\therefore \mathbf{B}^2(capt, t_2, \mathrm{Land}(capt, t_2, teb))$ [$\mathbf{B}^2$-def] ✓

“LGA unreachable”

“TEB”

TEB

“LGA”

ATC

LGA

$\mathbf{S}(atc, capt, t_3, \mathrm{Land}(capt, t_3, teb_1))$

$\therefore \mathbf{B}\big(capt, t_3, \mathbf{B}(atc, t_3, \mathrm{Land}(capt, t_3, teb_1))\big)$ $[I_{12}]$ ✓

$\therefore \mathbf{B}^1(capt, t_3, \mathrm{Land}(capt, t_3, teb_1))$ $[\mathbf{B}^1\text{-def}]$ ✓

$\therefore \mathrm{Land}(capt, t_3, hud) \succ_{t_3}^{capt} \mathrm{Land}(capt, t_3, teb_1)$ $[\succ_t^a\text{-def}]$ ✓

$\therefore \mathbf{B}^2(capt, t_3, \mathrm{Land}(capt, t_3, hud))$ $[\mathbf{B}^2\text{-def}]$ ✓

“TEB unreachable”

“Hudson”

TEB

“Okay, TEB”

ATC

LGA

$$\mathbf{S}(atc, capt, t_3, \mathrm{Land}(capt, t_3, teb_1))$$

$$\therefore \mathbf{B}\big(capt, t_3, \mathbf{B}(atc, t_3, \mathrm{Land}(capt, t_3, teb_1))\big) \quad [I_{12}] \checkmark$$

$$\therefore \mathbf{B}^1(capt, t_3, \mathrm{Land}(capt, t_3, teb_1)) \quad [\mathbf{B}^1\text{-def}] \checkmark$$

$$\therefore \mathrm{Land}(capt, t_3, hud) \succ_{t_3}^{capt} \mathrm{Land}(capt, t_3, teb_1) \quad [\succ_t^a \text{-def}] \checkmark$$

$$\therefore \mathbf{B}^2(capt, t_3, \mathrm{Land}(capt, t_3, hud)) \quad [\mathbf{B}^2\text{-def}] \checkmark$$

TEB

“Okay, TEB”

ATC

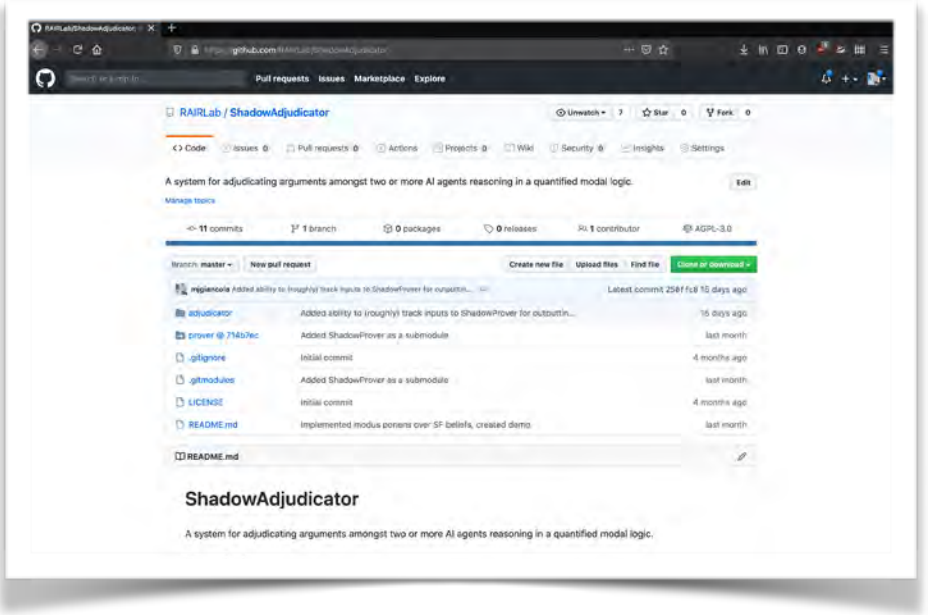
LGA

4Q

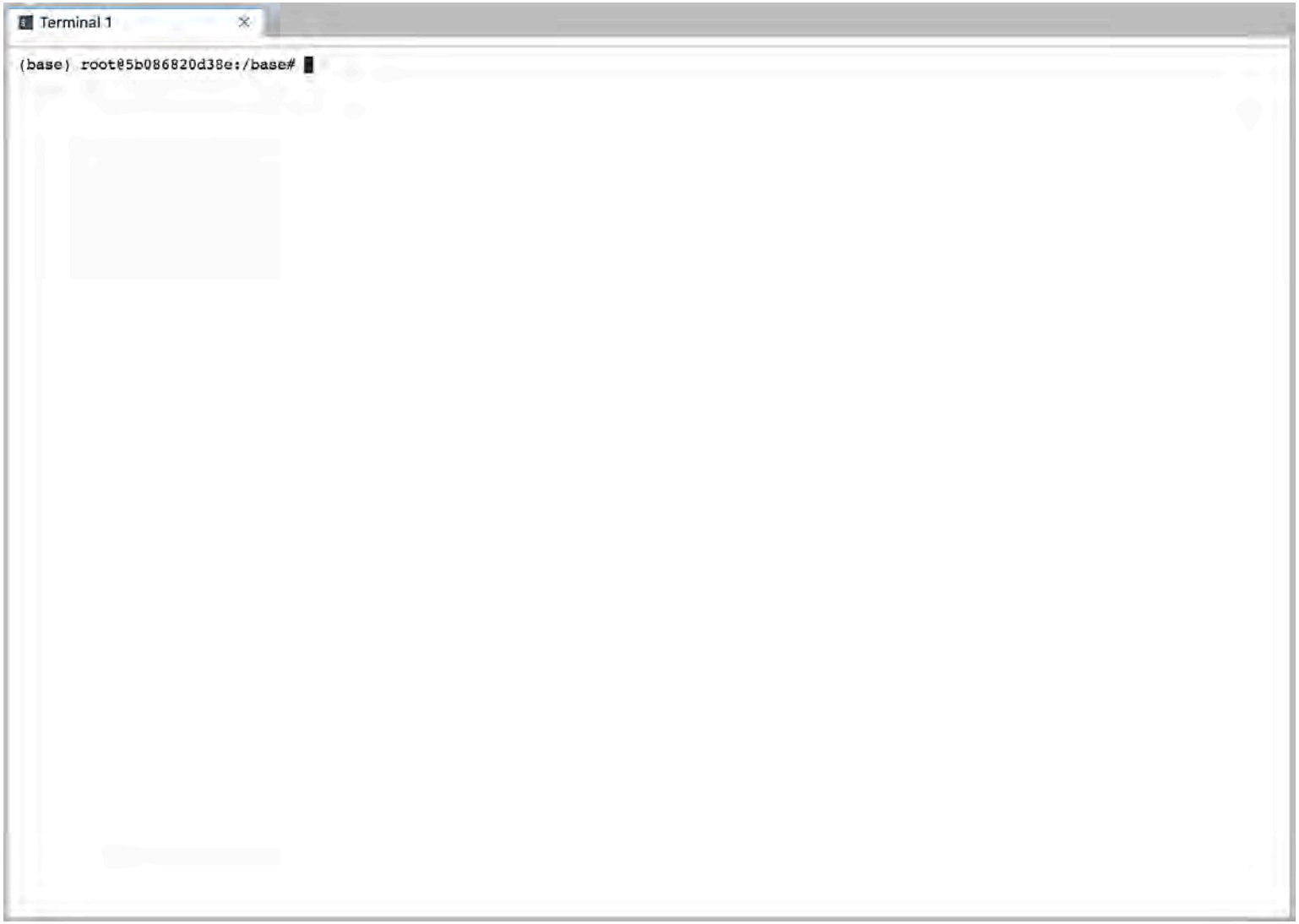
ShadowAdjudicator

Mike Giancola

- A nascent automated reasoner for generating and adjudicating arguments
- Builds upon ShadowProver
  - Uses ShadowProver for sub-proofs of modal/FOL/PL formulae
  - Implements an algorithm and inference schemata for generating arguments with strength factors



Simulation [rather < 48 hrs]



ICRES 2020: 5th International Conference on Robot
Ethics and Standards, Taipei, Taiwan, 28-29 September 2020.

ETHICAL REASONING FOR AUTONOMOUS AGENTS UNDER UNCERTAINTY

MICHAEL GIANCOLA* and SELMER BRINGSJORD† and NAVEEN SUNDAR GOVINDARAJULU* and CARLOS VARELA‡

Rensselaer AI & Reasoning (RAIR) Lab†; Worldwide Computing Lab (WCL)‡*
Department of Computer Science†‡; Department of Cognitive Science†*
Rensselaer Polytechnic Institute
Troy, NY 12180, USA
www.rpi.edu
E-mail: { *mike.j.giancola, selmer.bringsjord, naveen.sundar.g* } *@gmail.com, cvarela@cs.rpi.edu*

Autonomous (and partially autonomous) agents are beginning to play significant roles in safety-critical and privacy-critical domains, such as driving and healthcare. When humans operate in these spaces, not only are there regulations and laws dictating proper behavior, but crucially, neurobiologically normal humans can be expected to comprehend how to reason with certain principles to ensure that their actions are legally/ethically/prudentially correct (whether or not these humans choose to abide by the principles in question). It seems reasonable that we should hold autonomous agents to, minimally, the same standard we hold humans to. In this paper, we present a framework for autonomous aircraft piloting agents to reason about ethical problems in the context of emergency landings. In particular, we are concerned with ethical problems in which every option is equally unethical with regard to the ethical principles the options violate; and the only distinguishing factor is the *likelihood* that a plan will violate an ethical principle. We conclude by discussing why, in general, we find an inference-theoretic approach to ethical reasoning to be superior to the model-theoretic approach of prior work.

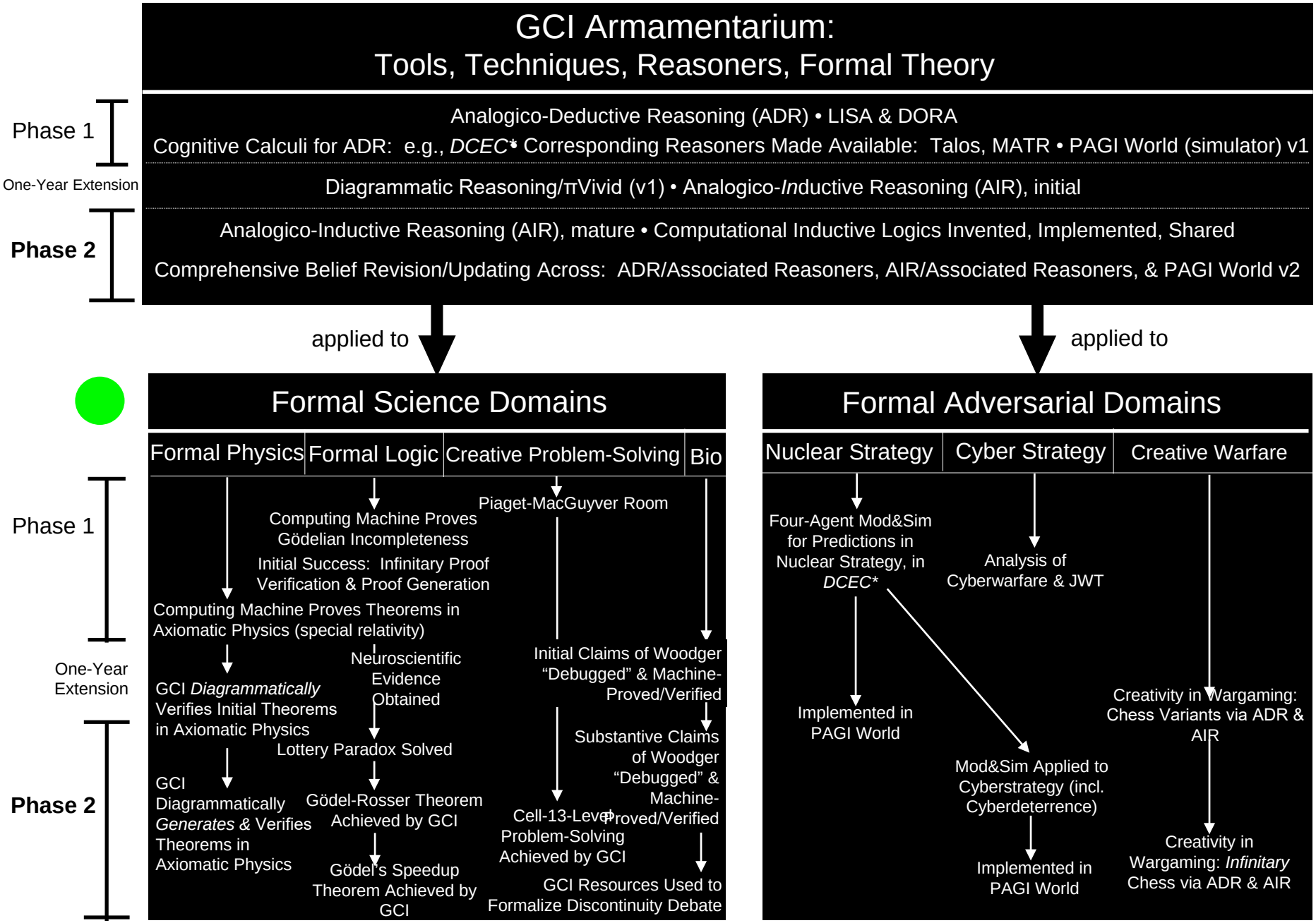
Keywords: Ethical reasoning; Reasoning under uncertainty; Modal logic.

1. Introduction

Autonomous (and partially autonomous) agents are beginning to play significant roles in safety-critical and privacy-critical domains, such as driving and healthcare. When humans operate in these spaces, not only are there regulations and laws dictating proper behavior, but crucially, neurobiologically normal humans can be expected to comprehend how to reason with regulations to ensure that their actions remain within the law (regardless of whether or not they choose to abide by the law). It seems reasonable that we should hold autonomous agents to, minimally, the same standard we hold humans to. That is, autonomous agents must be aware of relevant ethical constraints (those actions which are e.g. *obligatory*, *permissible*, *forbidden*, *supererogatory*, etc.; see [1] for a full discussion of these concepts in deontic logic), and be able to reason with them to determine actions which will satisfy those constraints. In addition, we require of our autonomous agents to not only verify their behavior to be ethical, but to output a proof which can be inspected by a human. By our lights, this requires an inference-theoretic approach to ethical reasoning. We will discuss in greater detail why we find an inference-theoretic approach to be superior to a model-theoretic approach in §4.

4. Progress on Speedup! …

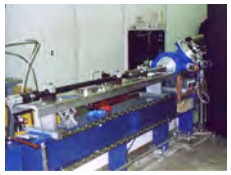
Great Computational Intelligence (GCI), Consummated & Applied



Ascending Acceleration

2 sec: 60 mph 5.5 sec: 100 mph 7.5 sec: 150 mph

20 sec: 268 mph 520 sec: 17,000 mph



1 sec: 20,000 mph

light-gas gun

$$PrRec{:}h(x,0) = f(x); h(x,y') = g(x,y,h(x,y))$$

exponentiation: $x^y = x \cdot x \cdot \ldots \cdot x (row\ of yxs)$

super-exponentiation (tetration): $x \uparrow (x \uparrow (x \uparrow \cdots \uparrow x))(yxs)$

Ackermann Function

$\alpha(x,y,z) = x\langle y \rangle z$ and $\gamma(x) = \alpha(x,x,x)$; then:

$\gamma(0) = 0 + 0 = 0$
$\gamma(1) = 1 \cdot 1 = 0$
$\gamma(2) = 2^2 = 4$
$\gamma(3) = 3^{3^3} = 3 \uparrow\uparrow 3 = 7,625,597,484,987$
$\gamma(4) = 4 \uparrow\uparrow 4 => 10^{1000}$ (note: 10^{100} is googol)

$$\Sigma: \mathbb{Z}^+ \mapsto \mathbb{Z}^+ where \Sigma(k) = max\ productivity\ of\ ak-state\ TM$$

Gödel's Speedup Theorem

Let $i \geq 0$, and let f be any recursive function.

Then there is an infinite family $\mathcal{F}$ of Π_1^0 formulae such that:

1. $\forall \phi \in \mathcal{F}, Z_i \vdash \phi$; and
2. $\forall \phi \in \mathcal{F}$, if k is the least integer s.t. $Z_{i+1} \vdash^{k \text{ symbols}} \phi$, then $Z_i \nvdash^{f(k) \text{ symbols}} \phi$.

AI's Moving Beyond FOL to SOL to … Is Highly Advisable!

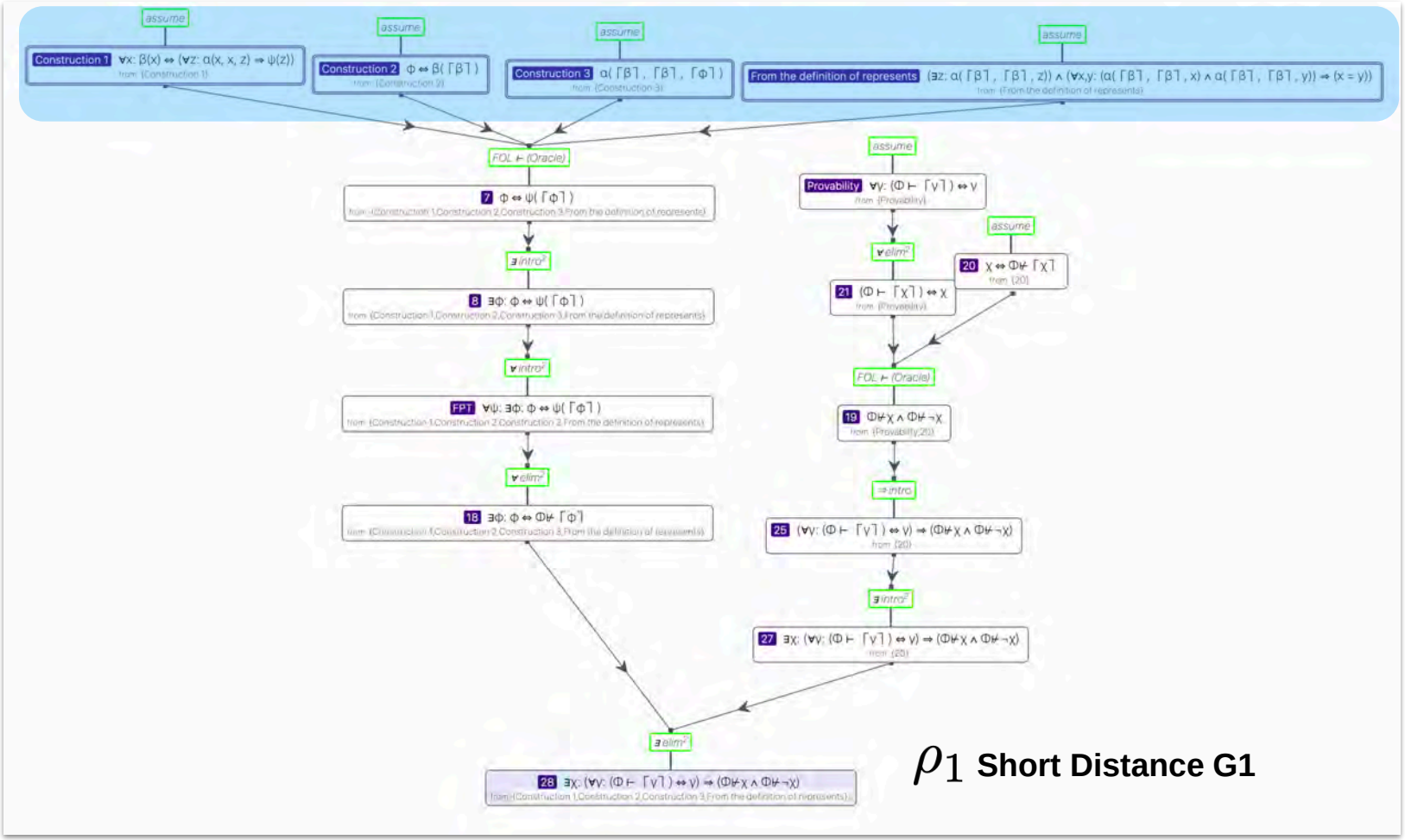
- Hummel, J. E. (2010) "Symbolic vs. Associative Learning" *Cognitive Science* 34: 958–965.
- Hummel, J. E., & Holyoak, K. J. (2003) "A Symbolic-Connectionist theory of Relational Inference and Generalization" *Psychological Review* **110**: 220–264.
- Markman, A & Gentner, D. (2001) "Thinking" *Annual Rev. Psychol.* 52: 223–247.
- & the reverse mathematicians!

Generalization of Inductive Calculus

- The inductive calculus shown in the previous slides applies only to first-order modal calculi.
- GST requires an inductive calculus that is second-order.

Apply IC to go from G1 to GST
(work in progress)

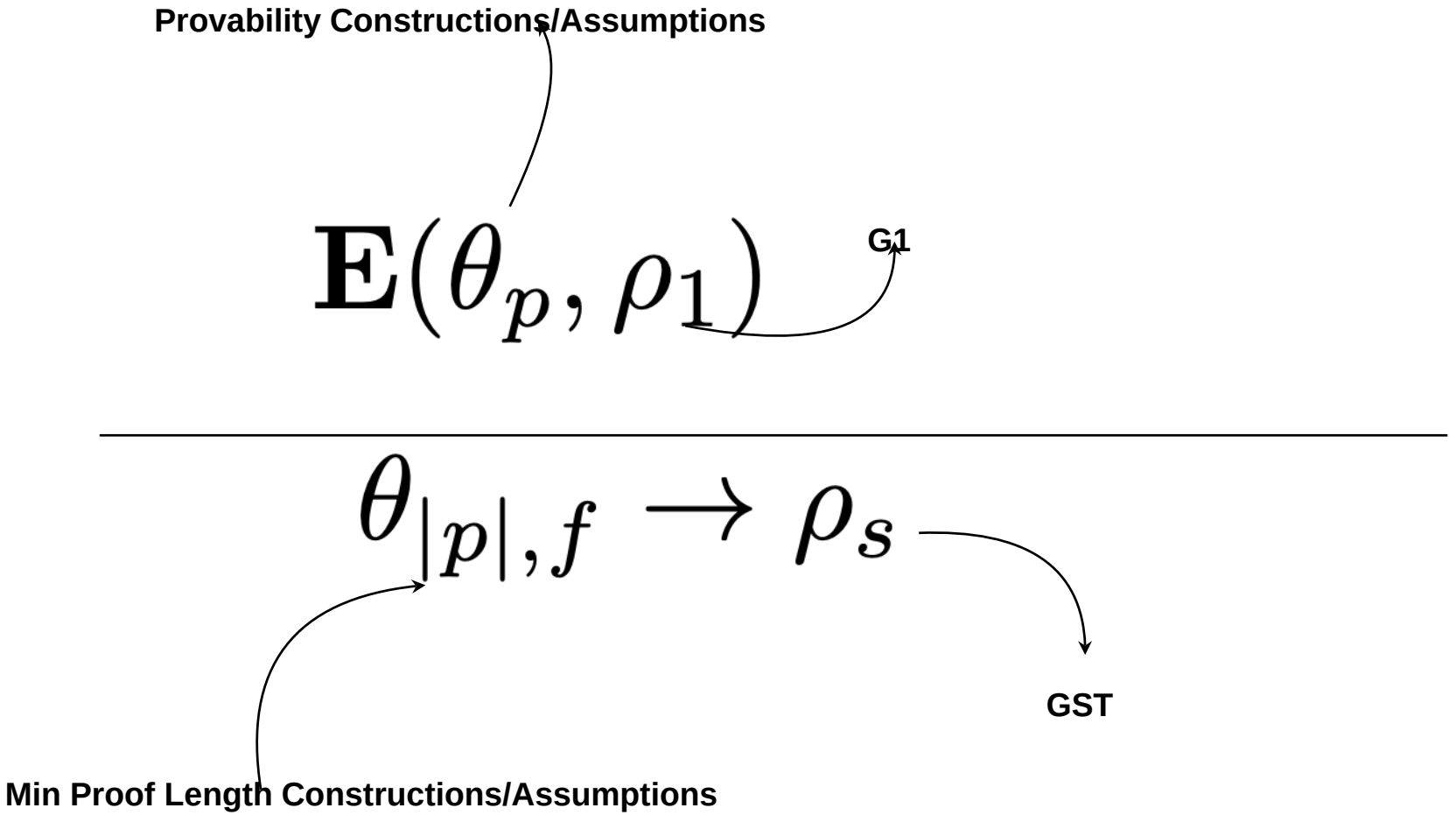
θ_p **Provability Constructions/Assumptions**



ρ_1 **Short Distance G1**

Apply IC to go from G1 to GST

(work in progress)



Second-order Inductive Calculus

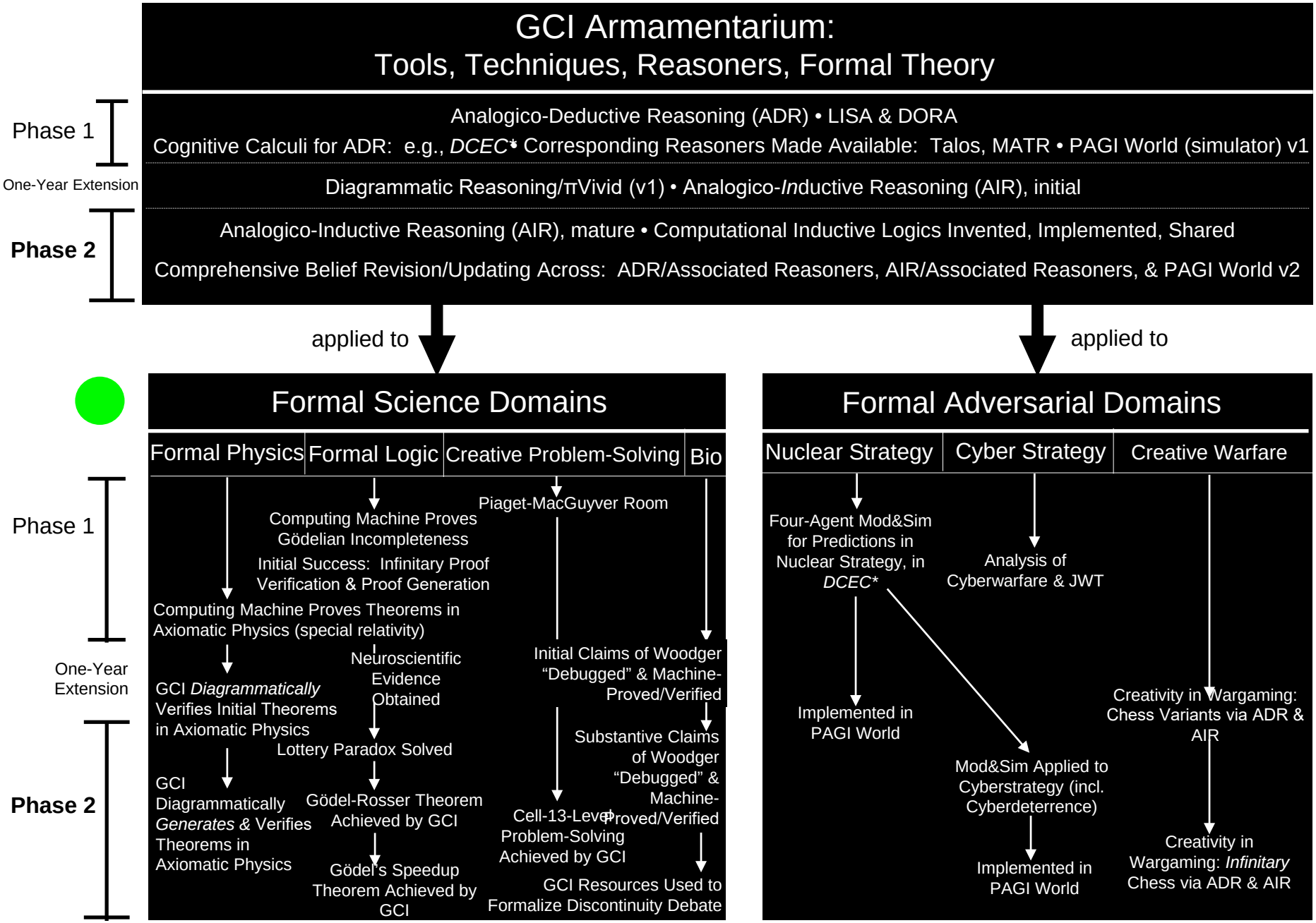
Same as first-order IC but we use

$g(\phi, \vec{\nu})$ Second-order generalization operator that can work with second-order logic

	Formulation Progress	**Implementation Progress**
First-order IC	~80%	~75%
Second-order IC	~60%	~40%

5. Progress Toward Resolving the Discontinuity Debate …

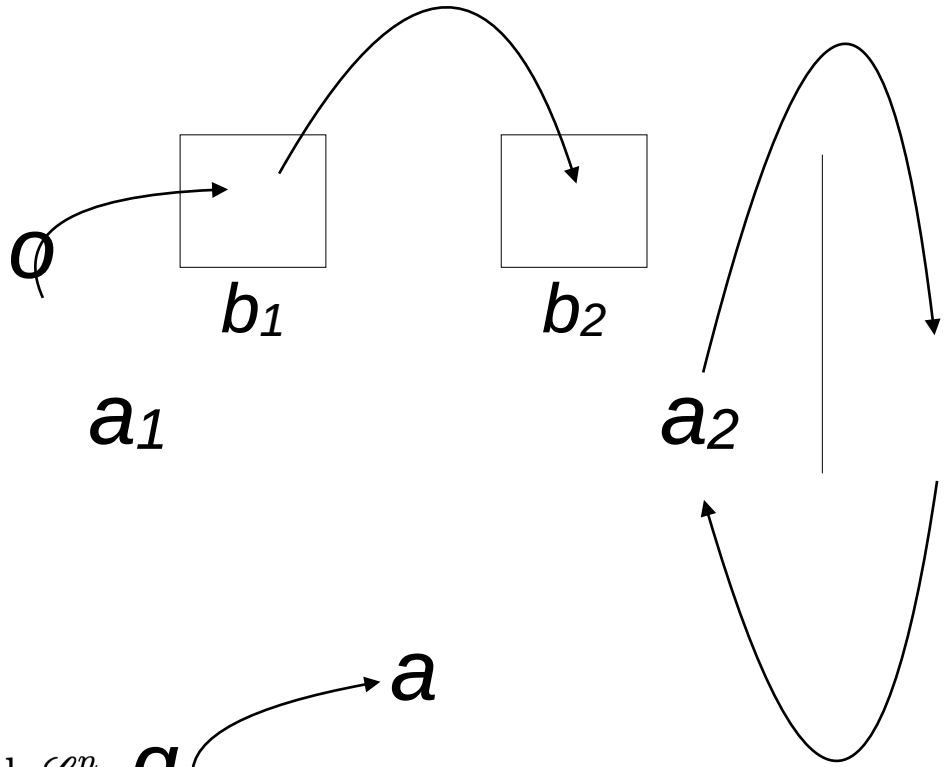
Great Computational Intelligence (GCI), Consummated & Applied



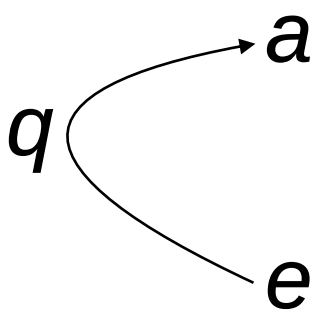
(requires high Λ)

Infinitary False Belief Task …

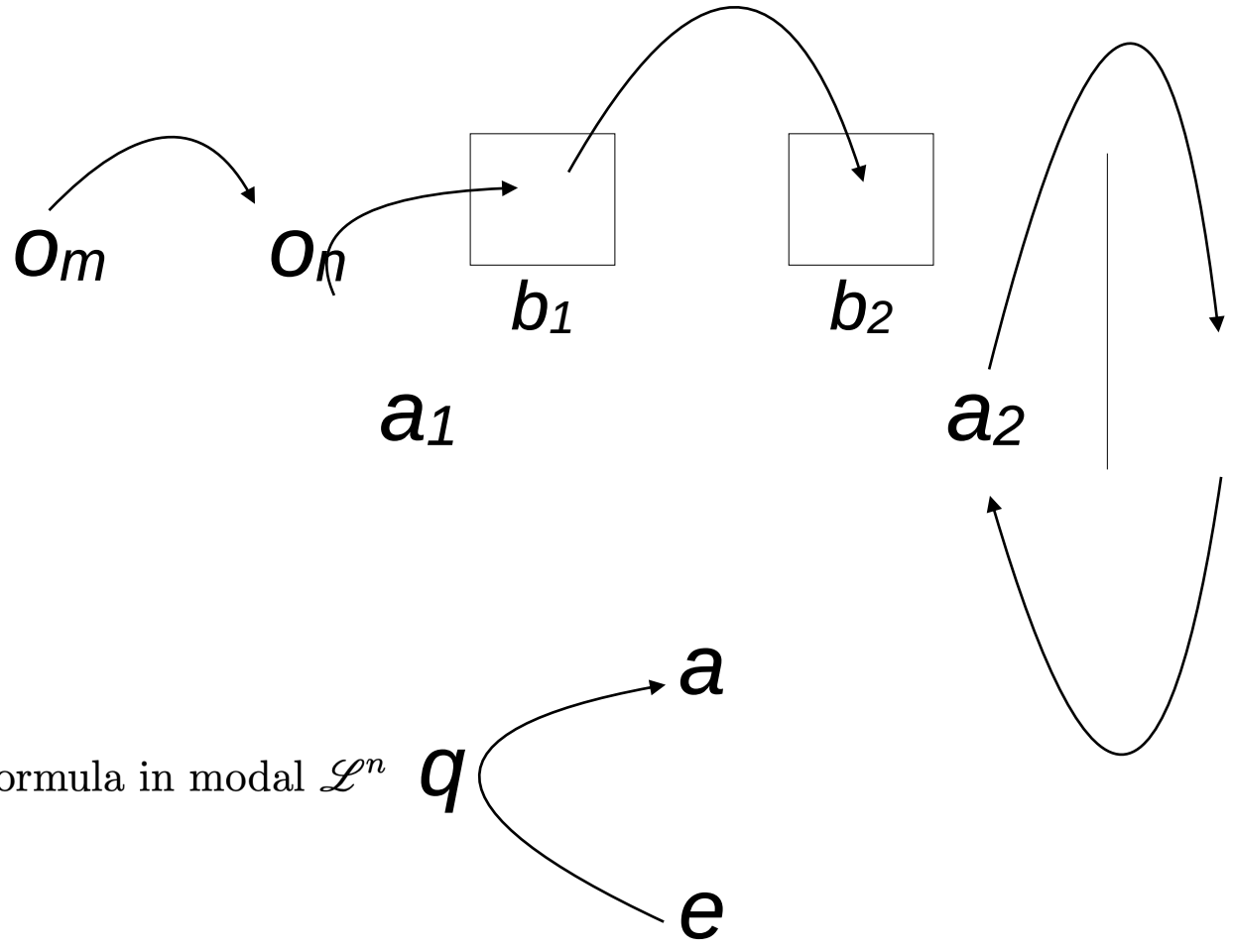
Framework for FBT^0_1
(five timepoints)



q a formula in modal $\mathscr{L}^n$



Framework for FBT^1_1
(six timepoints)



Done, a Decade Ago, Formally & Implementation/Simulation

Arkoudas, K. & Bringsjord, S. (2009) “Propositional Attitudes and Causation” *International Journal of Software and Informatics* **3.1**: 47–65.

http://kryten.mm.rpi.edu/PRICAI_w_sequentcalc_041709.pdf

Propositional attitudes and causation

Konstantine Arkoudas and Selmer Bringsjord

Cognitive Science and Computer Science Departments, RPI
arkouk@rpi.edu, brings@rpi.edu

Abstract. Predicting and explaining the behavior of others in terms of mental states is indispensable for everyday life. It will be equally important for artificial agents. We present an inference system for representing and reasoning about mental states, and use it to provide a formal analysis of the false-belief task. The system allows for the representation of information about events, causation, and perceptual, doxastic, and epistemic states (vision, belief, and knowledge), incorporating ideas from the event calculus and multi-agent epistemic logic. Unlike previous AI formalisms, our focus here is on mechanized proofs and proof programmability, not on metamathematical results. Reasoning is performed via relatively cognitively plausible inference rules, and a degree of automation is achieved by general-purpose inference methods and by a syntactic embedding of the system in first-order logic.

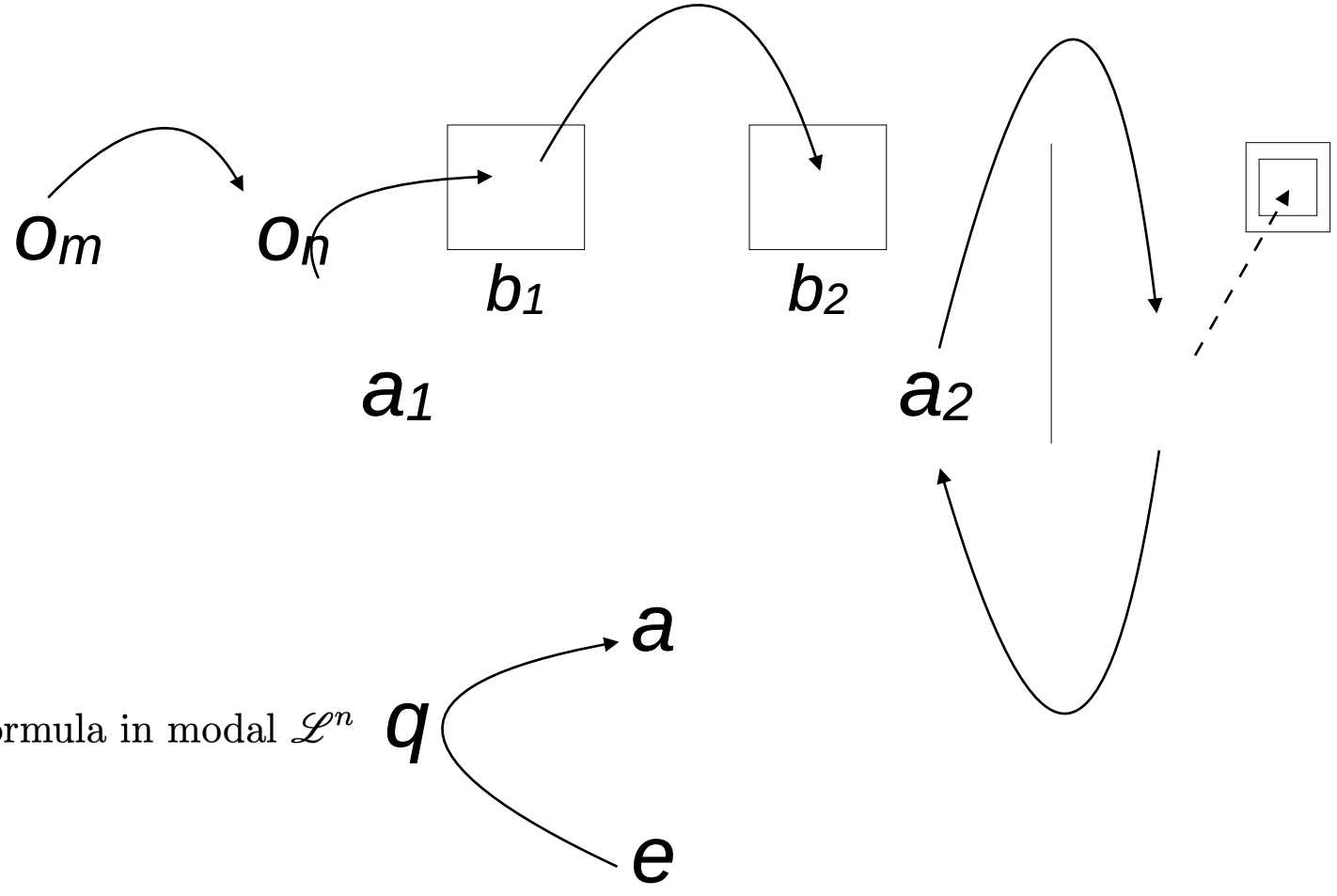
1 Introduction

Interpreting the behavior of other people is indispensable for everyday life. It is something that we do constantly, on a daily basis, and it helps us not only to make sense of human behavior, but also to predict it and—to a certain extent—to control it. How exactly do we manage that? That is not currently known, but many have argued that the ability to ascribe mental states to others and to reason about such mental states is a key component of our capacity to understand human behavior. In particular, all social transactions, from engaging in commerce and negotiating to making jokes and empathizing with other people's pain or joy, appear to require at least a rudimentary grasp of common-sense psychology (CSP), i.e., a large body of truisms such as the following: When an agent a (1) wants to achieve a certain state of affairs p, and (2) believes that some action c can bring about p, and (3) a knows how to carry out c; then, ceteris paribus,[1] a will carry out c; when a sees that p, a knows that p; when a fears that p and a discovers that p is the case, a is disappointed; and so on.

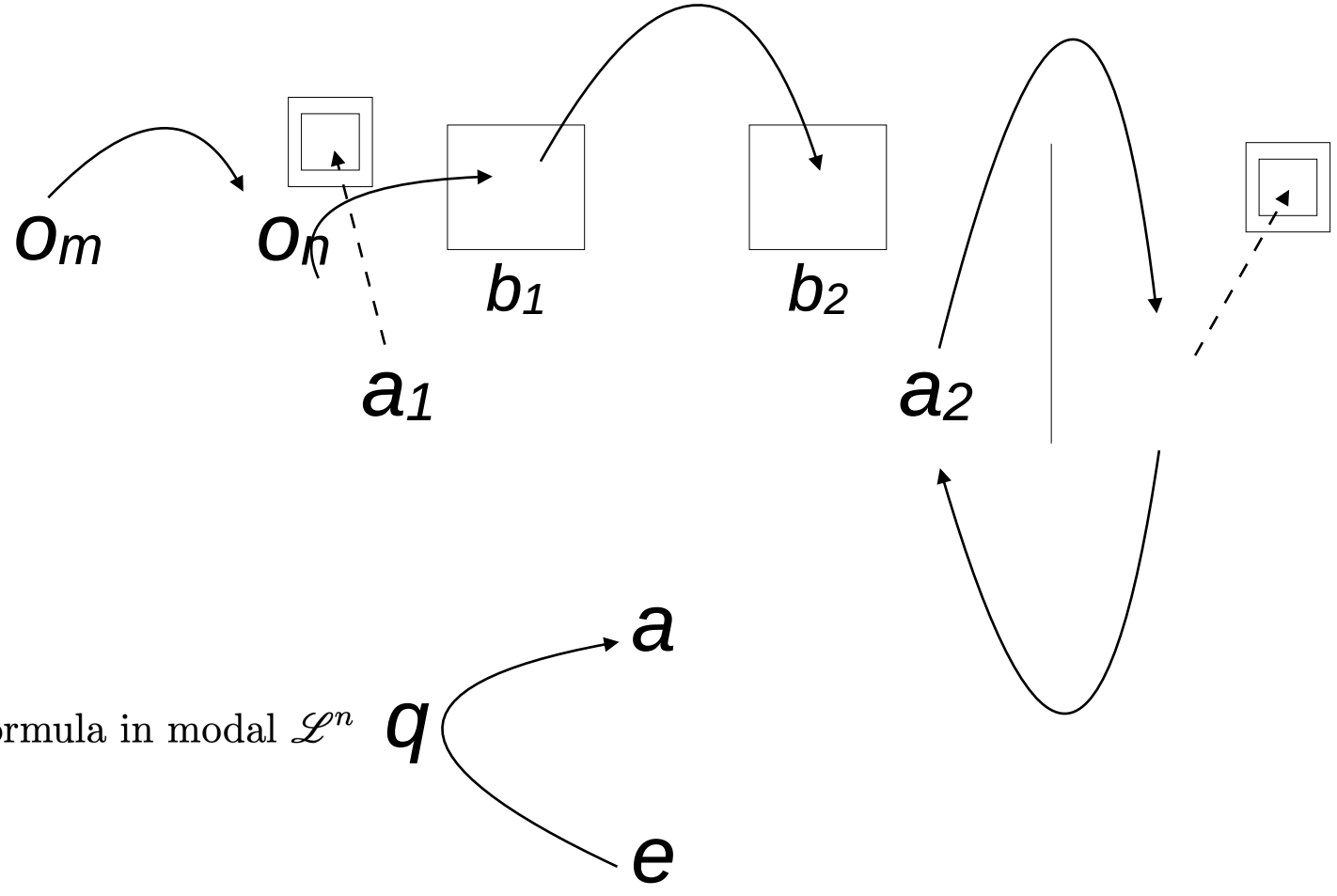
Artificial agents without a mastery of CSP would be severely handicapped in their interactions with humans. This could present problems not only for artificial agents trying to interpret human behavior, but also for artificial agents trying to interpret the behavior of one another. When a system exhibits a complex but rational behavior, and detailed knowledge of its internal structure is not

[1] Assuming that a is able to carry out c, that a has no conflicting desires that override his goal that p; and so on.

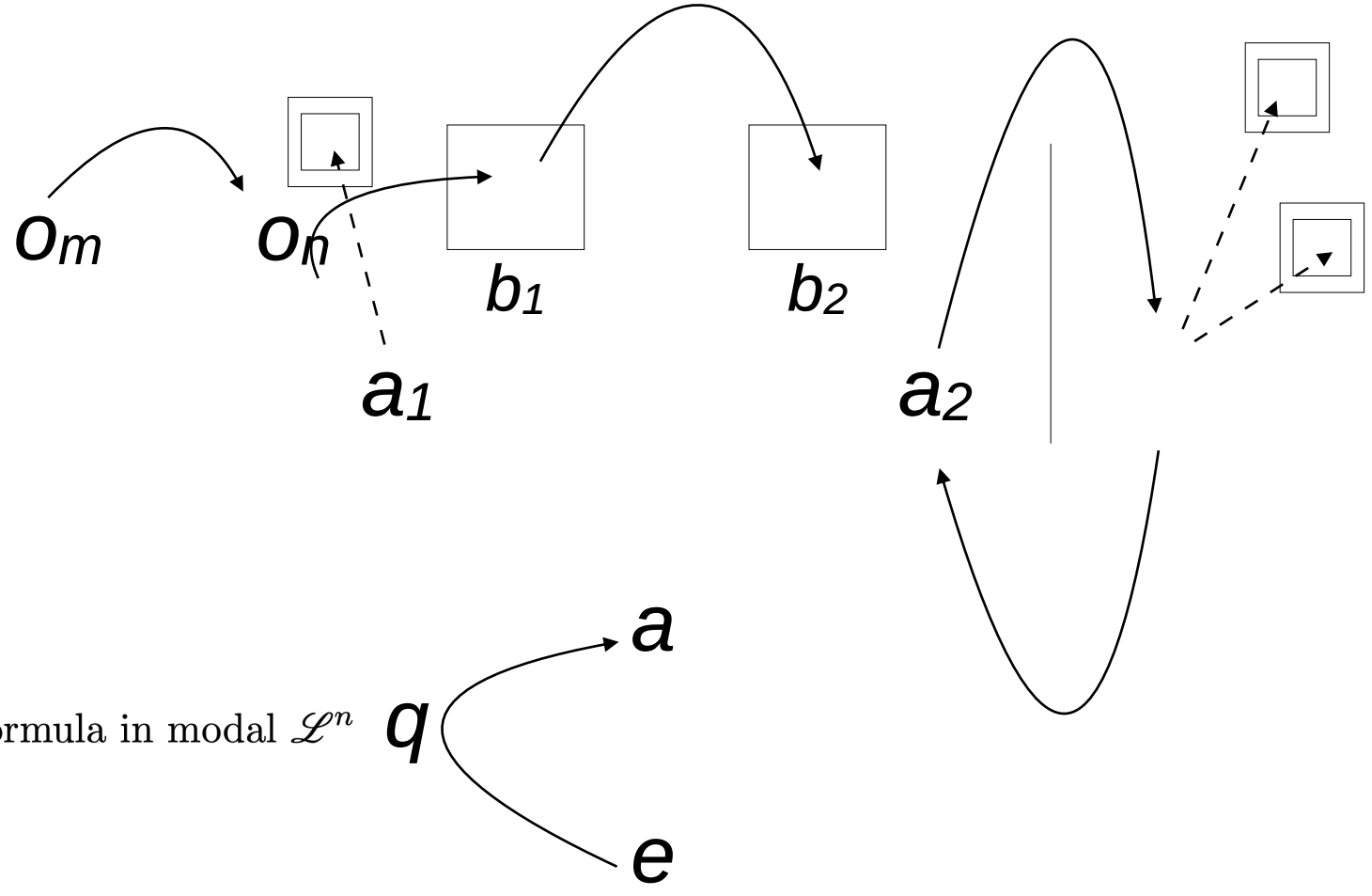
Framework for FBT^1_2
(seven timepoints)



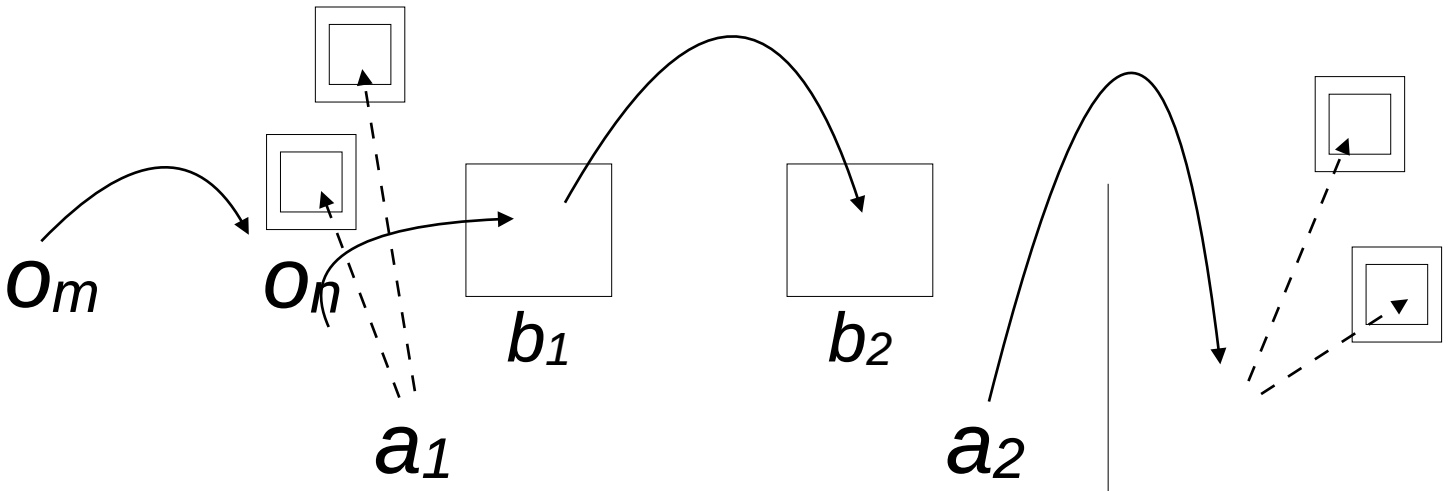
Framework for FBT^1_3
(eight timepoints)



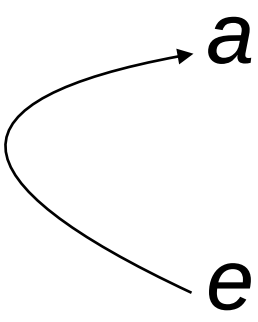
Framework for FBT^1_4
(nine timepoints)



Framework for FBT^1_5
(ten timepoints)



q a formula in modal $\mathscr{L}^n$



Humans Can Succeed

Neurobiologically normal, nurtured, educated, and sufficiently motivated humans can correctly answer any relevant query *q* for the infinite progression, and prove that their answer is correct. For the obvious subclass of queries (the form of which appear in the box below), they can prove and exploit the following lemma.

Lemma: Suppose $\mathrm{FBT}_k, k \in \mathbb{Z}^+$, holds; (i.e. that level k of FBT holds). Then, if k is even, $\mathbf{B}_2\mathbf{B}_1 \ldots \mathbf{B}_2\ \iota$, where there are $k+1$ iterated $\mathbf{B}_i$ operators; otherwise $\mathbf{B}_1\mathbf{B}_2 \ldots \mathbf{B}_1\mathbf{B}_2\ \iota$, where there again there are $k+1$ iterated $\mathbf{B}_i$ operators.

Passing to Probing Mastery of the Specific Subclass

Experimenter to *a*: "At level *k*, from which box will a_2 attempt to retrieve the objects o_n? Prove it!"

Theoretical Machine Success on Infinite FBT!

Theorem: $\forall q \in \mathscr{CC}, \mathfrak{M}$ can correctly answer and justify q. I.e., $\mathfrak{M}$ can pass FBT_{ω}.

Ok, so this logic machine exists in the *mathematical* universe; but does there exist an *implemented* machine with this power?

Yes …

Simulation Courtesy of …

ShadowProver!

Level 1

```
:name        "Level 1: False Belief Task "

:description "Agent a1 puts an object o into b1 in plain view of a2.
              Agent a2 then leaves, and in the absence of a2, a1 moves o
              from b1 into b2 ; this movement isn't perceived by a2 . Agent
              a2 now returns, and a is asked by the experimenter e: "If a2
              desires to retrieve o, which box will a2 look in?" If younger
              than four or five, a will reply "In b " (which of course fails 2
              the task); after this age subjects respond with the correct "In b1."

              Level1 Belief: a1 believes a2 believes o is in b1.
              "

:date        "Monday July 22, 2019"

:assumptions {
               :P1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1)))

               :P2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1))))))

               :P3 (holds (In o b1) t1)

               :C1 (Common! t0 (forall [?f ?t2 ?t2]
                                       (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                         (holds ?f ?t2))))

               :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))
               }

:goal        (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3)))}
```

Level 2

```
{:name        "Level 2: False Belief Task "

 :description "Agent a1 puts an object o into b1 in plain view of a2.
               Agent a2 then leaves, and in the absence of a2, a1 moves o
               from b1 into b2 ; this movement isn't perceived by a2 . Agent
               a2 now returns, and a is asked by the experimenter e: "If a2
               desires to retrieve o, which box will a2 look in?" If younger
               than four or five, a will reply "In b " (which of course fails 2
               the task); after this age subjects respond with the correct "In b1."

               Level2 Belief: a2 believes a1 believes a2 believes o is in b1.
               "

 :date        "Monday July 22, 2019"

 :assumptions {

                :P1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1))))

                :P2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1)))))))

                :P3 (holds (In o b1) t1)


                :C1 (Common! t0
                             (forall [?f ?t2 ?t2]
                                     (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                       (holds ?f ?t2))))

               :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}

 :goal        (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3))))}
```

Level 3

```
{:name        "Level 3: False Belief Task "

 :description "Agent a1 puts an object o into b1 in plain view of a2.
               Agent a2 then leaves, and in the absence of a2, a1 moves o
               from b1 into b2 ; this movement isn't perceived by a2 . Agent
               a2 now returns, and a is asked by the experimenter e: "If a2
               desires to retrieve o, which box will a2 look in?" If younger
               than four or five, a will reply "In b " (which of course fails 2
               the task); after this age subjects respond with the correct "In b1."

               Level3 Belief: a2 believes a1 believes a2 believes o is in b1.
               "

 :date        "Monday July 22, 2019"

 :assumptions {

                :P1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1)))))
                :P2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1)))))))))

                :P3 (holds (In o b1) t1)


                :C1 (Common! t0
                             (forall [?f ?t2 ?t2]
                                     (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                       (holds ?f ?t2))))

                :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}

 :goal        (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3)))))}
```

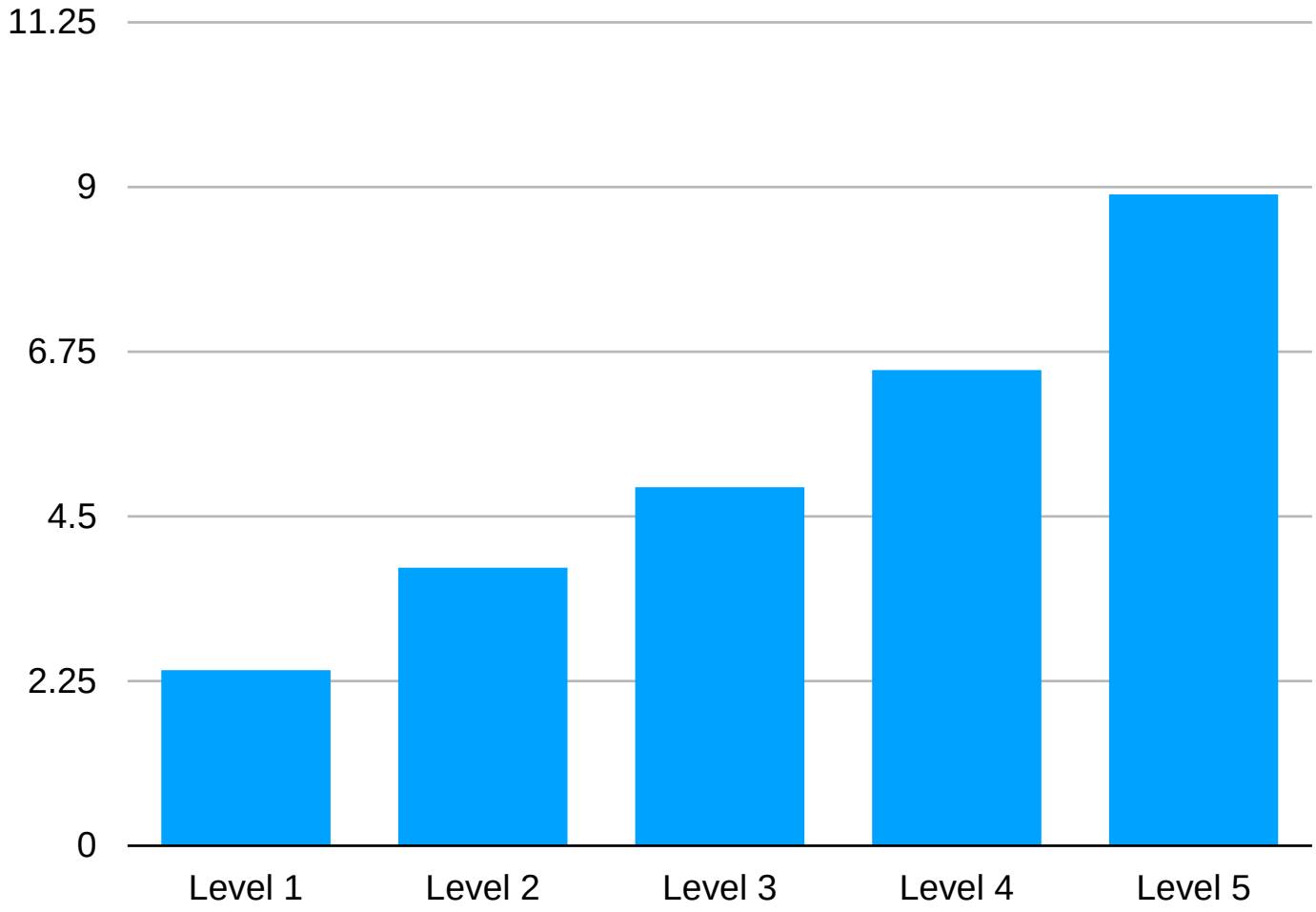
Level 4

```
{:name        "Level 4: False Belief Task "

 :description "Agent a1 puts an object o into b1 in plain view of a2.
               Agent a2 then leaves, and in the absence of a2, a1 moves o
               from b1 into b2 ; this movement isn't perceived by a2 . Agent
               a2 now returns, and a is asked by the experimenter e: "If a2
               desires to retrieve o, which box will a2 look in?" If younger
               than four or five, a will reply "In b " (which of course fails 2
               the task); after this age subjects respond with the correct "In b1."

               Level4 Belief: a2 believes a1 believes a2 believes a1 believes a2 believes o is in b1.
               "

 :date        "Monday July 22, 2019"

 :assumptions {

                :P1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1))))))
                :P2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1)))))))))

                :P3 (holds (In o b1) t1)


                :C1 (Common! t0
                             (forall [?f ?t2 ?t2]
                                     (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                       (holds ?f ?t2))))

                :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}

 :goal        (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3))))))}
```

Level 5

```
{:name        "Level 5: False Belief Task "

 :description "Agent a1 puts an object o into b1 in plain view of a2.
               Agent a2 then leaves, and in the absence of a2, a1 moves o
               from b1 into b2 ; this movement isn't perceived by a2 . Agent
               a2 now returns, and a is asked by the experimenter e: "If a2
               desires to retrieve o, which box will a2 look in?" If younger
               than four or five, a will reply "In b " (which of course fails 2
               the task); after this age subjects respond with the correct "In b1."

               Level5 Belief: a1 believes a2 believes a1 believes a2 believes a1 believes a2 believes o is in b1.
               "

 :date        "Monday July 22, 2019"

 :assumptions {

               :P1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (Perceives! a1 t1 (Perceives! a2 t1 (holds (In o b1) t1)))))))
               :P2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (Believes! a1 t2 (Believes! a2 t2 (not (exists [?e] (terminates ?e (In o b1)))))))))))

               :P3 (holds (In o b1) t1)


               :C1 (Common! t0
                            (forall [?f ?t2 ?t2]
                                    (if (and (not (exists [?e] (terminates ?e ?f))) (holds ?f ?t1) (< ?t1 ?t2))
                                      (holds ?f ?t2))))

               :C2 (Common! t0 (and (< t1 t2) (< t2 t3) (< t1 t3)))}

 :goal        (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (Believes! a1 t3 (Believes! a2 t3 (holds (In o b1) t3)))))))}
```

Time (in seconds) to Prove



Simulation of Level 5 in Real Time

```
/Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java ...
objc[16653]: Class JavaLaunchHelper is implemented in both /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java (0x102a2d4c0) and /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/jre/lib/libinstrument.dylib (0x102ab94e0)
---------- Level 5 ----------
```

6. Re Some Pubs

A novel Form of machine learning based on GCI formalisms:

Learning *Ex Nihilo*

(or Learning *Ex Minima*)

EPiC Series in Computing

Volume 72, 2020, Pages 1–27

GCAI 2020. 6th Global Conference on Artificial Intelligence (GCAI 2020)

Learning *Ex Nihilo*

Selmer Bringsjord[1], Naveen Sundar Govindarajulu[1], John Licato[2], and Michael Giancola[1]

[1] Rensselaer AI & Reasoning (RAIR) Lab,
Rensselaer Polytechnic Institute, Troy NY 12180 USA
[2] Advancing Machine and Human Reasoning (AMHR) Lab,
University of South Florida, Tampa FL 33620 USA
{ selmer.bringsjord, naveen.sundar.g, mike.j.giancola, 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 inference 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., here, 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 spirit 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 inference 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

[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 for us 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 which more will be said later), or inductive deductive techniques and schemas (e.g. mathematical induction, the induction schema in Peano Arithmetic, etc.). As we explain later, learning *ex nihilo* is in part powered by non-deductive inference schemata seen in inductive logic. An introductory overview of inductive logic is provided in [30].

Bringsjord, S., Govindarajulu, N.S., Licato, J. & Giancola, M. (2020) “Learning *Ex Nihilo*" *Proceedings of the 6th Global Conference on Artificial Intelligence* (GCAI 2020), within *International Conferences on Logic and Artificial Intelligence* at Zhejiang University (ZJULogAI), in Danoy, G., Pang, J. & Sutcliffe, G., eds., EPiC Series in Computing ***72***: 1–27 (Manchester, UK: EasyChair Ltd), ISSN: 2398-7340. https://easychair.org/publications/paper/NzWG

List of Publications Attributed to the Grants (selected)

- *Bringsjord, S., Govindarajulu, N.S., Licato, J. & Giancola, M. (2020) "Learning Ex Nihilo" Proceedings of the 6th Global Conference on Artificial Intelligence (GCAI 2020), within International Conferences on Logic and Artificial Intelligence at Zhejiang University (ZJULogAI), in Danoy, G., Pang, J. & Sutcliffe, G., eds., EPiC Series in Computing **72**: 1–27 (Manchester, UK: EasyChair Ltd), ISSN: 2398-7340. https://easychair.org/publications/paper/NzWG*
- *Bringsjord, S. & Govindarajulu, N.S. (2020) "The Theory of Cognitive Consciousness, and Λ (Lambda)." Journal of AI & Consciousness **7.2**: 155–181.*
- *Bringsjord, S. & Govindarajulu, N.S. (2020) "Review of Fundamental Proof Methods in Computer Science." Theory and Practice of Logic Programming. DOI: https://doi.org/10.1017/S1471068420000071. 8 pages. Online version now up from TPLP/Cambridge U Press.*
- *Giancola, M., Bringsjord, S., Govindarajulu, N.S. & Varela, C. (2020) "Ethical Reasoning for Autonomous Agents Under Uncertainty"' in Tokhi, M.O., Ferreira, M.I.A., Govindarajulu, N.S., Silva, M.F., Kadar, E.E., Wang, Jen-Chien, Kaur, A.P.,~eds., Smart Living and Quality Health with Robots, Proceedings of ICRES 2020, Taipai, Taiwan, September 28--29 2020 (London, UK: CLAWAR), pp. 26–41.*
- *Marji, Z., Nighojkar, A., & Licato, J. (2020) "Probing the Natural Language Inference Task with Automated Reasoning Tools." In Proceedings of The 33rd International Florida Artificial Intelligence Research Society Conference (FLAIRS-33). AAAI Press. https://arxiv.org/abs/2005.02573*
- *Malaby, E., Dragun, B., & Licato, J. (2020). "Towards Concise, Machine-discovered Proofs of Gödel's Two Incompleteness Theorems." In Proceedings of The 33rd International Florida Artificial Intelligence Research Society Conference (FLAIRS-33). AAAI Press. https://arxiv.org/abs/2005.02576*
- *Quandt, R. & Licato, J. (2020). "Problems of Autonomous Agents Following Informal, Open-Textured Rules." In Lawless, W.F., Mittu, R., & Sofge, D.A., eds. Human-Machine Shared Contexts. Academic Press.*
- *Bringsjord, S. (2020) "Computer Science as Immaterial Formal Logic" Philosophy & Technology **33.2**: 339--347. The link below is to a preprint: http://kryten.mm.rpi.edu/ComputerScienceAsImmaterialFormalLogic.pdf*
- *Bringsjord, S. & Govindarajulu, N.S. (2019) ``Introducing \Lambda for Measuring Cognitive Consciousness'' in Chella, A., Gamez, D., Lincoln, P., Manzotti, R., Pfautz, J., Proceedings of TOCAIS19 Toward Conscious AI Systems), Stanford, CA, March 25–27, 2019. ISSN 1613-0073, Vol-2287.*
- *BBringsjord, S.,~Govindarajulu,~Elmore C.\ (2019) ``Logicist Computational Cognitive Modeling of Infinitary False-Belief Tasks,'' In A.K. Goel, C.M. Seifert, & C. Freksa, eds. Proceedings of the 41st Annual Conference of the Cognitive Science Society} (Montreal, QB: Cognitive Science Society), pp. 43–45.*
- *Govindarajulu, N.S.\ \& Bringsjord, S. (2019) ``Towards a Computable \& Harnessable Model of Consciousness'' in Chella, A., Gamez, D., Lincoln, P., Manzotti, R., Pfautz, J. Proceedings of TOCAIS S19} (Toward Conscious AI Systems), Stanford, CA, March 25–27, 2019. ISSN 1613–0073, Vol–2287.*
- *Bringsjord, S., Govindarajulu, N.S., Banerjee, S., & Hummel, J. (2018) "Do Machine-Learning Machines Learn?" in Müller, V., ed., Philosophy and Theory of Artificial Intelligence 2017 (Berlin, Germany: Springer SAPERE), pp. 136–157, Vol. 44 in the book series.*
- *Heaton, R. F., & Hummel, J. E. (in press). Rapid Unsupervised Encoding of Object Files for Visual Reasoning. To appear in Proceedings of the 2019 Meeting of the Cognitive Science Society.*
- *Bringsjord, S. & Govindarajulu, N.S. (2018) "Artificial Intelligence" Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/artificial-intelligence/*
- *Licato, J. & Cooper, M. (2019). Evaluating Relevance in Analogical Arguments through Warrant-based Reasoning. In Proceedings of the European Conference on Argumentation (ECA 2019).*
- *Licato, J., Boger, M. & Zhang, Z. (2018) "Developing a Dataset for Personal Attacks and Other Indicators of Biases," Proceedings of the AAAI 2018 Spring Symposium on "Beyond Machine Intelligence.*
- *Bringsjord, S. & Govindarajulu, N.S. (2018) "Are Autonomous-and-Creative Machines Intrinsically Untrustworthy?" in Hussein Abbass, H.,~Scholz, J.,~Reid, D., eds. Foundations of Trusted Autonomy} (Springer: Cham, Switzerland), pp. 317-–335.*
- *Sen, A. (2017) Computational Axiomatic Science (PhD dissertation enabled by AFOSR support, computer science, RPI).*
- *Hummel, J. E. (2017). "Putting Distributed Representations into Context" Language, Cognition and Neuroscience, **32**, 3, 359–365. http://www.tandfonline.com/toc/plcp21/32/3*
- *Hummel, J., Licato, J. & Bringsjord, S. (2014) "Analogy, Explanation, and Proof" Frontiers in Neuroscience. Paper available here: https://doi.org/10.3389/fnhum.2014.00867*

List of Publications Attributed to the Grants (selected con.; older)

- *Sen, A., Bringsjord, S., Marton, N. & Licato, J. (2015) "Toward Diagrammatic Automated Discovery in Axiomatic Physics" Proceedings of Logic, Relativity, and Beyond II, Renyi Institute of Mathematics, Budapest, Hungary.*
- *Bringsjord, S., Govindarajulu, N.S., Ellis, S, McCarty, E. & Licato, J. J. (2014) "Nuclear Deterrence and Logic of Deliberative Mindreading." J. of Cognitive Systems Research 28: 20–43.*
- *2013: NATO CCD COE, IJCAI, UCNC papers (recall above); respectively:*
  - *By "Disanalogy, Cyberwarfare is Utterly New"; Bringsjord & Licato.*
    - *Greatly expanded target paper now under construction for NATO CCD COE. Will appear in Philosophy of Technology.*
  - *"Analogico-Deductive Generation of Gödel's First Incompleteness Theorem from the Liar Paradox"; Licato, Govindarajulu, Bringsjord, Pomeranz, Gittelson. IJCAI 2013.*
  - *"Small Steps Toward Hypercomputation via Infinitary Proof Verification and Proof Generation"; Govindarajulu, Licato, Bringsjord. UCNC.*
- *Bringsjord, S., Licato, J. & Hummel, J. (2012) "Psychometric Artificial General Intelligence: The Piaget-MacGuyver Room" in Wang, P. & Goertzel, B., eds., Theoretical Foundations of Artificial General Intelligence (Amsterdam, The Netherlands: Atlantis Press), pp. 25–47.*
- *Knowlton, B. J., Morrison, R. G., Hummel, J. E., & Holyoak, K. J. (2012) "A Neurocomputational System for Relational Reasoning" Trends in Cognitive Sciences 17: 373–381.*
- *Licato, J., Bringsjord, S., & Hummel, J.E. (2012) "Exploring the Role of Analogico-Deductive Reasoning in the Balance-Beam Task." In Rethinking Cognitive Development: Proceedings of the 42nd Annual Meeting of the Jean Piaget Society.*
  - *https://docs.google.com/open?id=0B1S661sacQp6NDJ0YzVXajJMWVU*
- *Govindarajulu, N.S. & Bringsjord, S. (2012) "Proof Verification and Proof Discovery for Relativity." Proceedings of First International Conference on Logic and Relativity, Budapest, Hungary, MTA Rényi Alfréd Matematikai Kutatóintézet. Full paper, with first substantive machine-verified and semi-automated theorem in relativity theory, available at*
  - *http://kryten.mm.rpi.edu/Govindarajulu-Bringsjord_proof_discovery_verification.pdf.*

Med nok penger, kan logikk løse alle problemer.