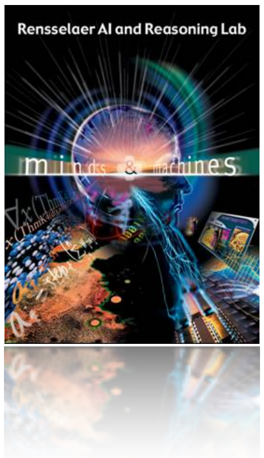
What Does *Watson 2.0* Tell us About the Philosophy, Theory, & Future of AI?

Selmer Bringsjord[1] • **Naveen Sundar G.**[2]
Rensselaer AI & Reasoning (RAIR) Lab[1,2]
Department of Cognitive Science[1]
Department of Computer Science[1,2]
Lally School of Management & Technology[1]
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

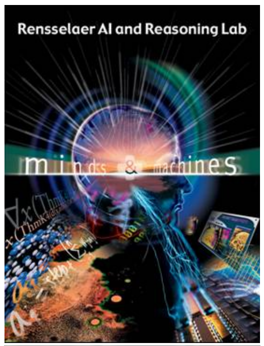
Oxford University
Oxford, UK
9/21/2013



What Does *Watson 2.0* Tell us About the Philosophy, Theory, & Future of AI?

Selmer Bringsjord[1] • **Naveen Sundar G.**[2]
Rensselaer AI & Reasoning (RAIR) Lab[1,2]
Department of Cognitive Science[1]
Department of Computer Science[1,2]
Lally School of Management & Technology[1]
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

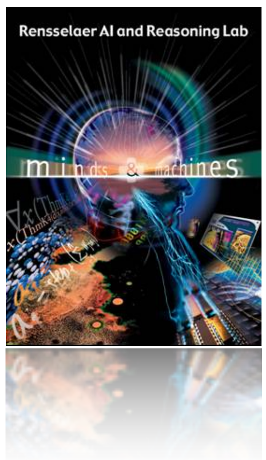
Oxford University
Oxford, UK
9/21/2013



What Does *Watson 2.0* Tell us About the Philosophy, Theory, & Future of AI?

Selmer Bringsjord[1] • **Naveen Sundar G.**[2]
Rensselaer AI & Reasoning (RAIR) Lab[1,2]
Department of Cognitive Science[1]
Department of Computer Science[1,2]
Lally School of Management & Technology[1]
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

Oxford University
Oxford, UK
9/21/2013



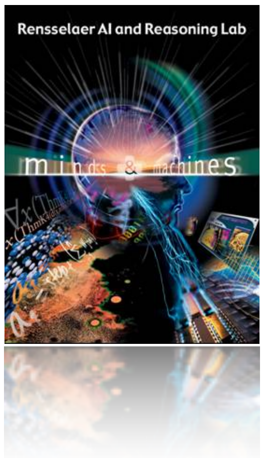
Made possible by the generosity of:

IBM
RPI's pursuit of ITS/IILE
NSF
JTF
AFOSR
ONR x 2
DARPA

What Does *Watson 2.0* Tell us About the Philosophy, Theory, & Future of AI?

Selmer Bringsjord[1] • **Naveen Sundar G.**[2]
Rensselaer AI & Reasoning (RAIR) Lab[1,2]
Department of Cognitive Science[1]
Department of Computer Science[1,2]
Lally School of Management & Technology[1]
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

Oxford University
Oxford, UK
9/21/2013

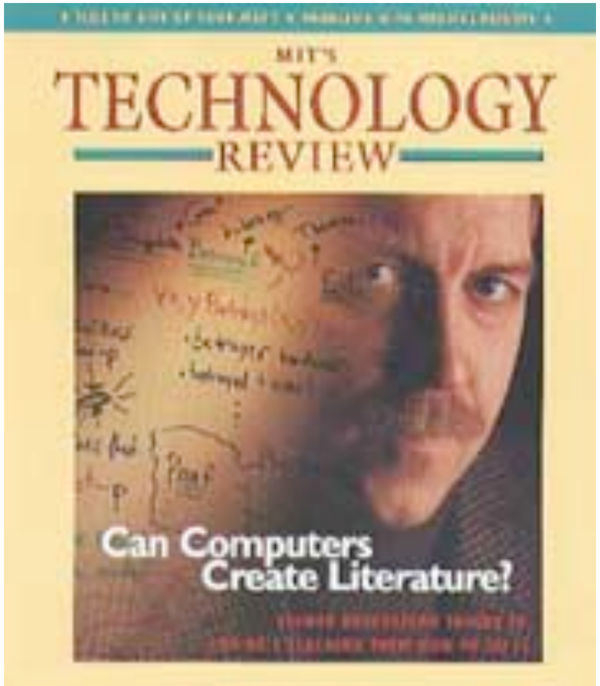


Made possible by the generosity of:

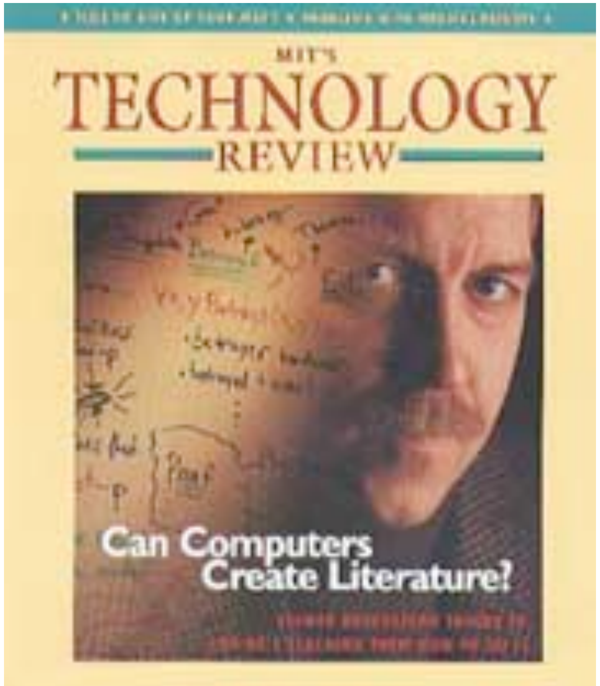
IBM
RPI's pursuit of ITS/IILE
NSF
JTF
AFOSR
ONR x 2
DARPA

Simon: “Chess is a lofty target for AI, yes. Go for it!”

“Chess is Too Easy”



“Chess is Too Easy”



I was right :).

What about this, then?

What about this, then?

What about this, then?

Experts to IBM: “Can’t be done!”

What about this, then?

Experts to IBM: “Can’t be done!”

No one asked me.

What about this, then?

Experts to IBM: “Can’t be done!”

No one asked me.

Bits

MAY 6, 2013, 3:37 PM | 2 Comments

David Ferrucci: Life After Watson

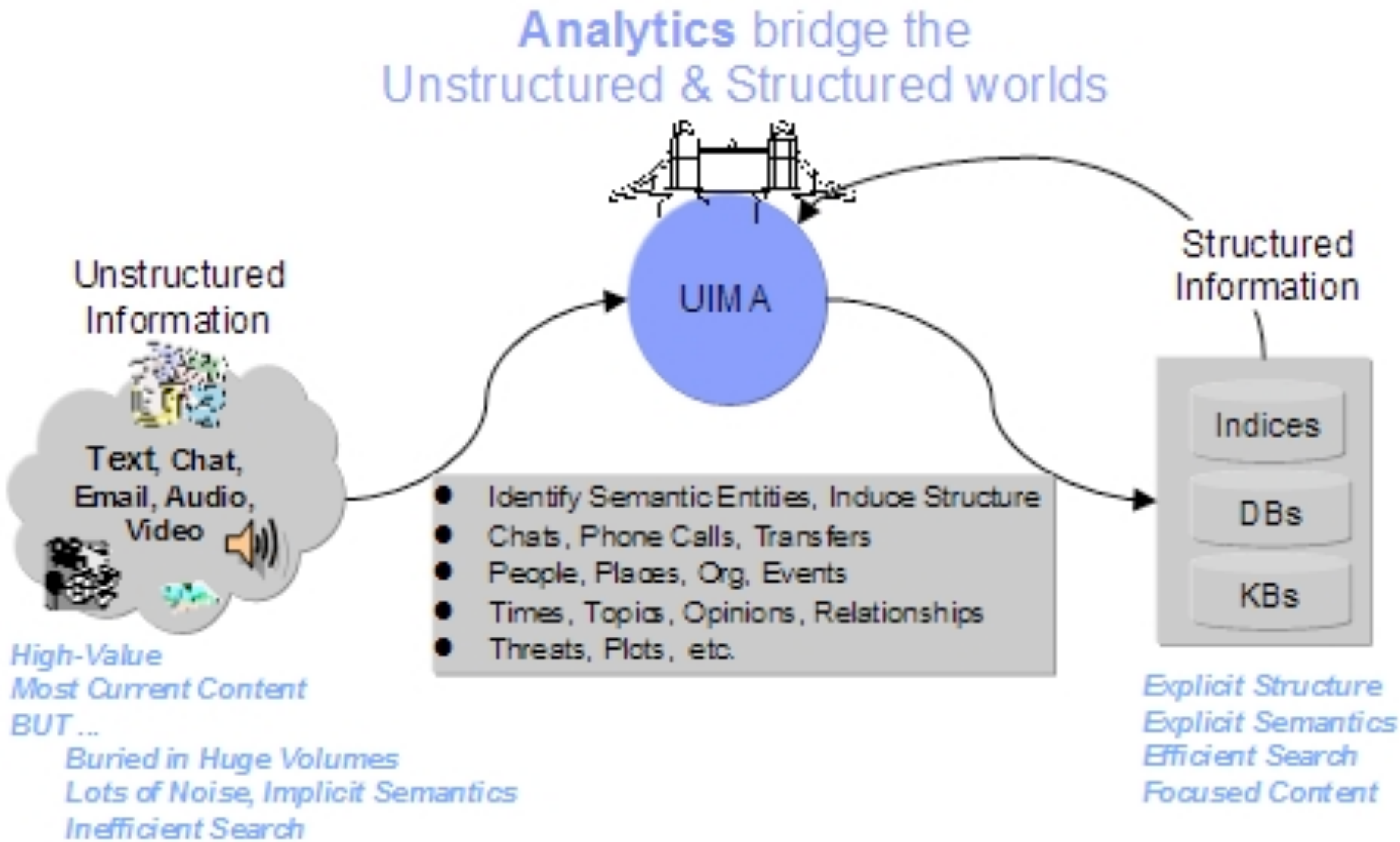
By STEVE LOHR

Suzanne DeChillo/The New York Times

David Ferrucci has left I.B.M., and Watson, and joined the hedge fund, Bridgewater Associates.

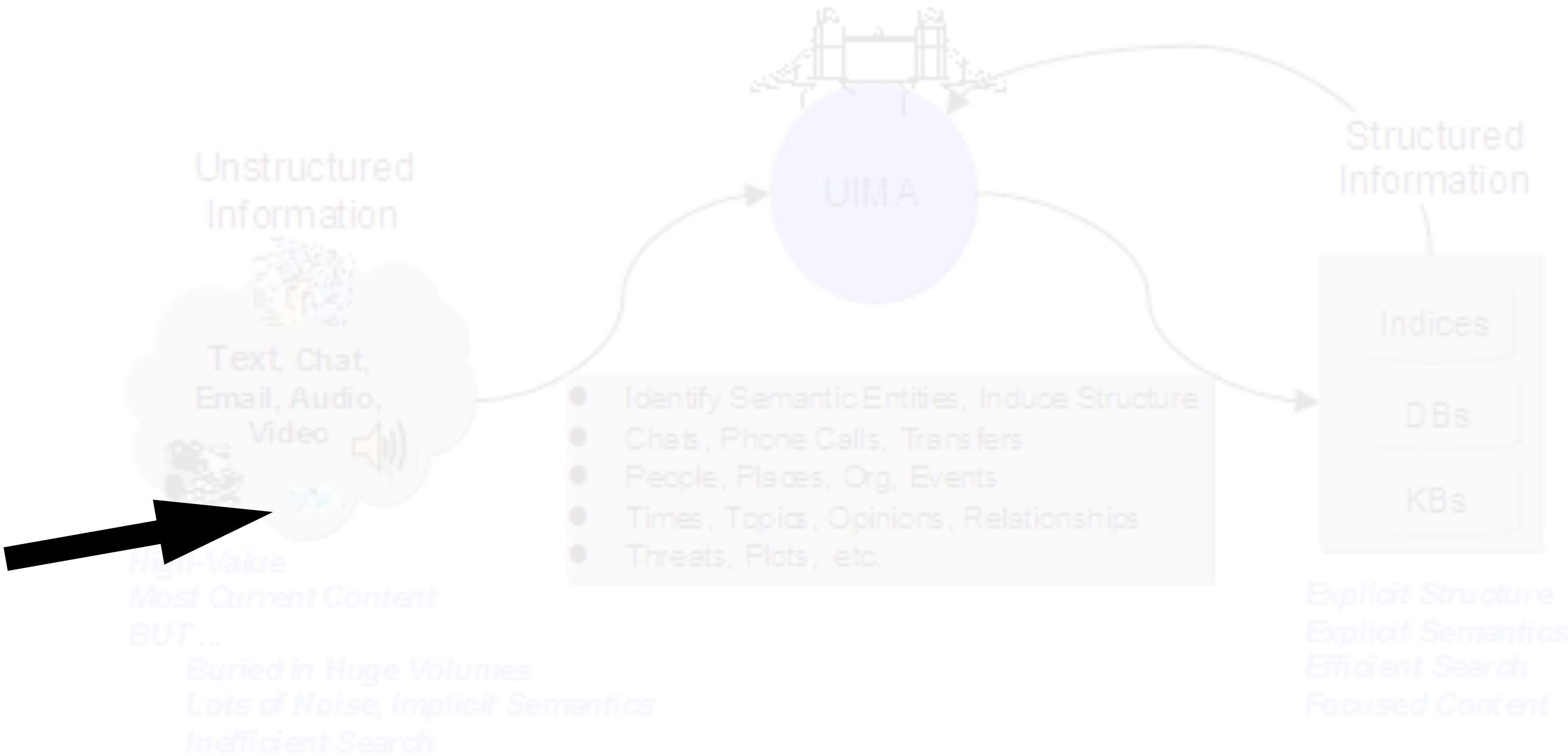
To the degree there was a human face of Watson, the "Jeopardy!" computer champion, it was David Ferrucci. He was the I.B.M. researcher who led the development of Watson, an artificial intelligence engine. The goateed computer scientist was always articulate and at ease in front of a camera or a microphone.

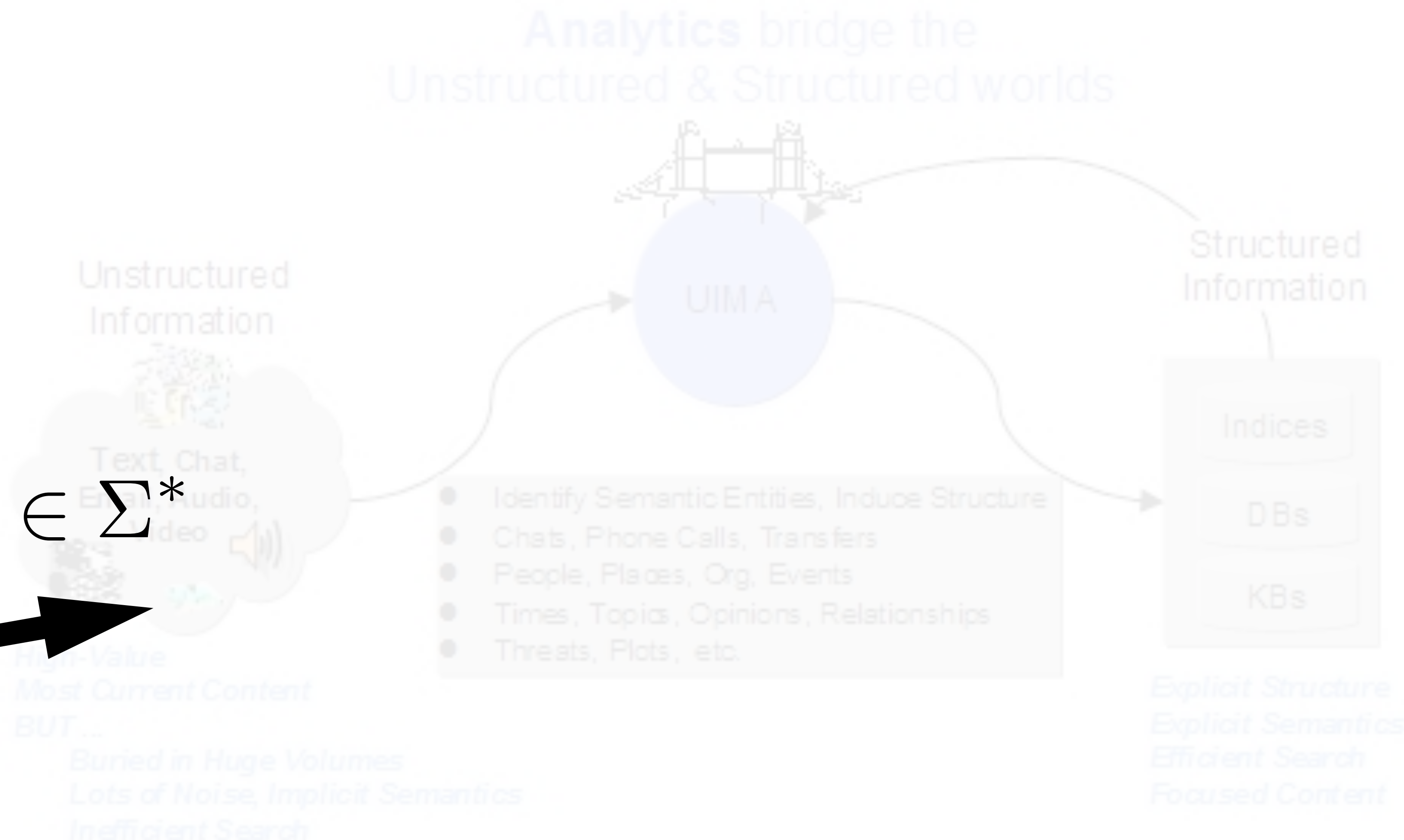
Dr. Ferrucci has left I.B.M. to join the giant hedge fund Bridgewater Associates. And the weight of the Watson-related fame, it seems, played a role. "I was so linked to the Watson achievement, and where I.B.M. was taking it, that I felt I was almost losing my identity," he said in a recent interview.


Analytics bridge the
Unstructured & Structured worlds
Unstructured
Information
Text, Chat,
Email, Audio,
Video
UIMA
Structured
Information
Indices
DBs
KBs
Identify Semantic Entities, Induce Structure
Chats, Phone Calls, Transfers
People, Places, Org, Events
Times, Topics, Opinions, Relationships
Threats, Plots, etc.
High-Value
Most Current Content
BUT ...
Buried in Huge Volumes
Lots of Noise, Implicit Semantics
Inefficient Search
Explicit Structure
Explicit Semantics
Efficient Search
Focused Content

Analytics bridge the Unstructured & Structured worlds

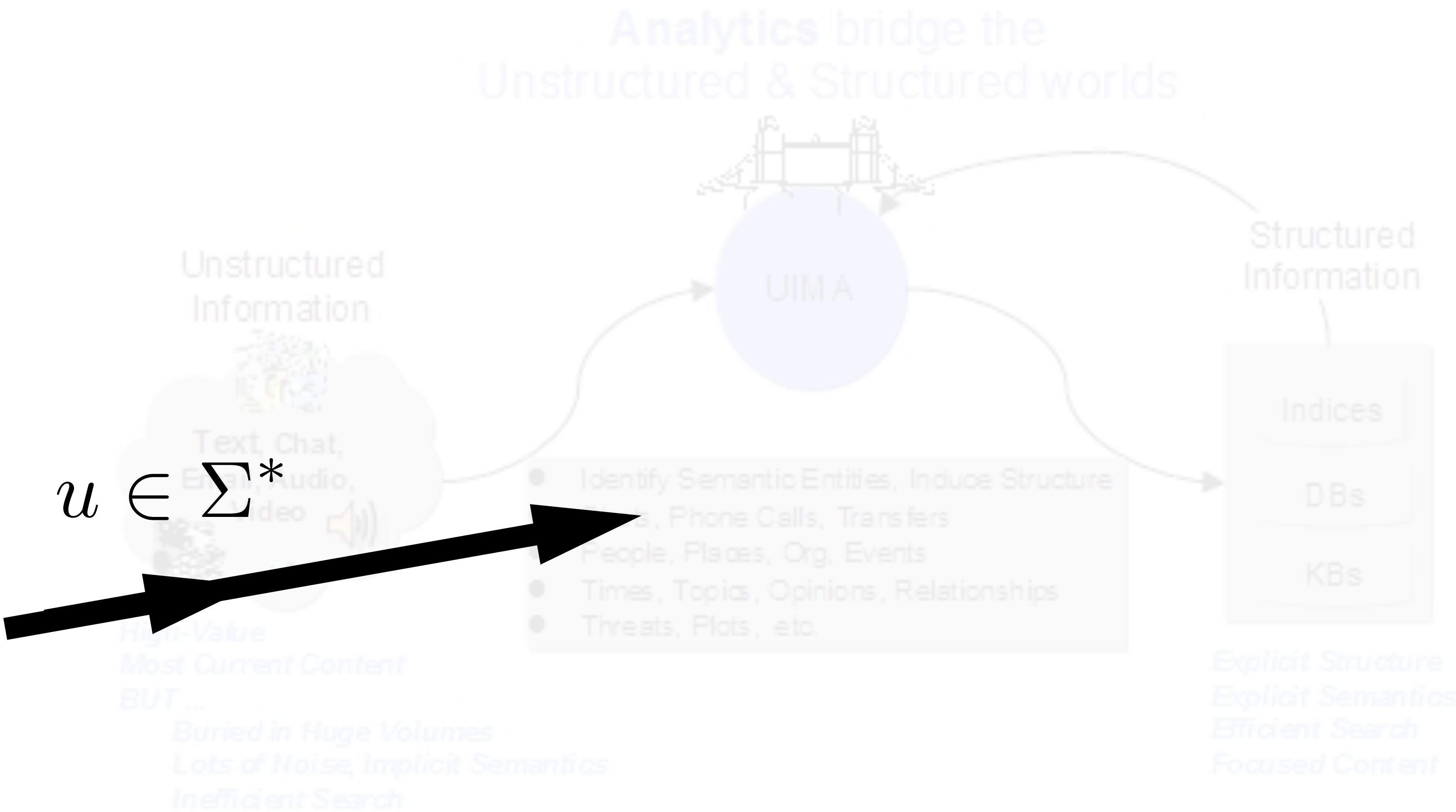
Analytics bridge the Unstructured & Structured worlds



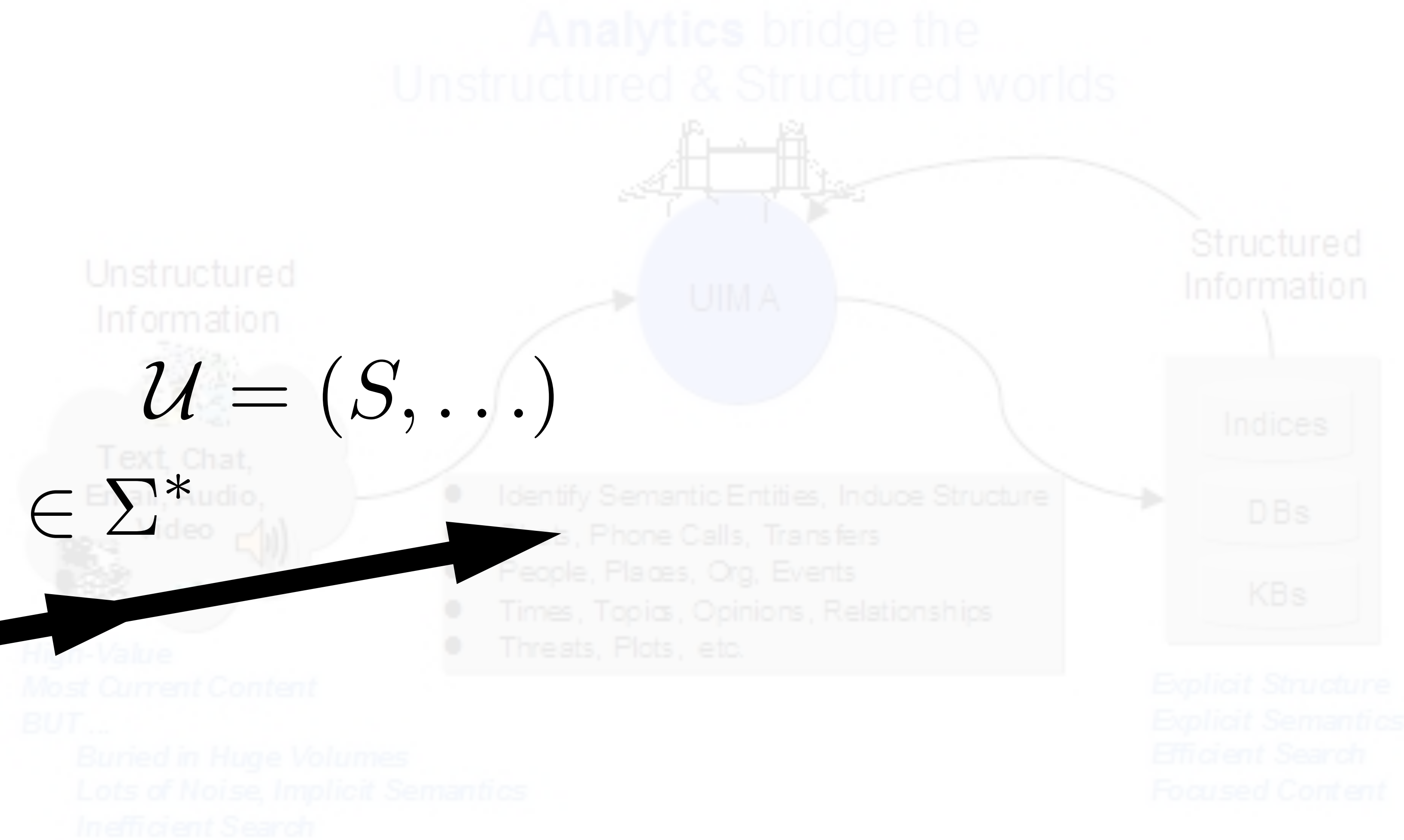


Analytics bridge the
Unstructured & Structured worlds
Unstructured Information
Text, Chat, Email, Audio, Video
UIMA
Structured Information
Indices
DBs
KBs
Identify Semantic Entities, Induce Structure
Chats, Phone Calls, Transfers
People, Places, Org, Events
Times, Topics, Opinions, Relationships
Threats, Plots, etc.
High-Value
Most Current Content
BUT…
Buried in Huge Volumes
Lots of Noise, Implicit Semantics
Inefficient Search
Explicit Structure
Explicit Semantics
Efficient Search
Focused Content

$u \in \Sigma^*$


Analytics bridge the
Unstructured & Structured worlds
Unstructured Information
UIMA
Structured Information
Text, Chat,
Indices
DBs
KBs
Identify Semantic Entities, Induce Structure
People, Places, Org, Events
Times, Topics, Opinions, Relationships
Threats, Plots, etc.
Most Current Content
Buried in Huge Volumes
Lots of Noise, Implicit Semantics
Inefficient Search
Explicit Structure
Explicit Semantics
Efficient Search
Focused Content

$u \in \Sigma^*$



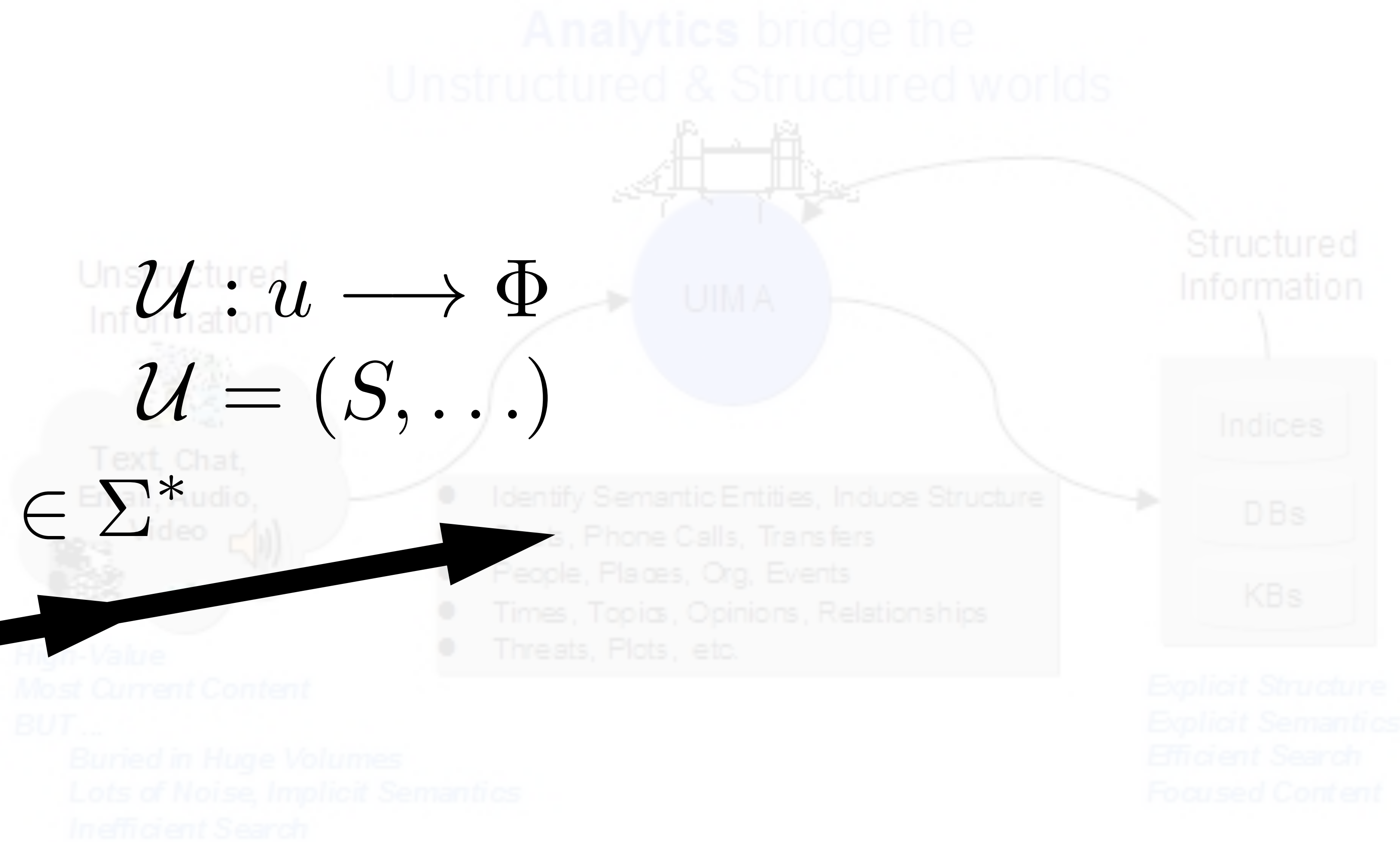
$$\mathcal{U} = (S, \ldots)$$

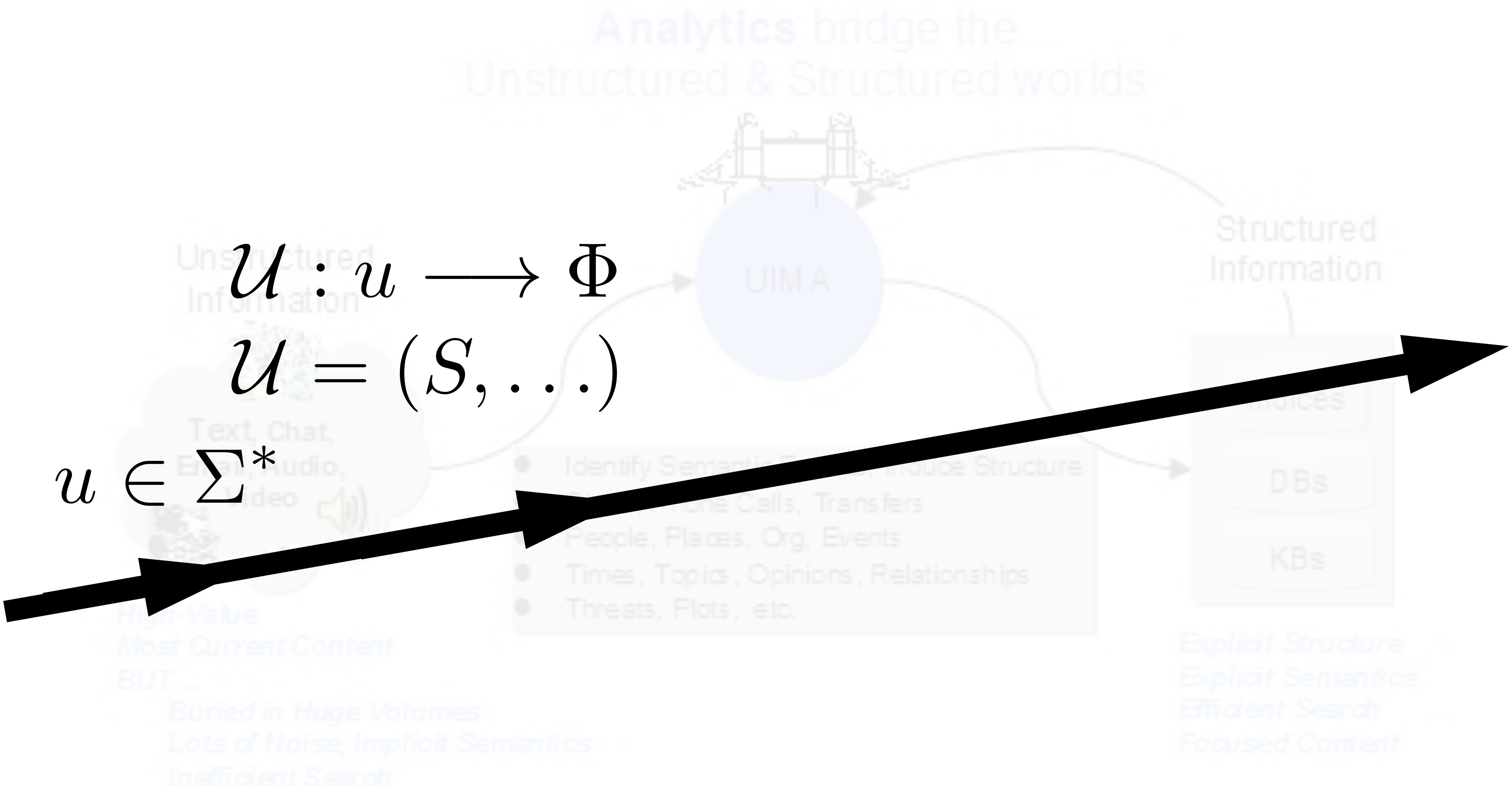
$$u \in \Sigma^*$$

$$\mathcal{U} : u \longrightarrow \Phi$$

$$\mathcal{U} = (S, \ldots)$$

$$u \in \Sigma^*$$

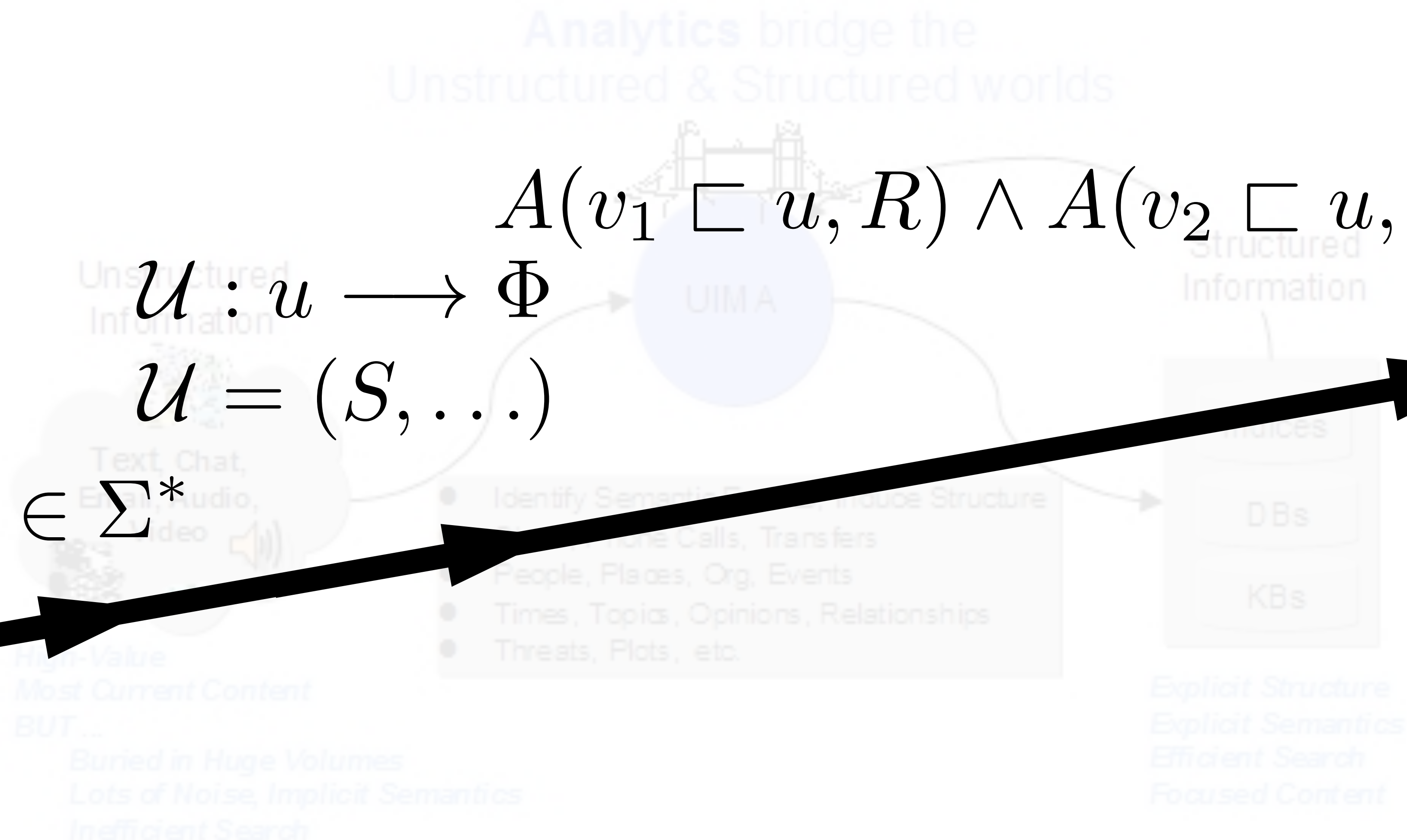




$$\mathcal{U} : u \longrightarrow \Phi$$

$$\mathcal{U} = (S, \ldots)$$

$$u \in \Sigma^*$$



$$A(v_1 \sqsubset u, R) \wedge A(v_2 \sqsubset u, R)$$

$$\mathcal{U} : u \longrightarrow \Phi$$

$$\mathcal{U} = (S, \ldots)$$

$$u \in \Sigma^*$$

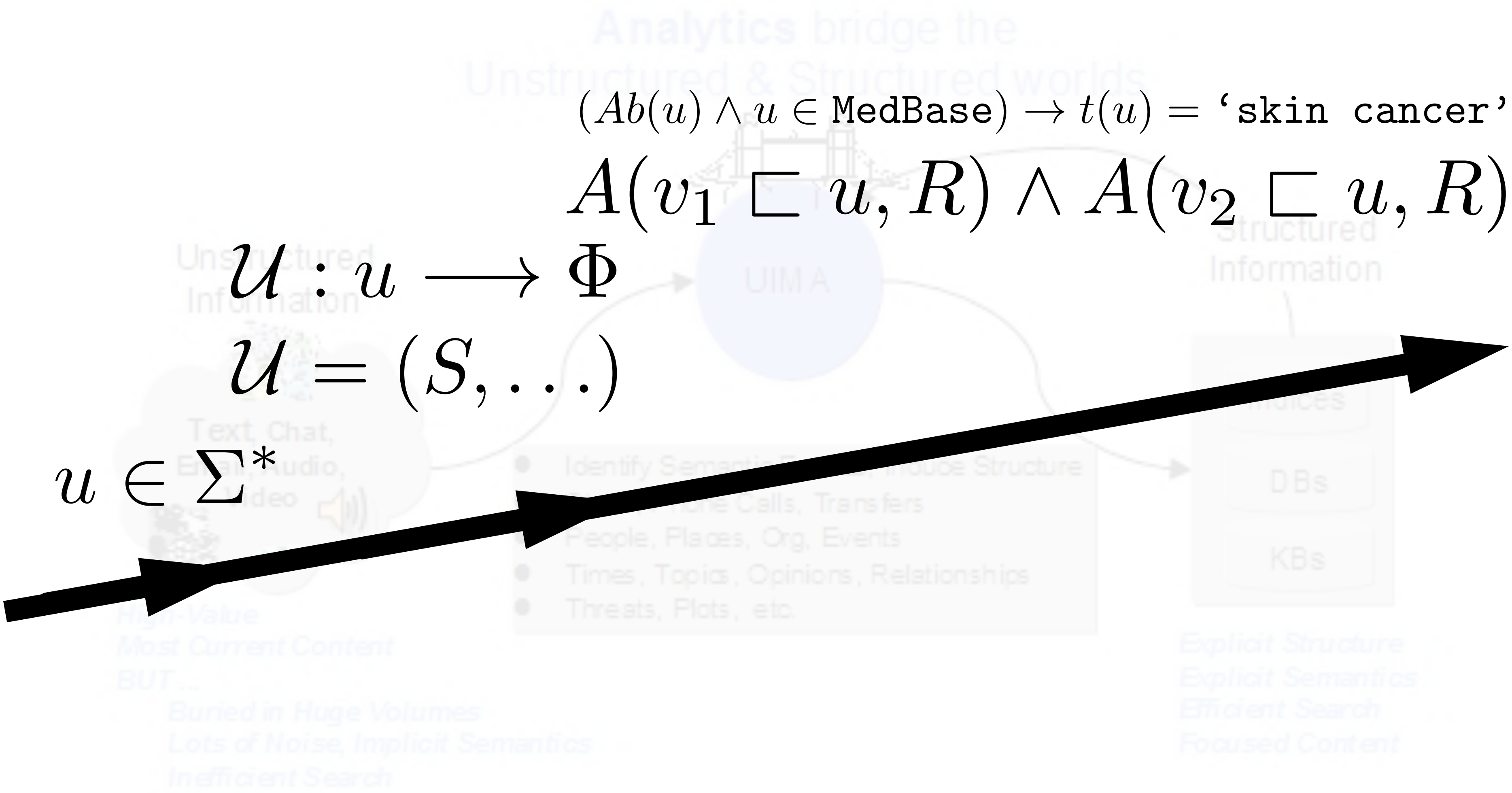
$$(Ab(u) \wedge u \in \texttt{MedBase}) \rightarrow t(u) = \text{‘}\texttt{skin cancer}\text{’}$$

$$A(v_1 \sqsubset u, R) \wedge A(v_2 \sqsubset u, R)$$

$$\mathcal{U} : u \longrightarrow \Phi$$

$$\mathcal{U} = (S, \ldots)$$

$$u \in \Sigma^*$$



$

$(Ab(u) \wedge u \in \texttt{MedBase}) \rightarrow t(u) = \text{`}\texttt{skin cancer}\text{'}$

$$A(v_1 \sqsubset u, R) \wedge A(v_2 \sqsubset u, R)$$

$$\mathcal{U} : u \longrightarrow \Phi$$

$$\mathcal{U} = (S, \ldots)$$

$$u \in \Sigma^*$$

Many thanks to *A* Bringsjord for this pair of slides, and discussion.

Meg Whitman: Autonomy Is An 'Almost Magical Technology'

JULIE BORT | APR. 10, 2013, 6:58 PM | 2,226 | 5

Even in the wake of the resignation of two HP board members last week, CEO Meg Whitman doesn't regret telling the public about Autonomy's problems, she said at a press conference in London this week.

HP

HP CEO Meg Whitman

About a year after acquiring Autonomy for $11 billion, HP wrote off $8.8 billion of the

Autonomy “Magic”

Autonomy “Magic”

Understanding 100% of Information
The characteristic that makes human information unstructured is its form — it does not fit neatly into the rows and columns of a database, but exists in various formats including books, email messages, surveillance video, chat streams, and phone calls that occur across networks, the web, the cloud, and numerous mobile devices. Growing at a rate three times that of structured data, the increasing deluge of unstructured information makes up approximately 90 percent of all information. The challenge for the modern enterprise is to understand and extract value from this rich sea of human information.

Today, with Autonomy, organizations can now process and understand in real time, the meaning of 100 percent of information.

Autonomy “Magic”

Understanding 100% of Information
The characteristic that makes human information unstructured is its form — it does not fit neatly into the rows and columns of a database, but exists in various formats including books, email messages, surveillance video, chat streams, and phone calls that occur across networks, the web, the cloud, and numerous mobile devices. Growing at a rate three times that of structured data, the increasing deluge of unstructured information makes up approximately 90 percent of all information. The challenge for the modern enterprise is to understand and extract value from this rich sea of human information.

Today, with Autonomy, organizations can now process and understand in real time, the meaning of 100 percent of information.

Autonomy “Magic”

Understanding 100% of Information
The characteristic that makes human information unstructured is its form — it does not fit neatly into the rows and columns of a database, but exists in various formats including books, email messages, surveillance video, chat streams, and phone calls that occur across networks, the web, the cloud, and numerous mobile devices. Growing at a rate three times that of structured data, the increasing deluge of unstructured information makes up approximately 90 percent of all information. The challenge for the modern enterprise is to understand and extract value from this rich sea of human information.

Today, with Autonomy, organizations can now process and understand in real time, the meaning of 100 percent of information.

!?!?

Autonomy "Magic"

Understanding 100% of Information
The characteristic that makes human information unstructured is its form — it does not fit neatly into the rows and columns of a database, but exists in various formats including books, email messages, surveillance video, chat streams, and phone calls that occur across networks, the web, the cloud, and numerous mobile devices. Growing at a rate three times that of structured data, the increasing deluge of unstructured information makes up approximately 90 percent of all information. The challenge for the modern enterprise is to understand and extract value from this rich sea of human information.

Today, with Autonomy, organizations can now process and understand in real time, the meaning of 100 percent of information.

!?!?

(They sure as heck never consulted with Floridi.)

Autonomy "Magic"

Understanding 100% of Information
The characteristic that makes human information unstructured is its form — it does not fit neatly into the rows and columns of a database, but exists in various formats including books, email messages, surveillance video, chat streams, and phone calls that occur across networks, the web, the cloud, and numerous mobile devices. Growing at a rate three times that of structured data, the increasing deluge of unstructured information makes up approximately 90 percent of all information. The challenge for the modern enterprise is to understand and extract value from this rich sea of human information.

Today, with Autonomy, organizations can now process and understand in real time, the meaning of 100 percent of information.

!?!?

(They sure as heck never consulted with Floridi.)

The little **O**UP book that could've saved HP $6 billion:

Autonomy “Magic”

Understanding 100% of Information
The characteristic that makes human information unstructured is its form — it does not fit neatly into the rows and columns of a database, but exists in various formats including books, email messages, surveillance video, chat streams, and phone calls that occur across networks, the web, the cloud, and numerous mobile devices. Growing at a rate three times that of structured data, the increasing deluge of unstructured information makes up approximately 90 percent of all information. The challenge for the modern enterprise is to understand and extract value from this rich sea of human information.

Today, with Autonomy, organizations can now process and understand in real time, the meaning of 100 percent of information.

!?!?

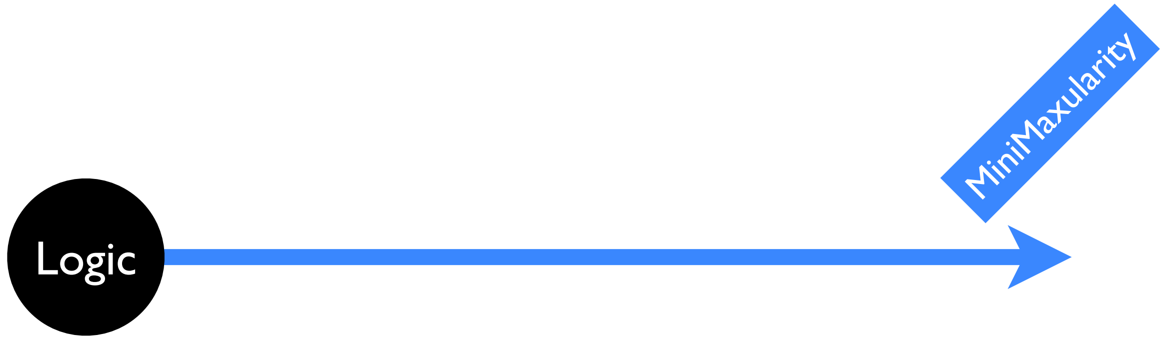
(They sure as heck never consulted with Floridi.)

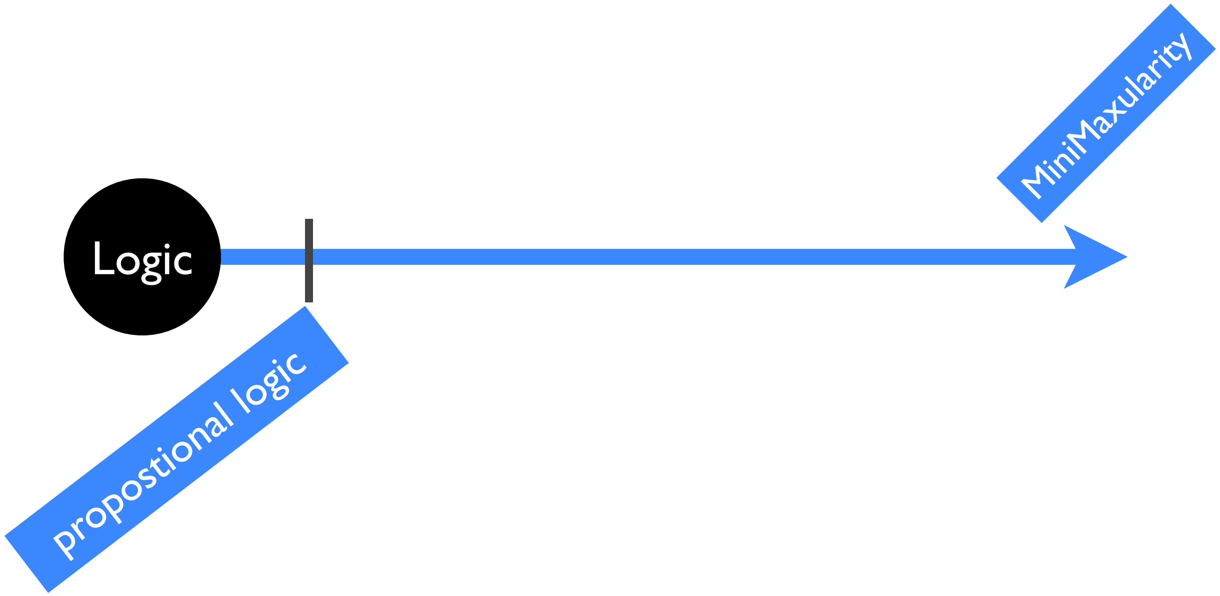
The little OUP book that could’ve saved HP $6 billion:

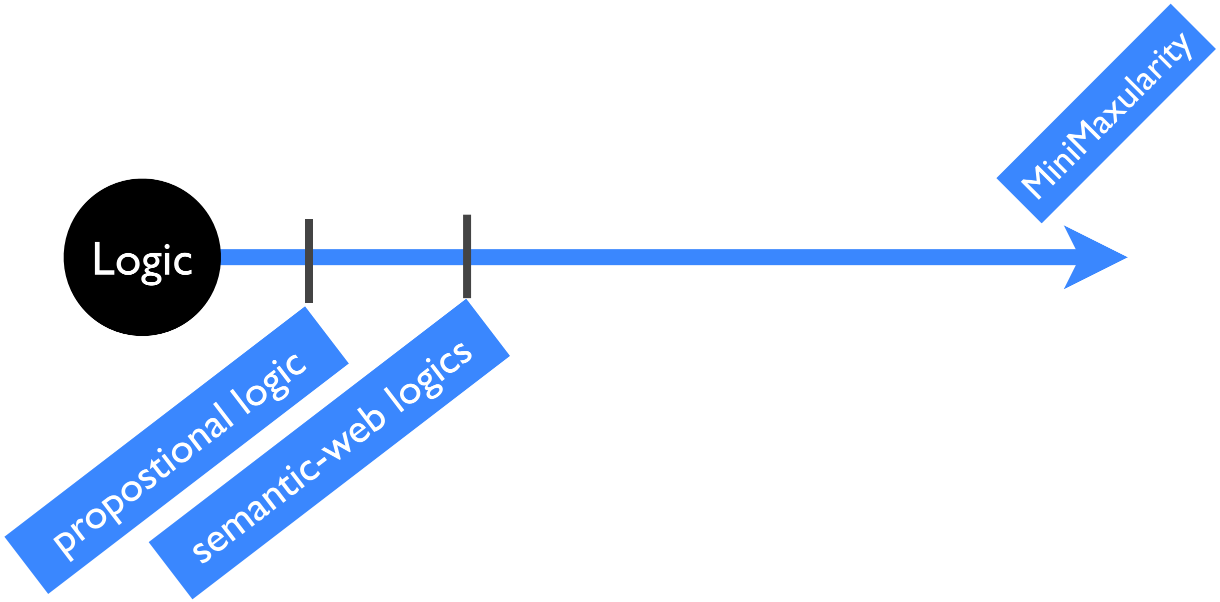
Let’s *in fact* talk about 100% of semantic information, using *Watson 2.0* as a springboard ...

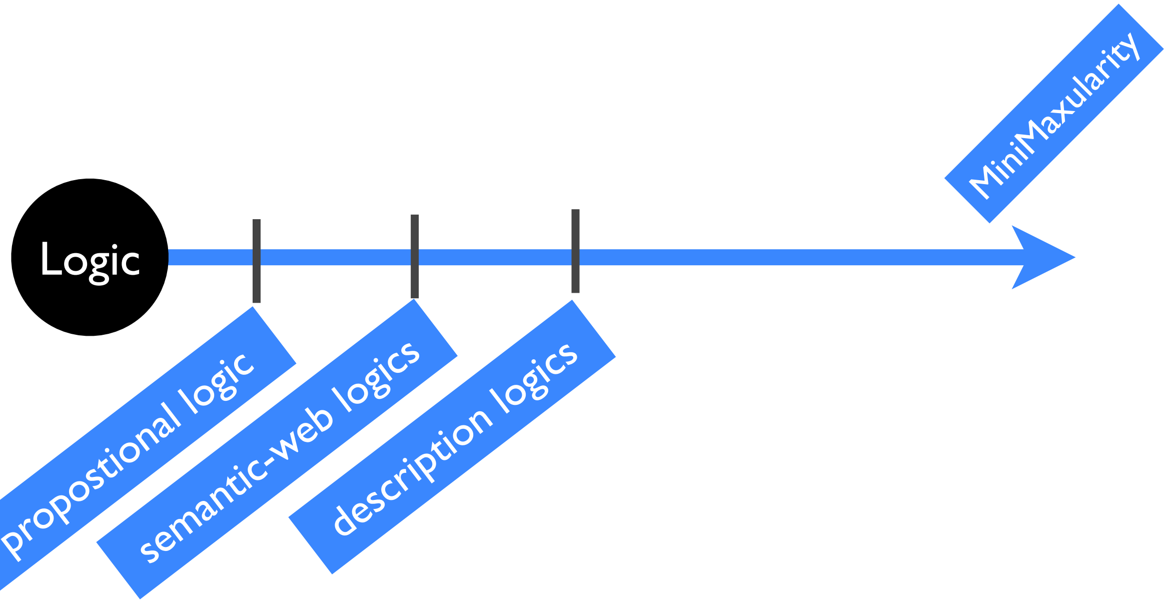
Logic

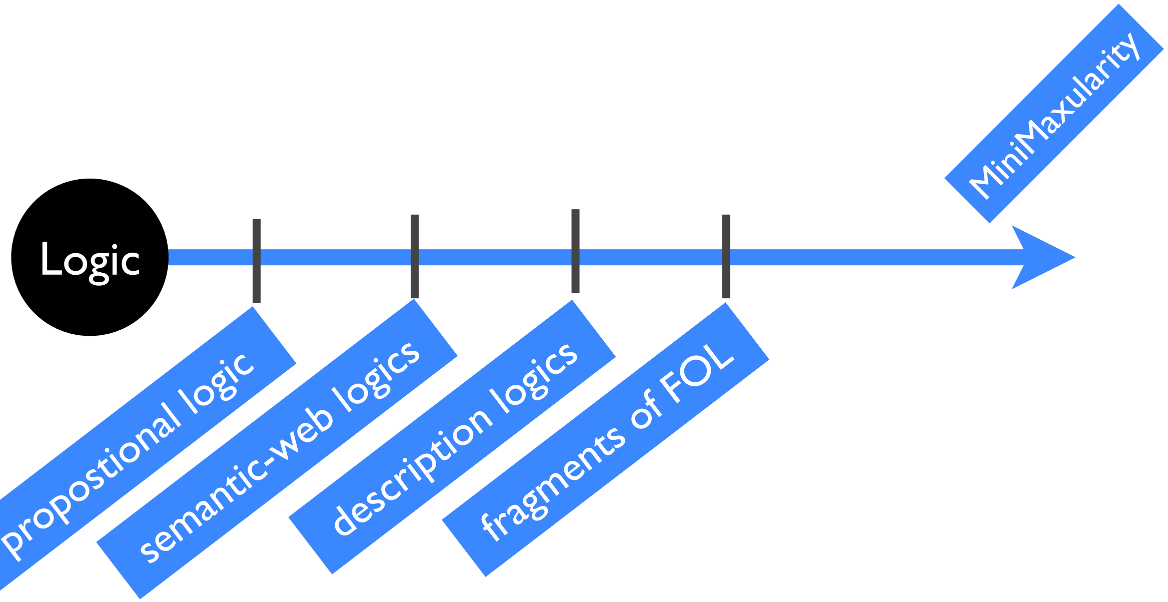

Logic

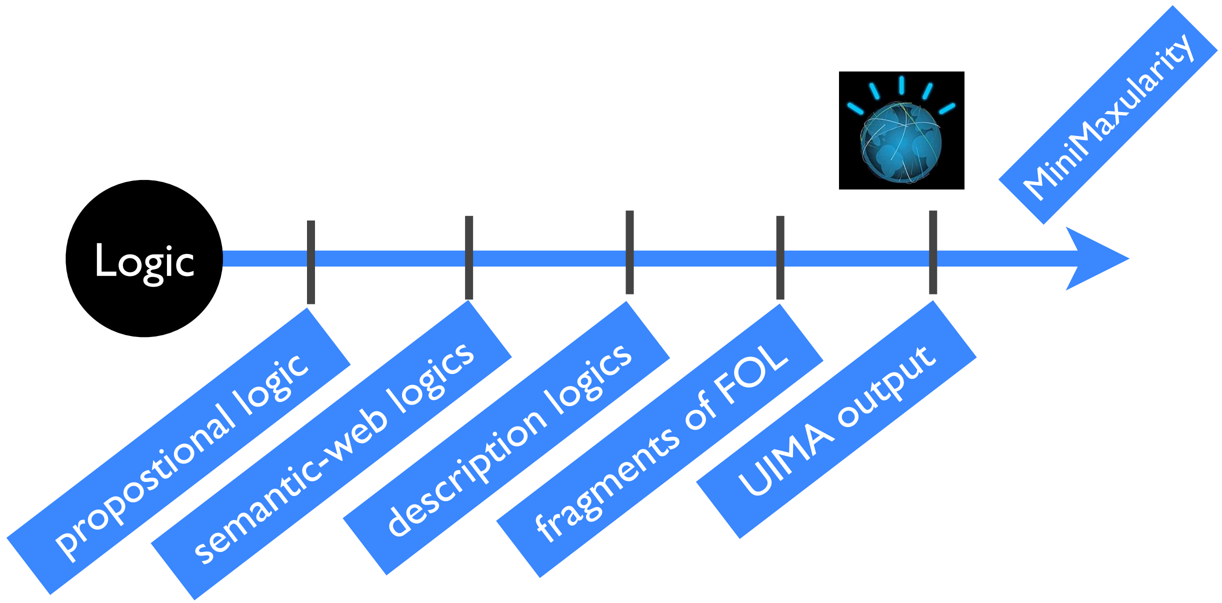

Logic
MiniMaxularity

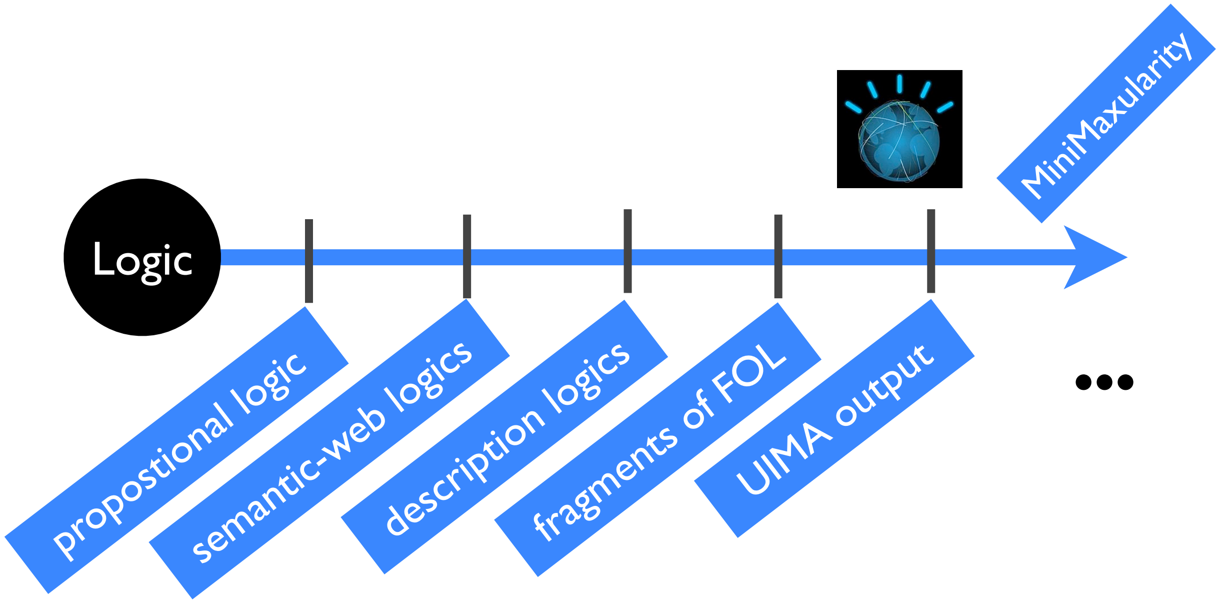

Logic
propostional logic
MiniMaxularity

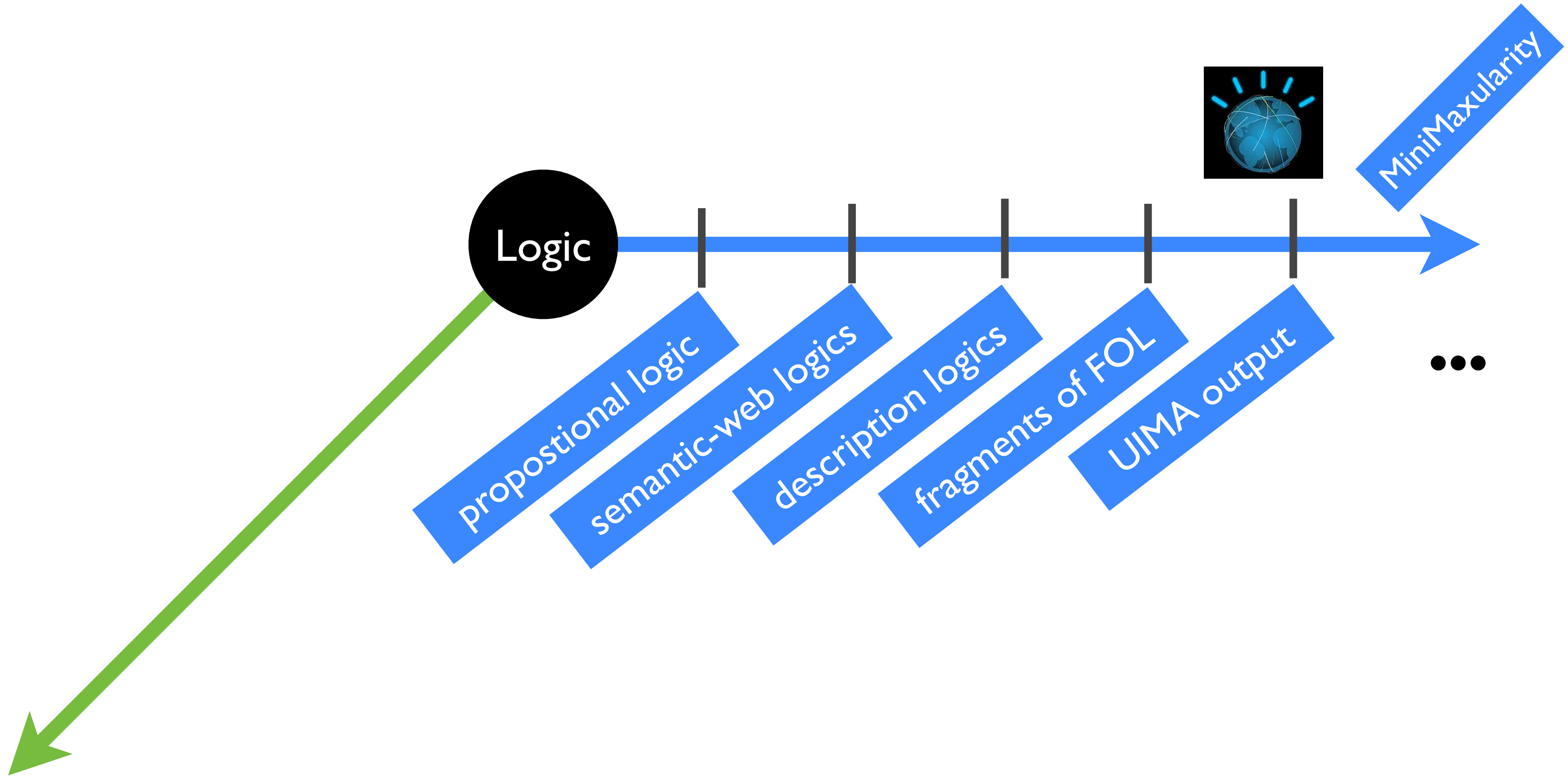

Logic
propostional logic
semantic-web logics
MiniMaxularity

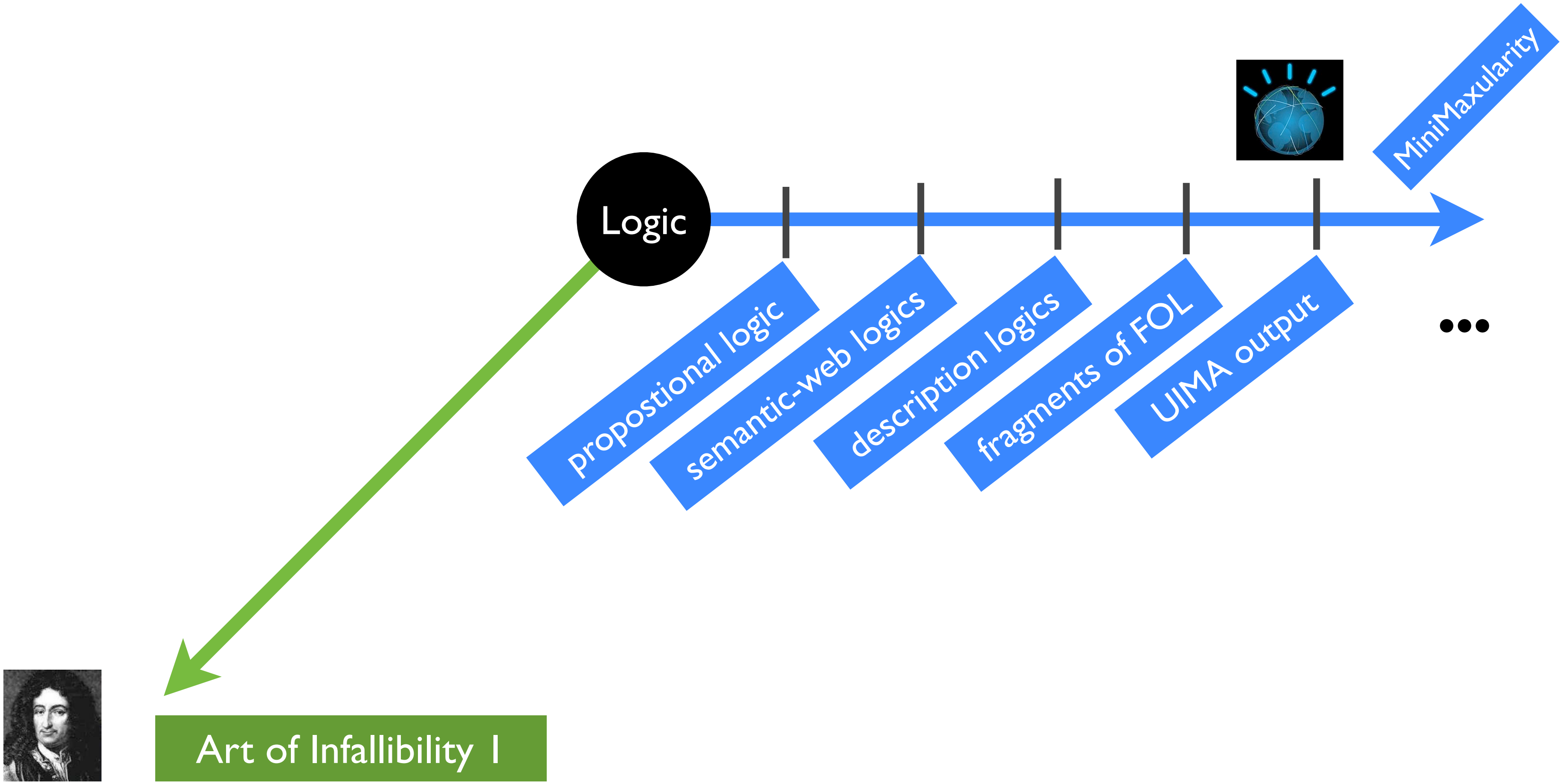

Logic
propostional logic
semantic-web logics
description logics
MiniMaxularity

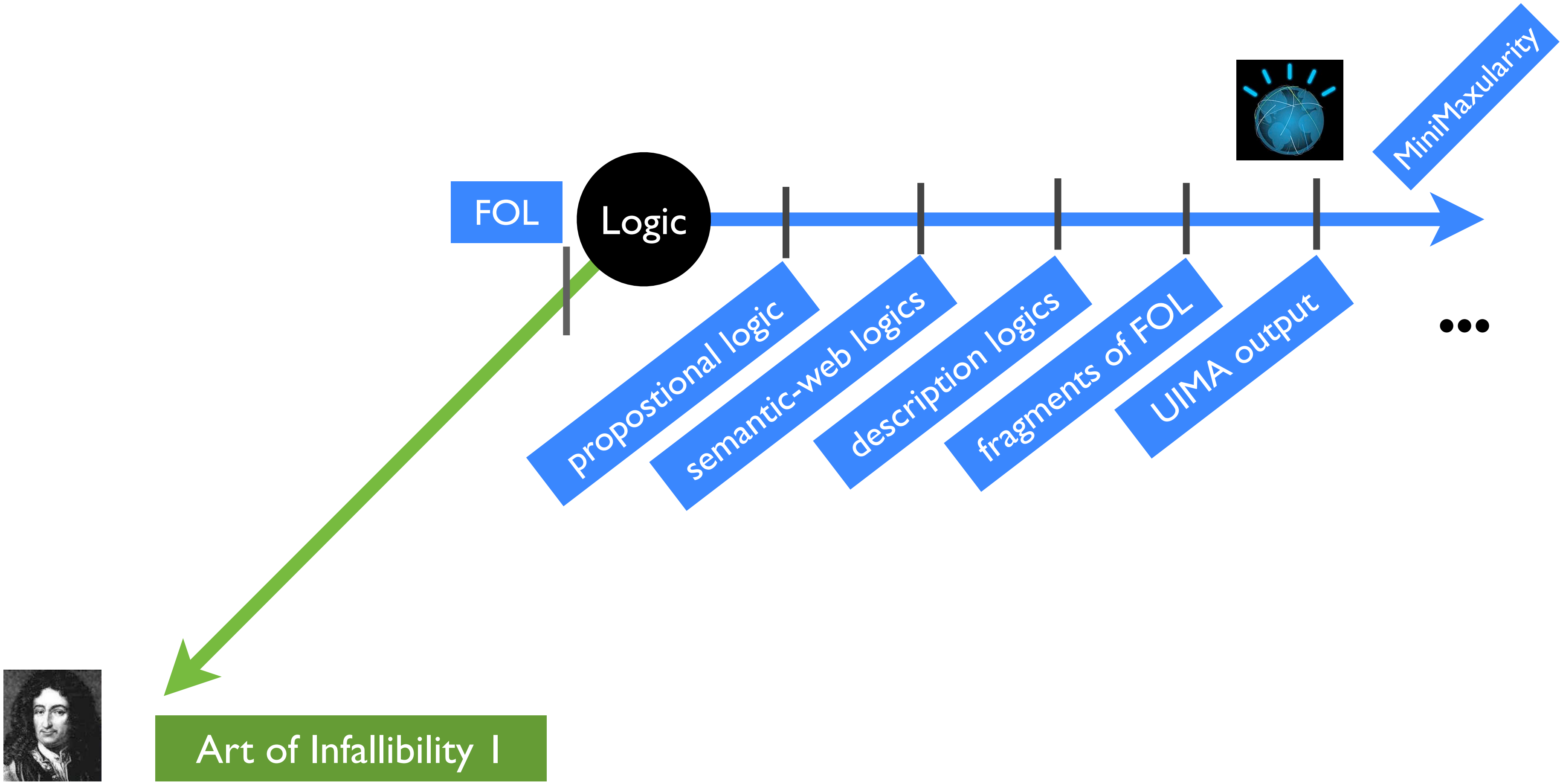

Logic
propostional logic
semantic-web logics
description logics
fragments of FOL
MiniMaxularity

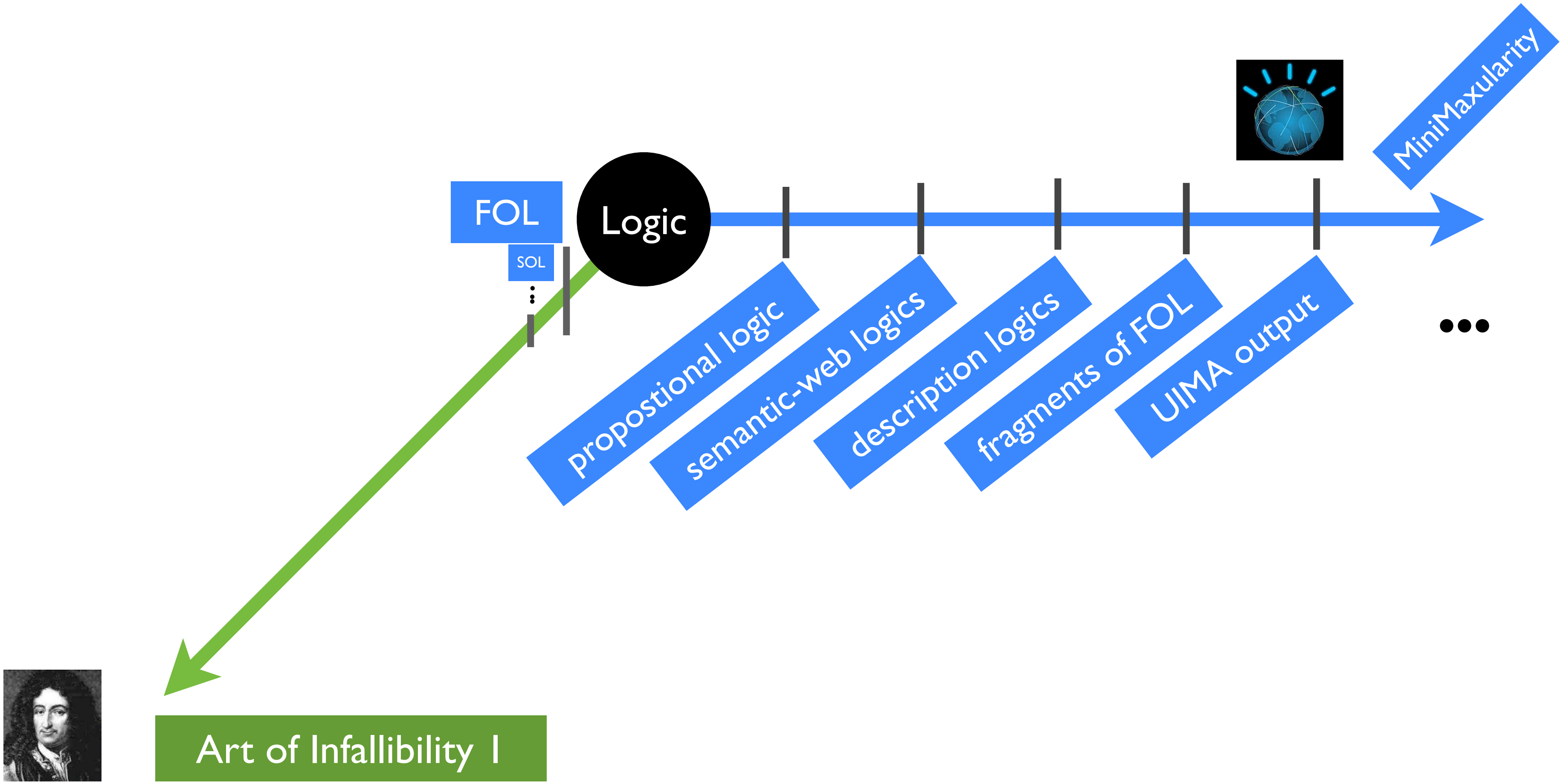

Logic
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity

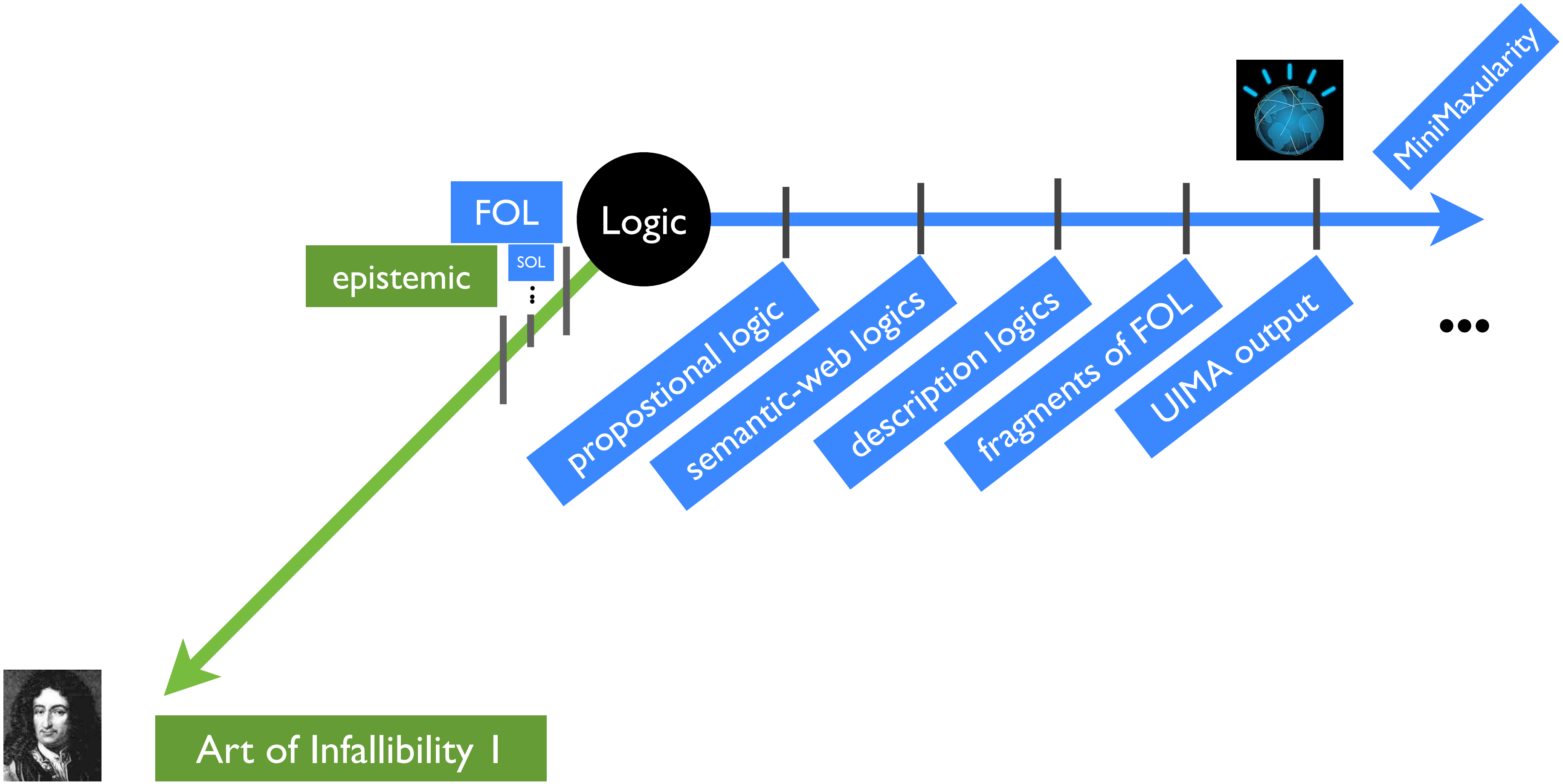

MiniMaxularity
Logic
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
...

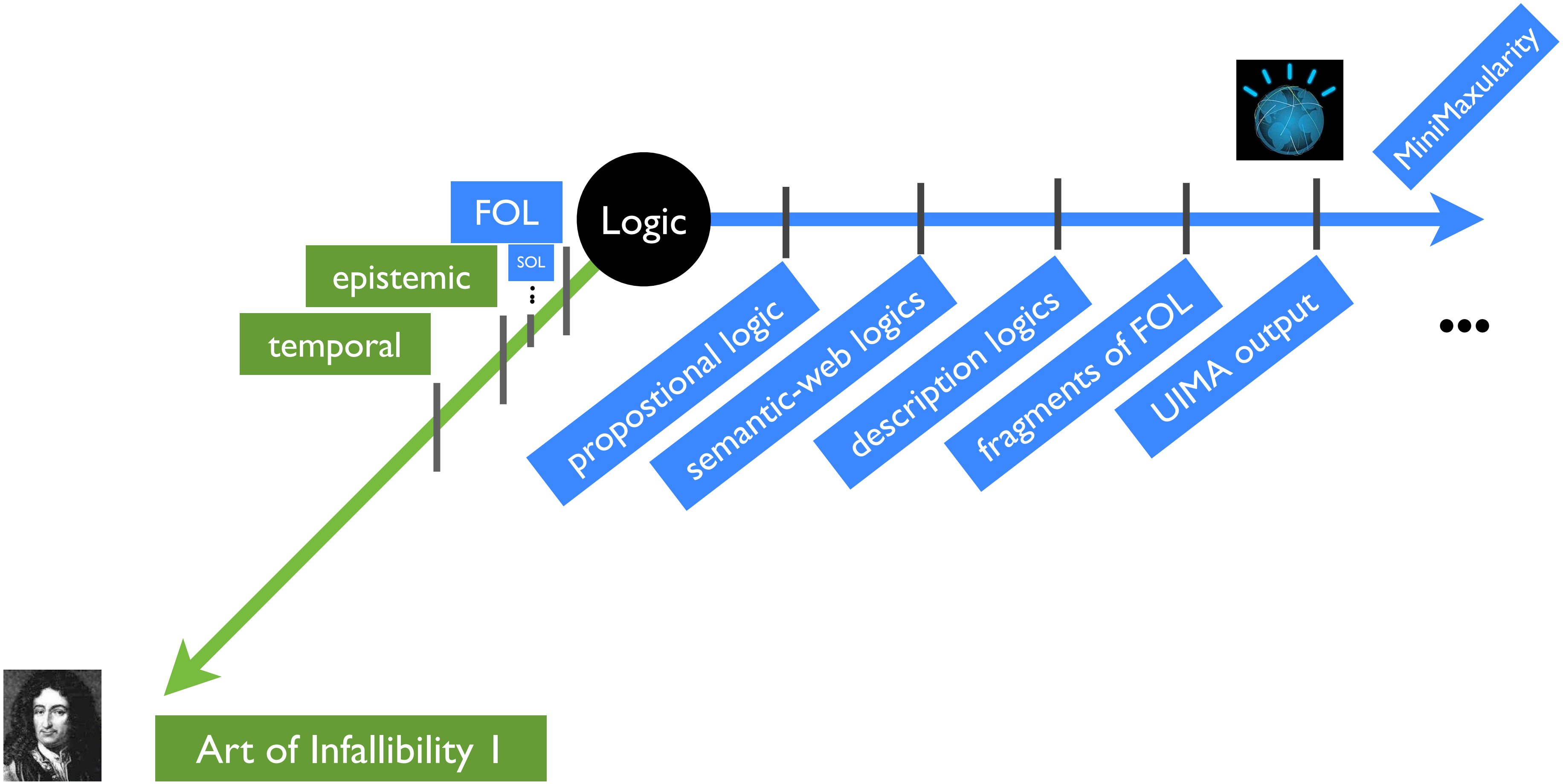

Logic
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity
...

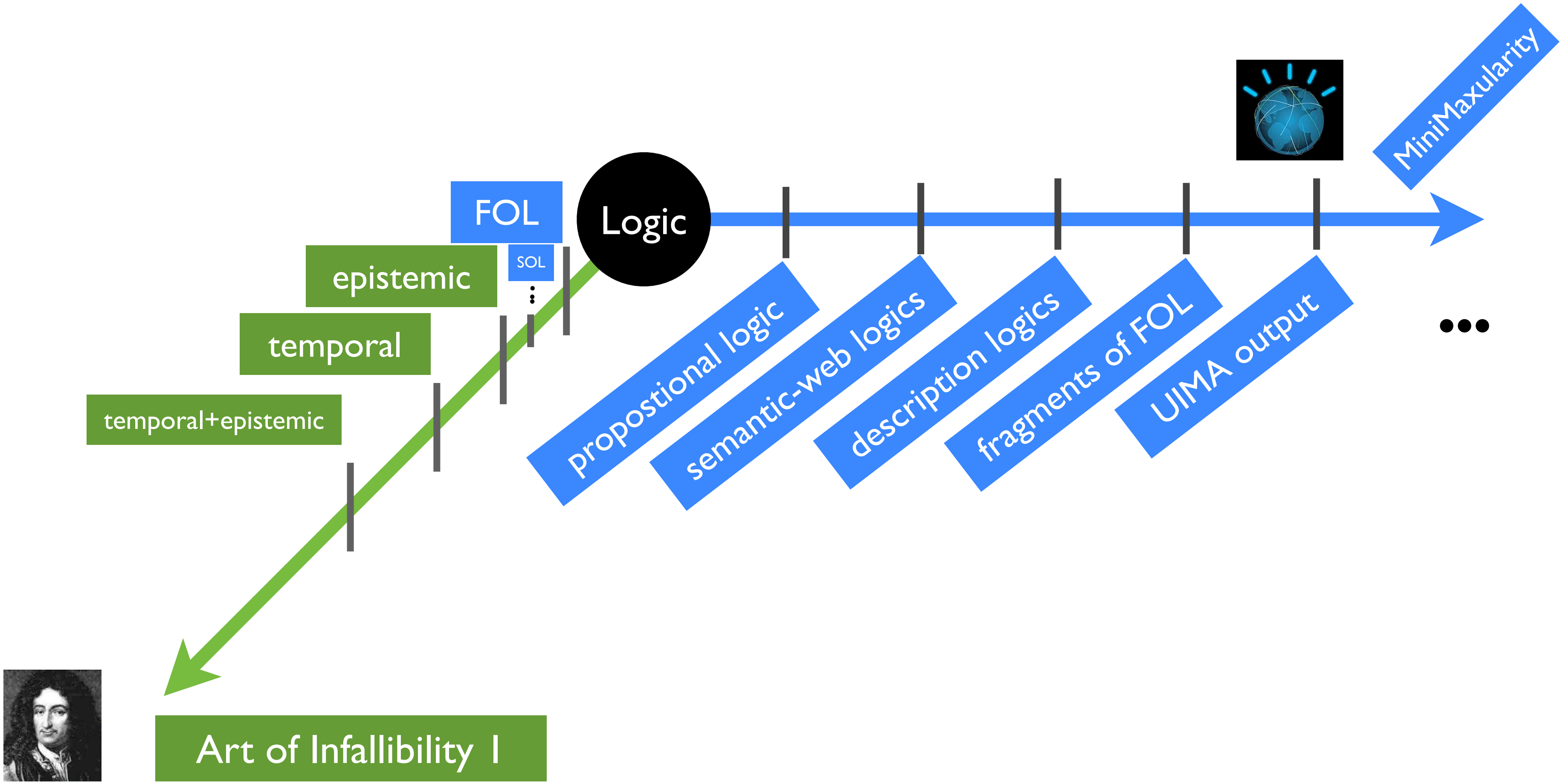

Logic
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity
...
Art of Infallibility I


FOL
Logic
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity
...
Art of Infallibility I

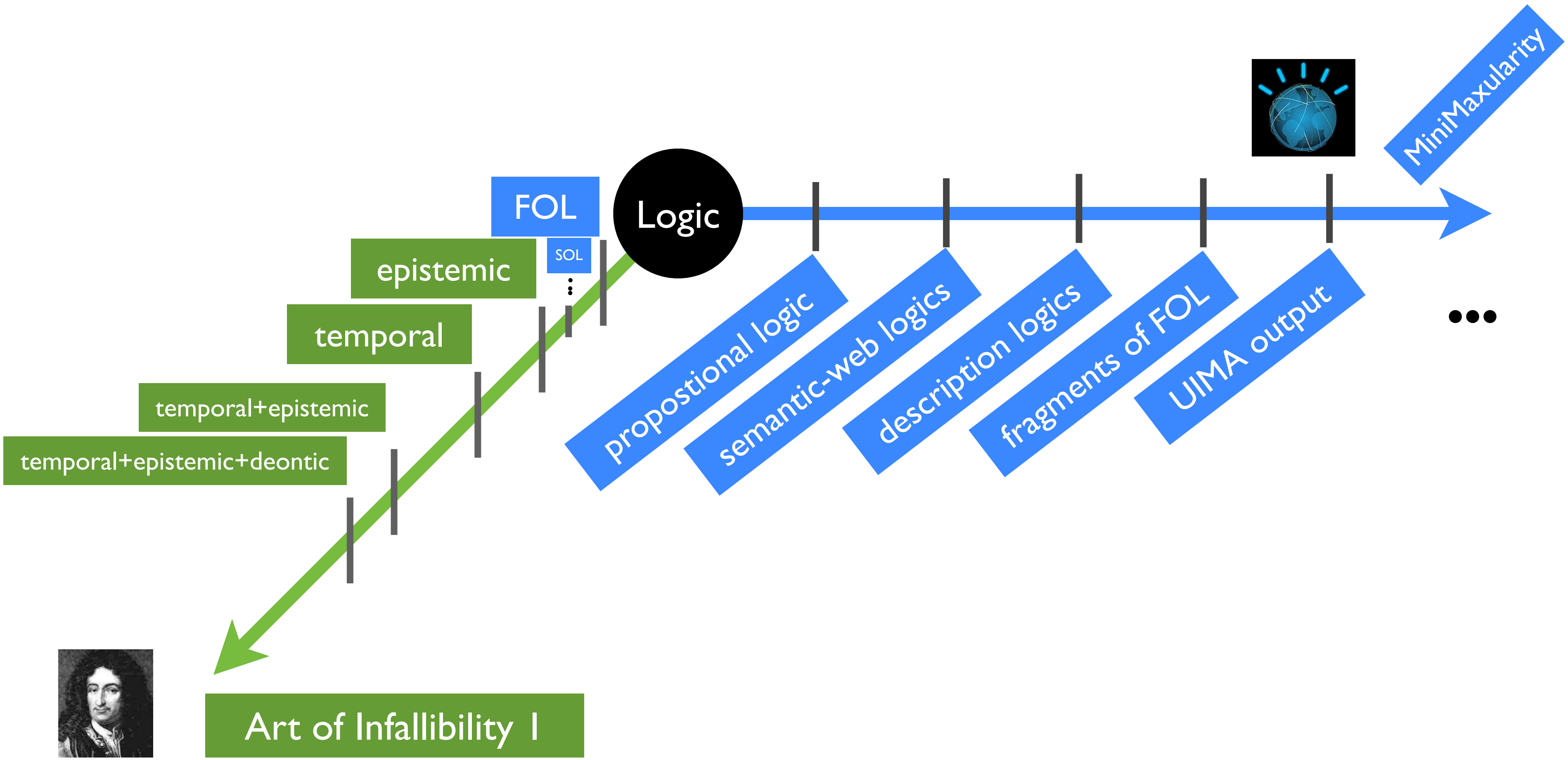

MiniMaxularity
FOL
Logic
SOL
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
•••
Art of Infallibility I

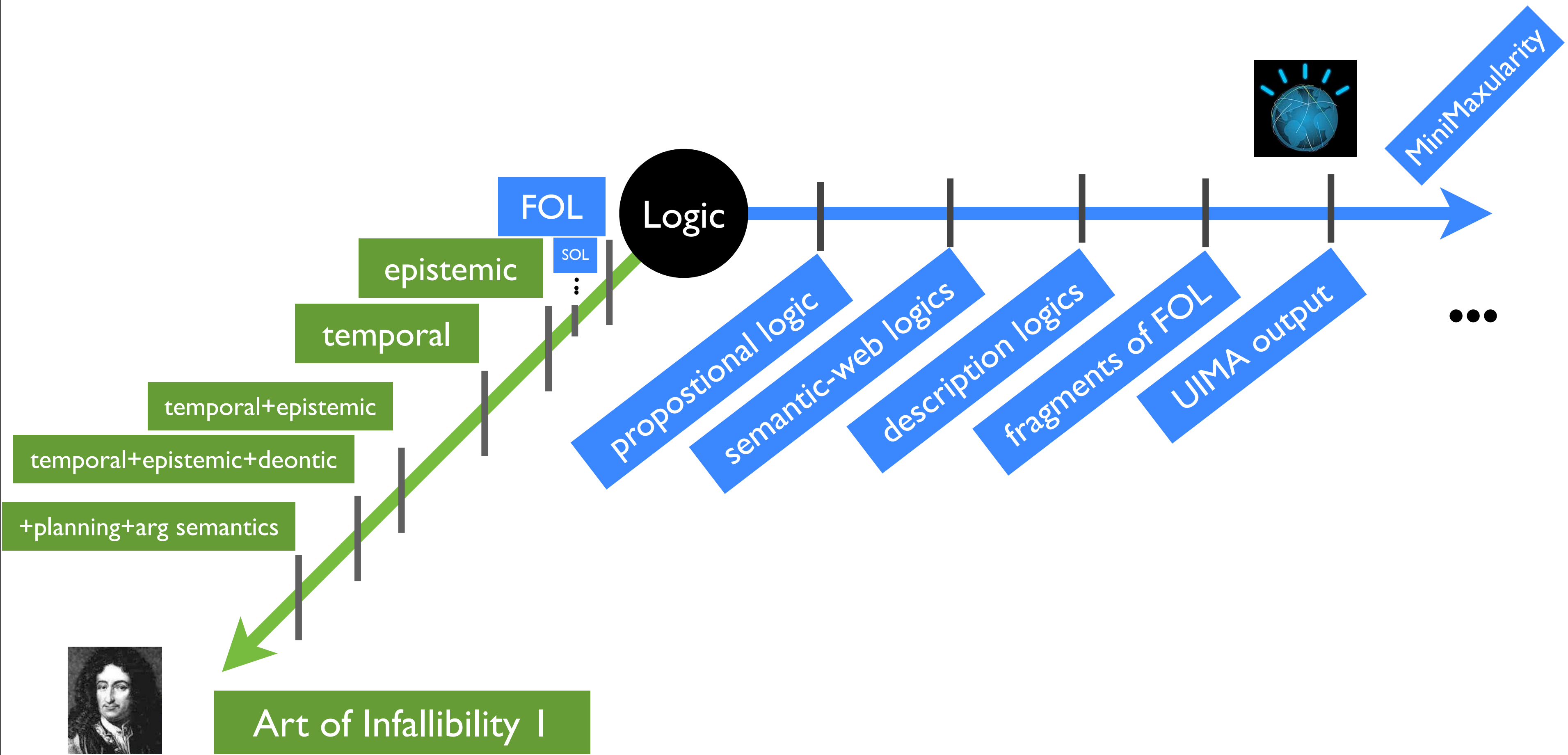

FOL
Logic
epistemic
SOL
⋮
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity
•••
Art of Infallibility I

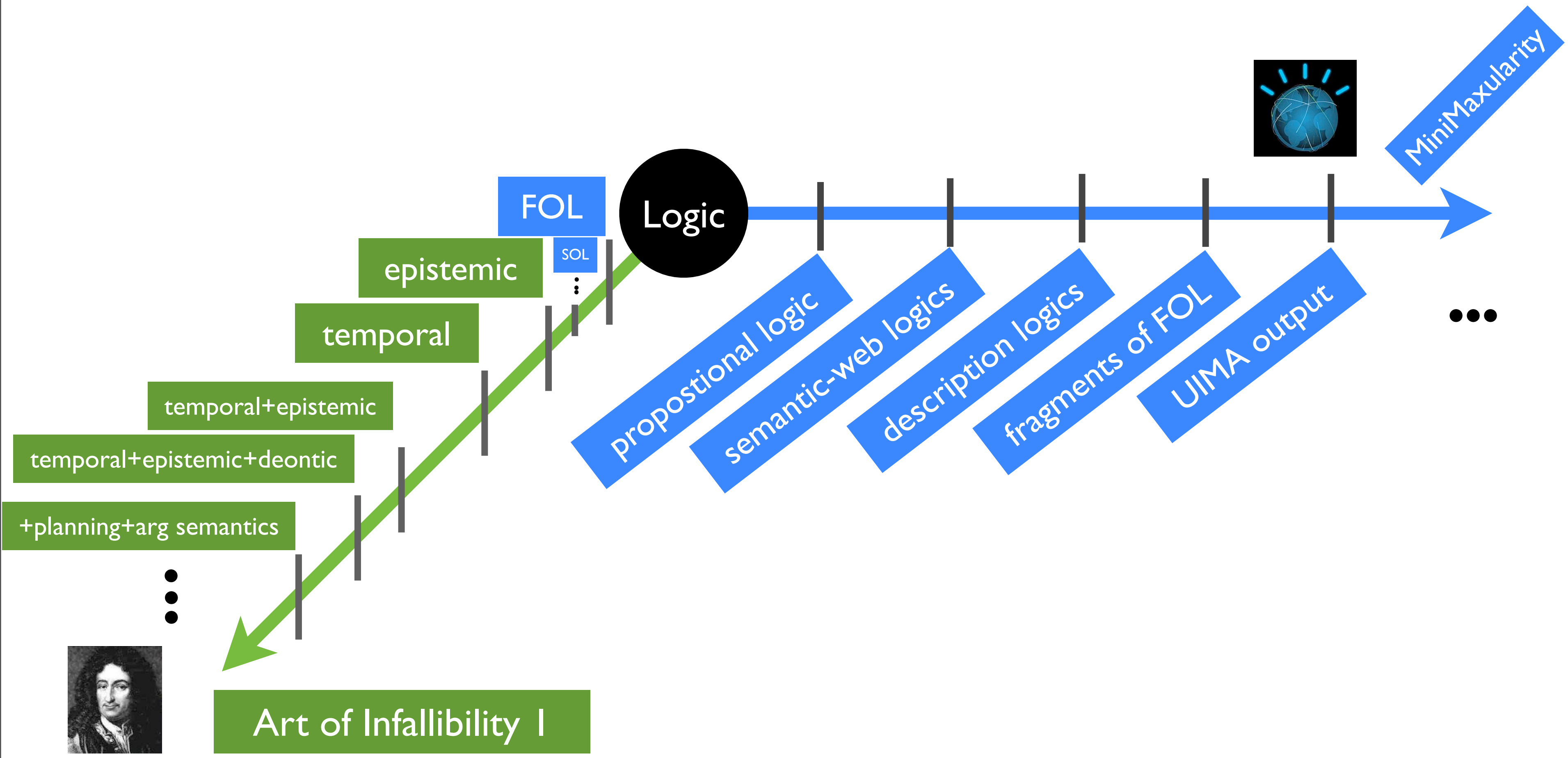

FOL
SOL
epistemic
temporal
Logic
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity
Art of Infallibility I

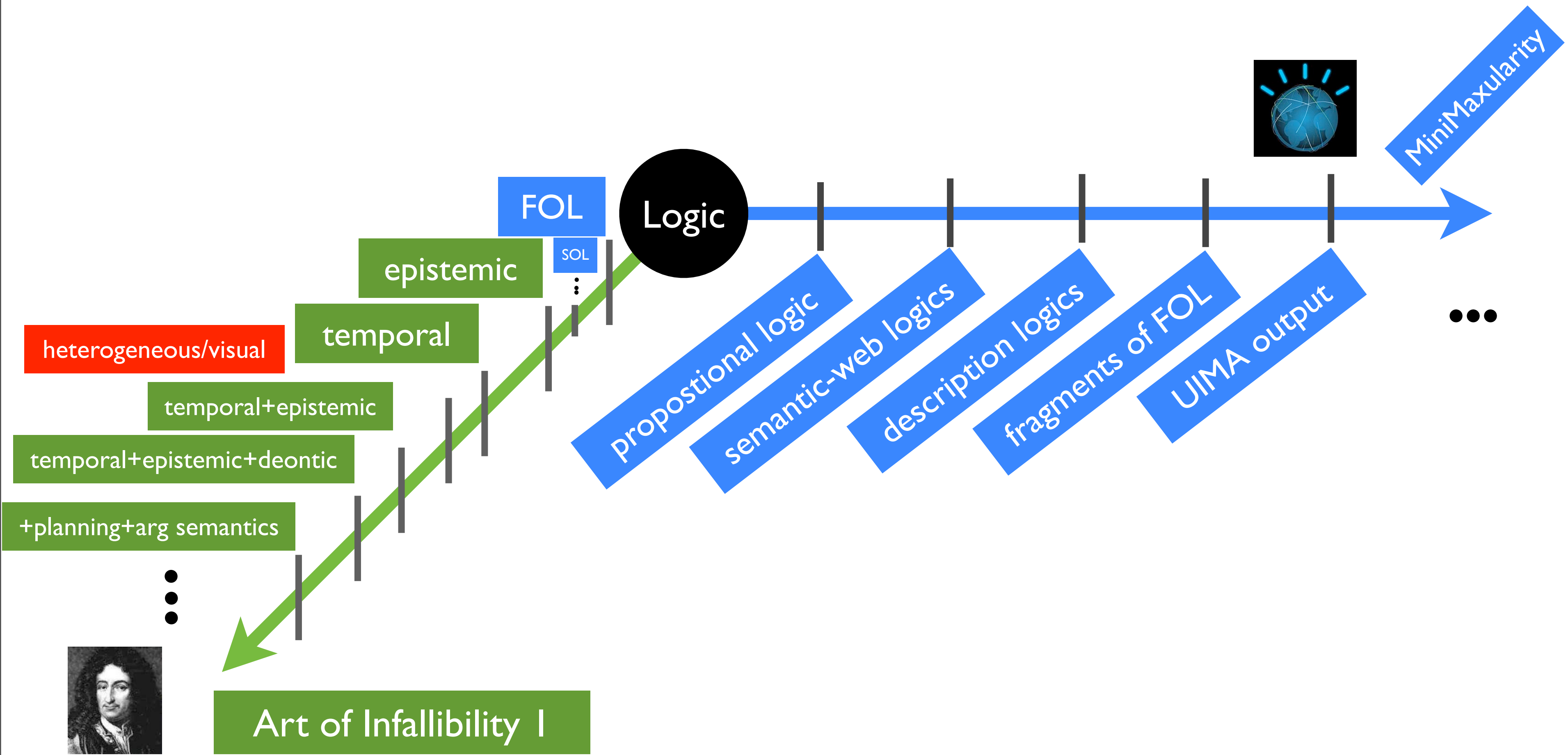


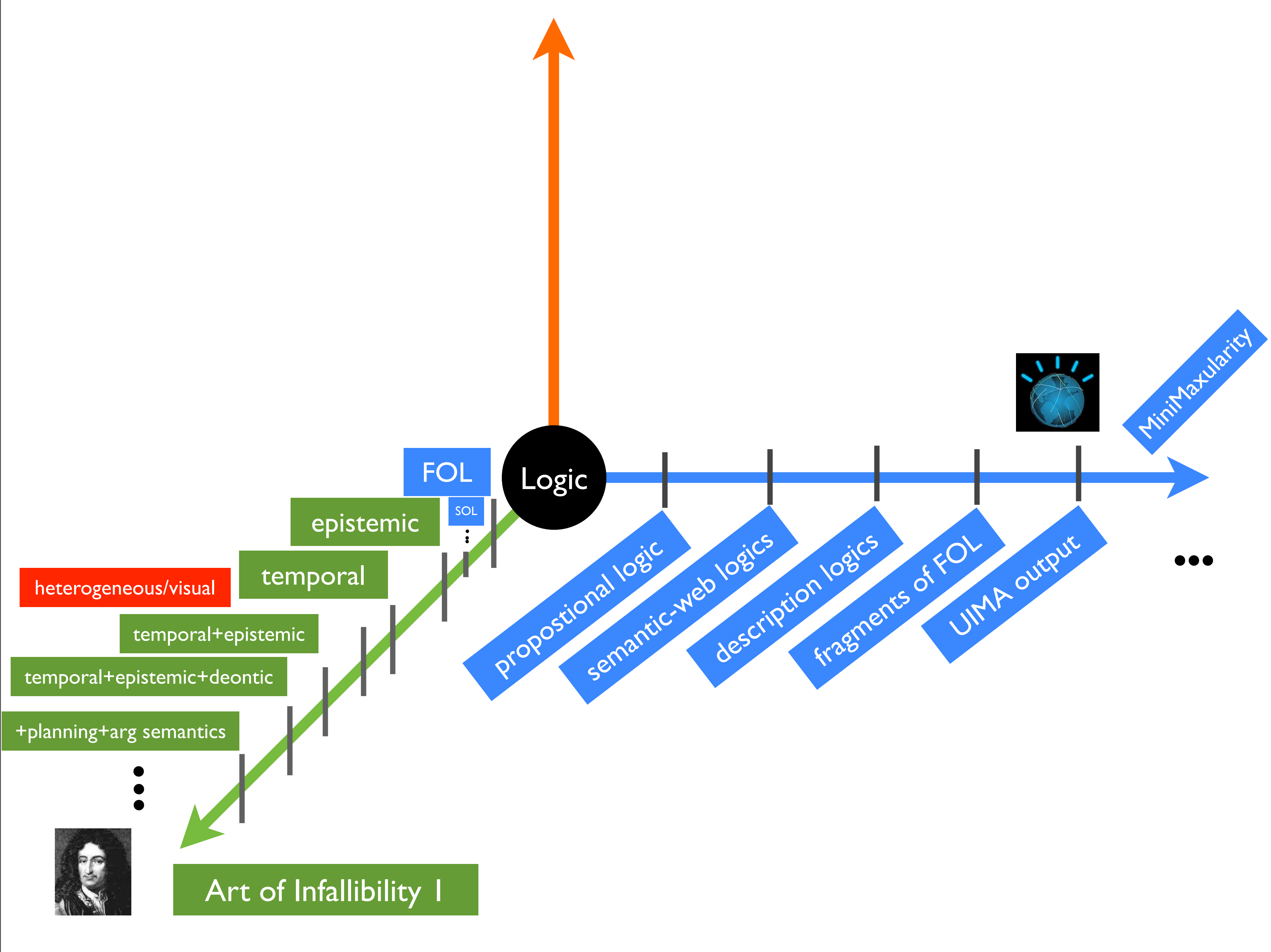
FOL
Logic
SOL
epistemic
temporal
temporal+epistemic
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity
•••
Art of Infallibility I

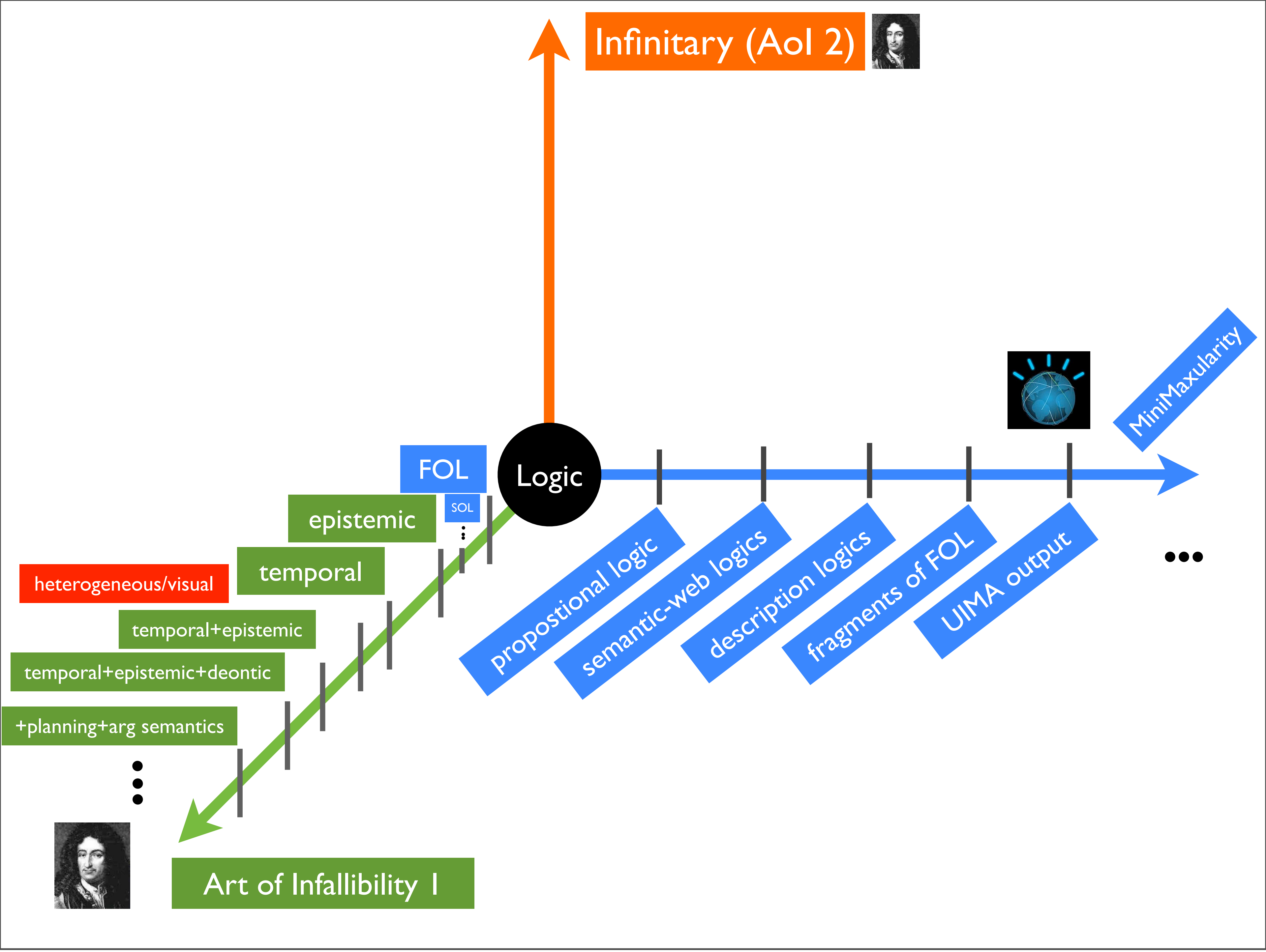

FOL
Logic
SOL
epistemic
temporal
temporal+epistemic
temporal+epistemic+deontic
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity
...


MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
•••
Art of Infallibility I

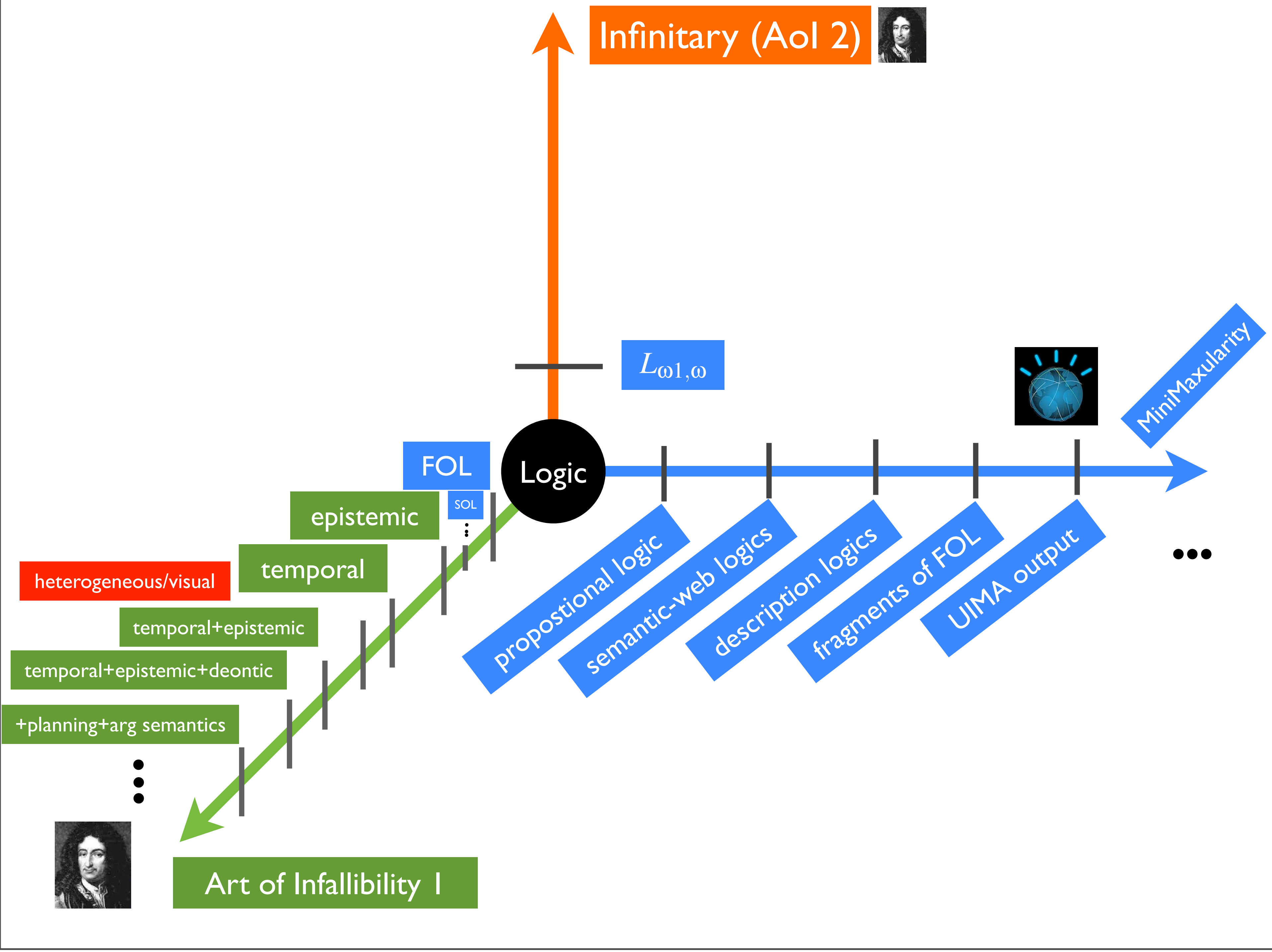

FOL
SOL
Logic
epistemic
temporal
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity

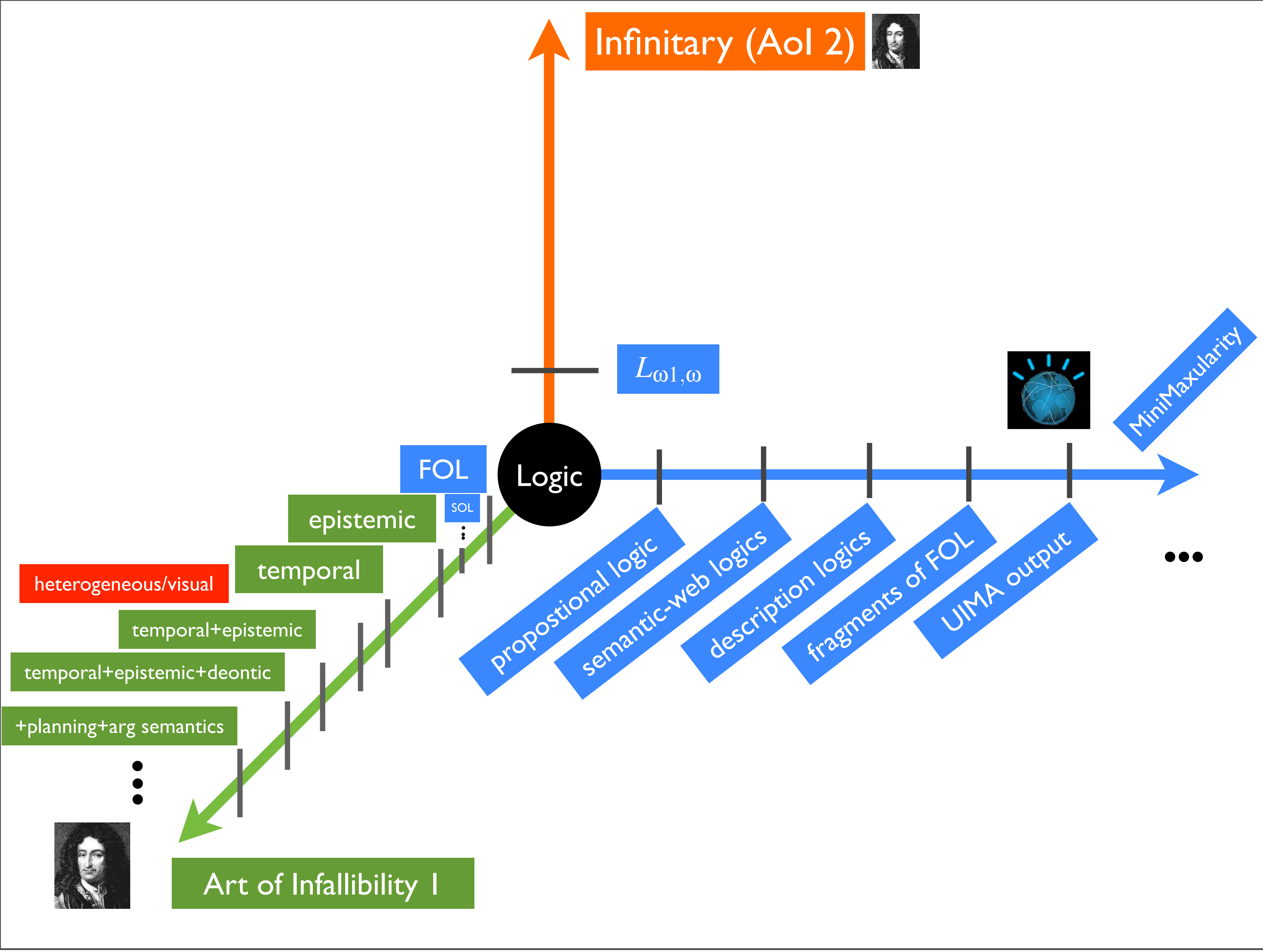

FOL
SOL
Logic
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity


FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity

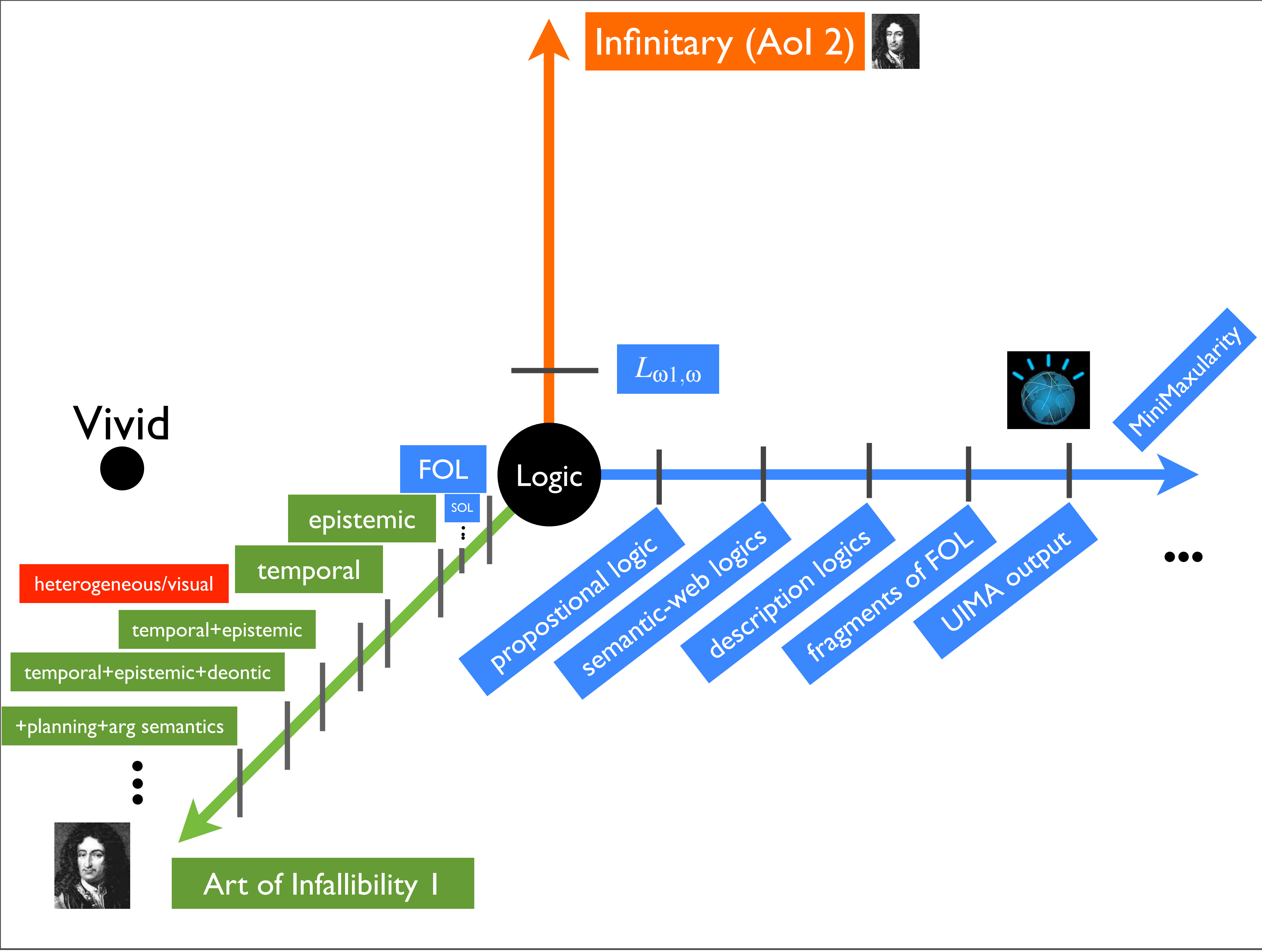


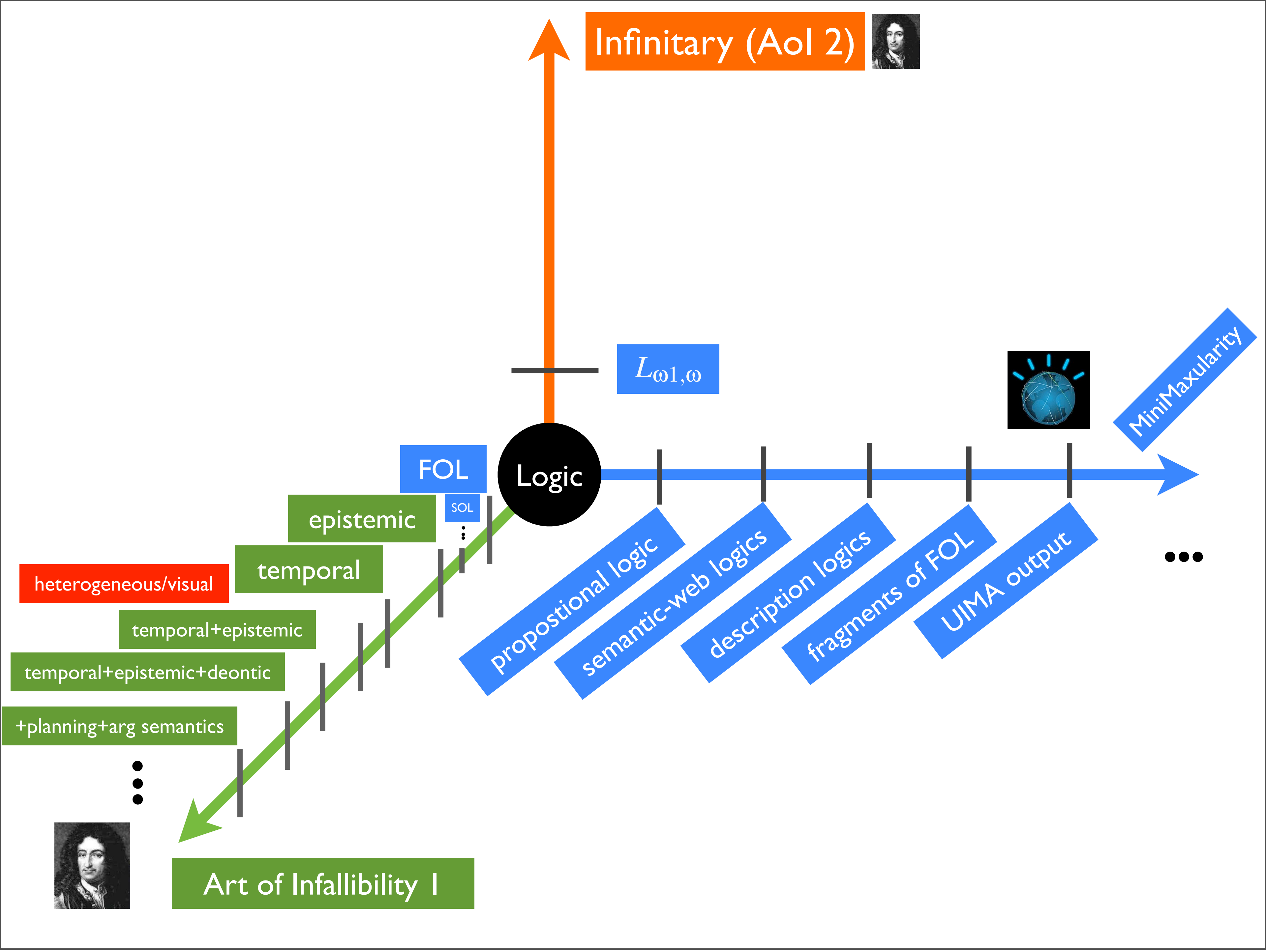
Infinitary (AoI 2)
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity

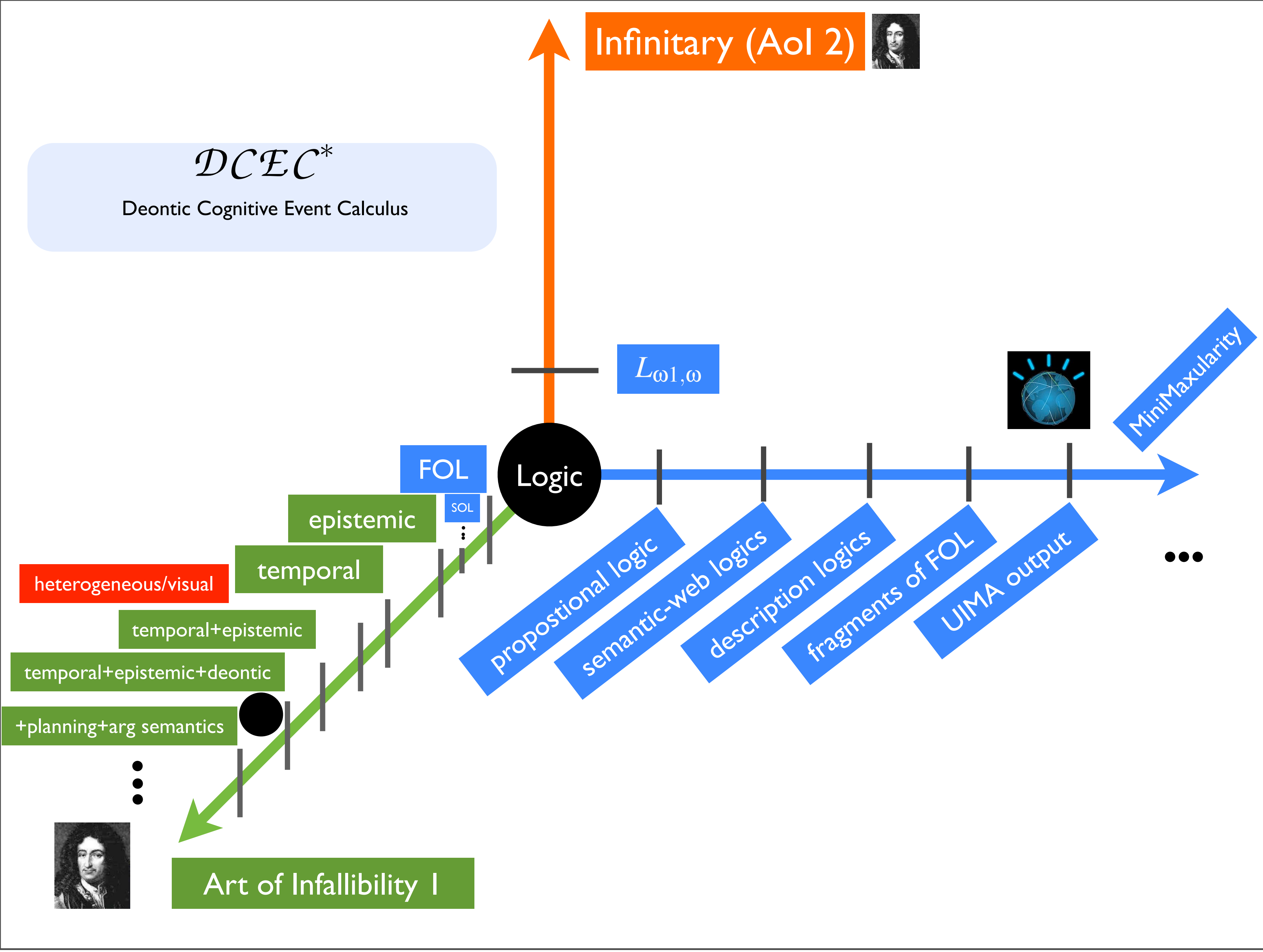

Infinitary (AoI 2)
$L_{\omega1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I



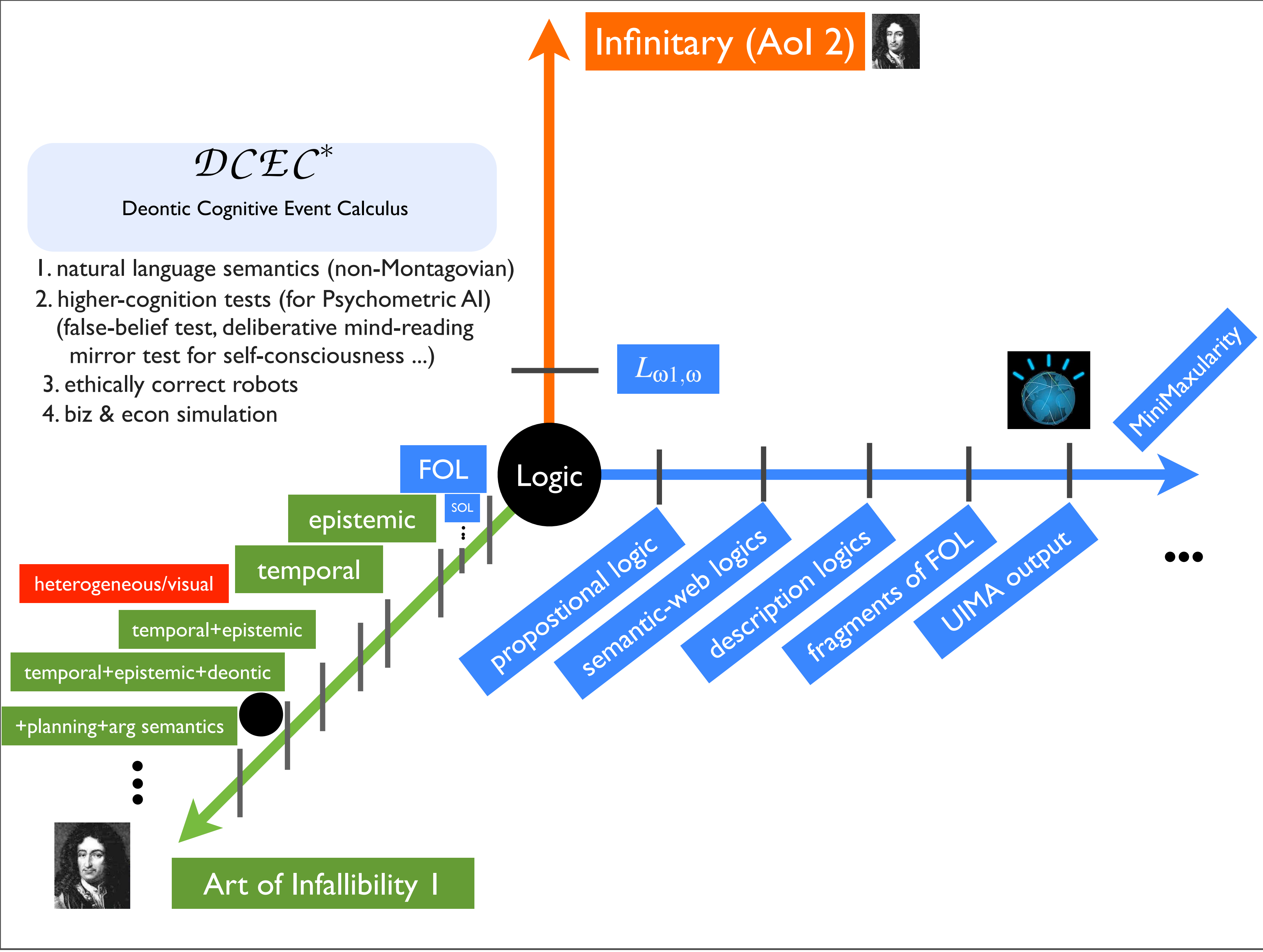
Infinitary (AoI 2)
$L_{\omega1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output

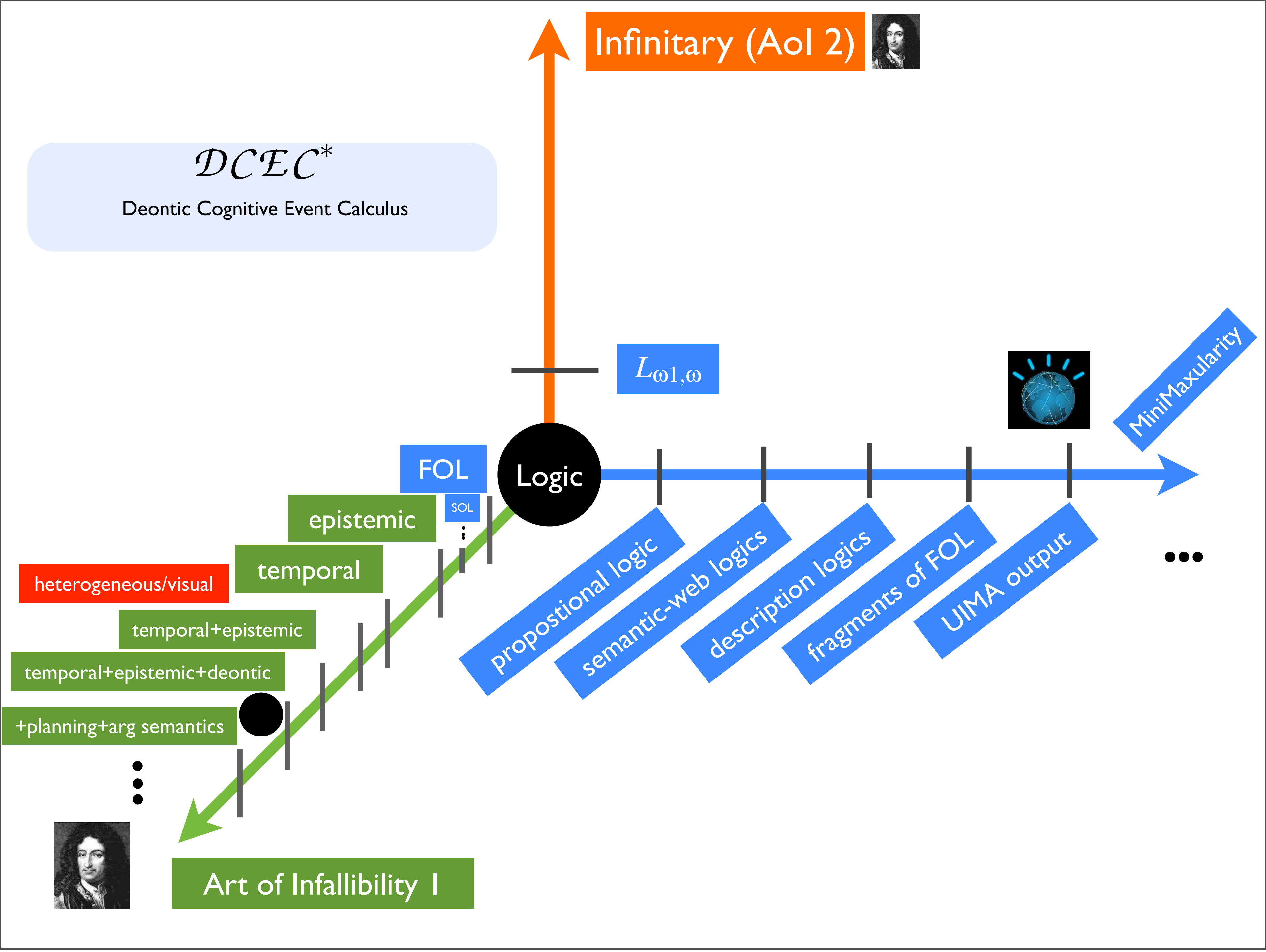

Infinitary (AoI 2)
$L_{\omega 1,\omega}$
Vivid
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity

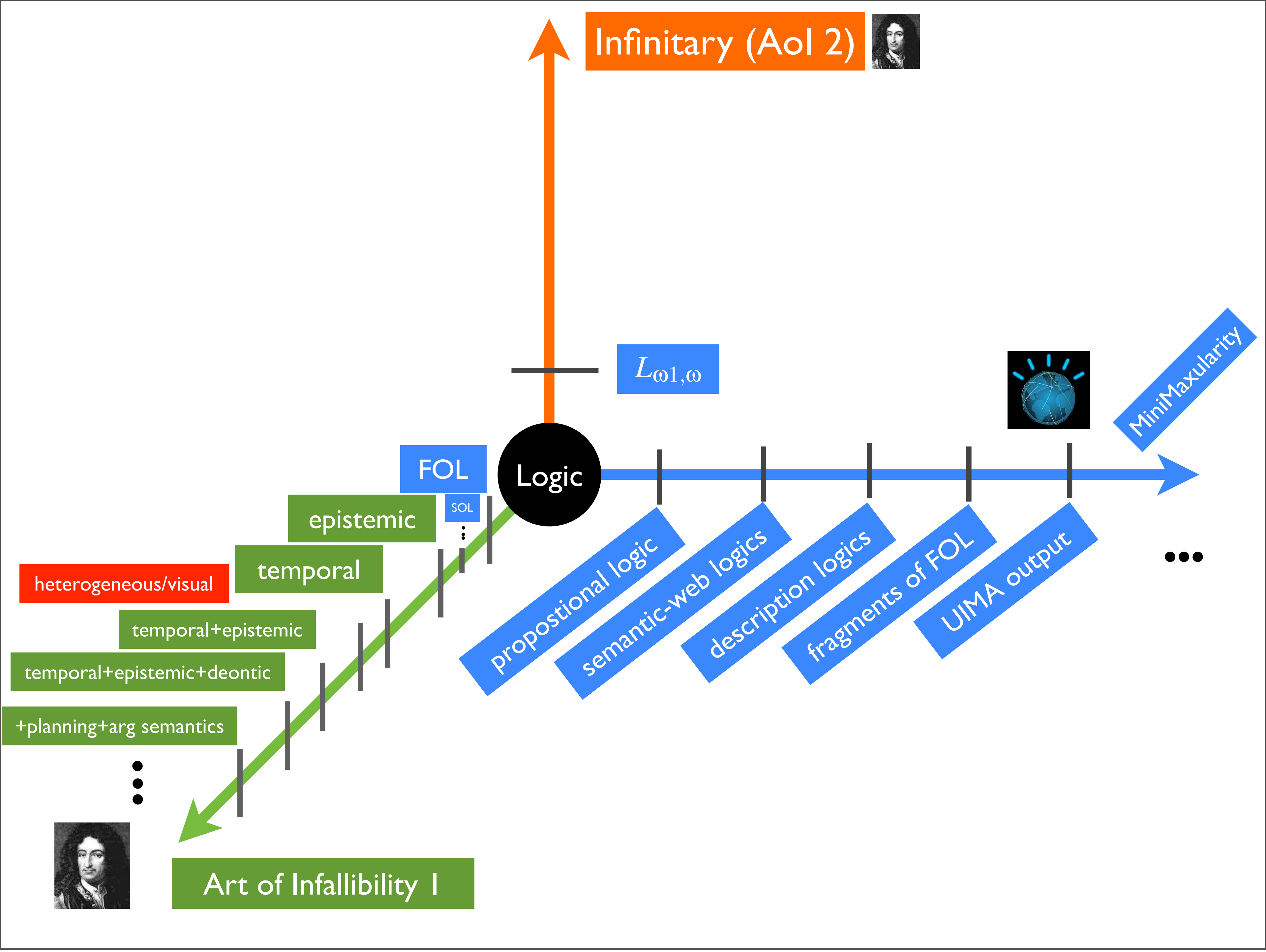

Infinitary (AoI 2)
$L_{\omega 1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I

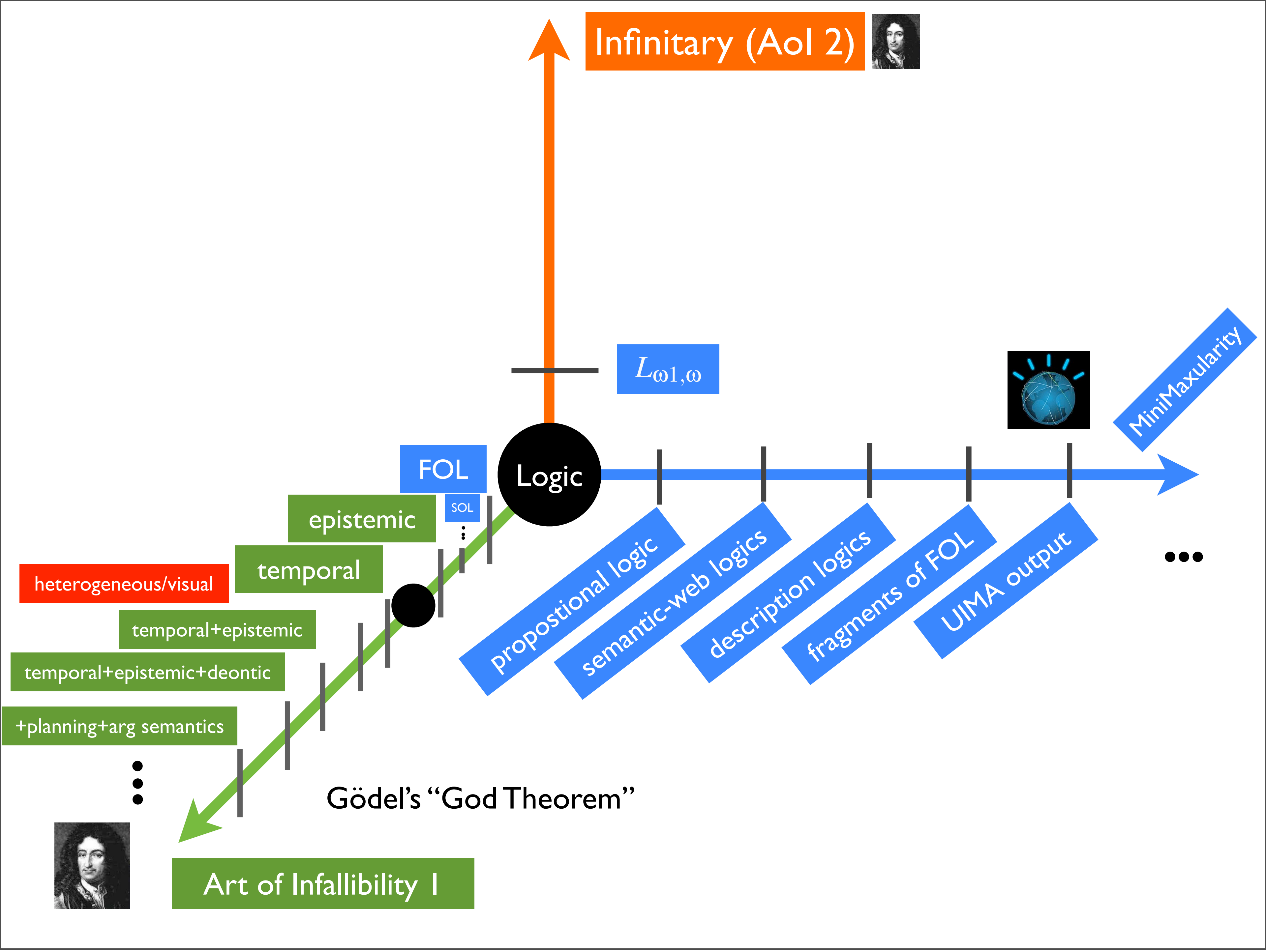


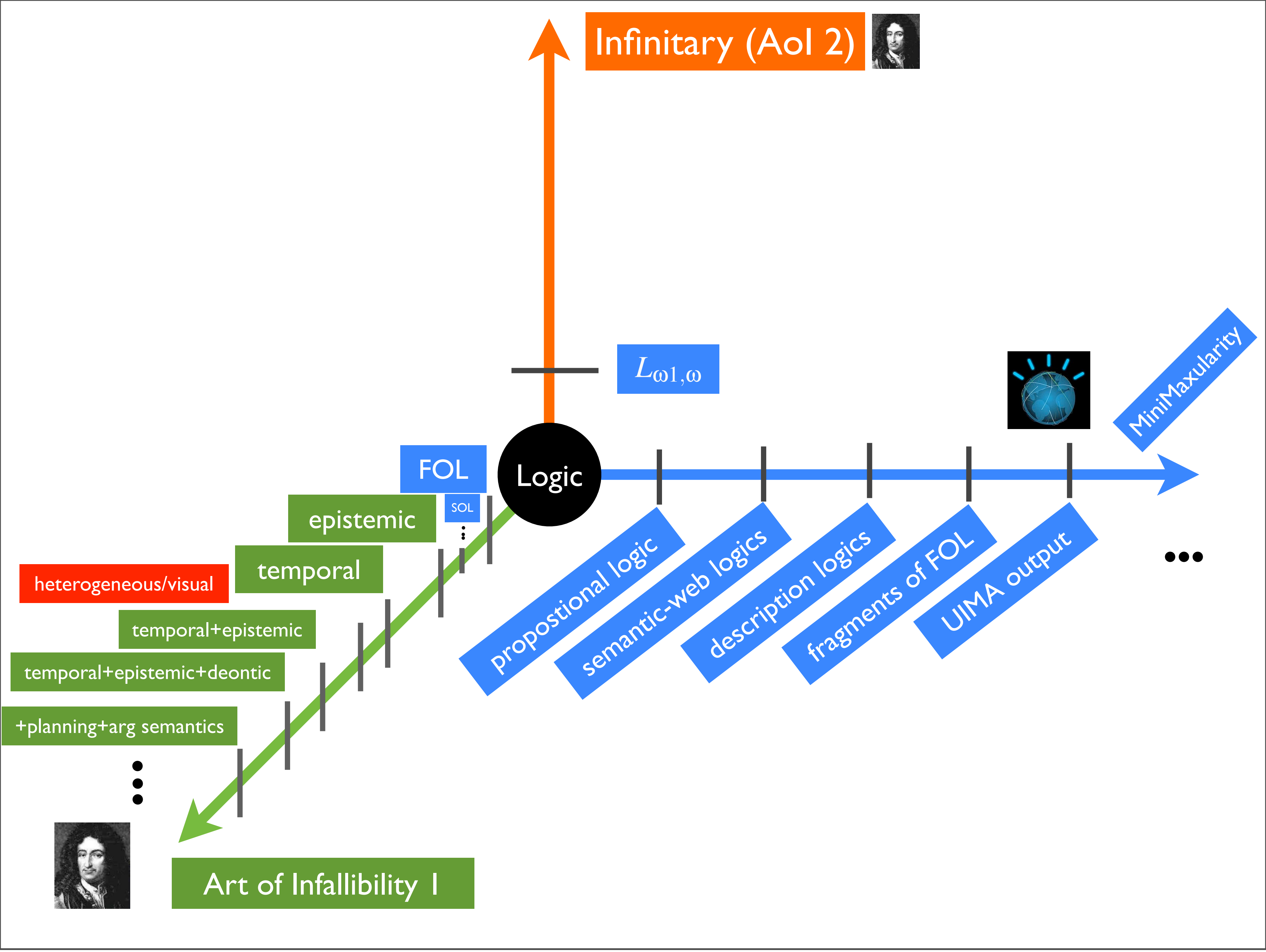
Infinitary (AoI 2)
$\mathcal{DCEC}^*$
Deontic Cognitive Event Calculus
$L_{\omega_1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I

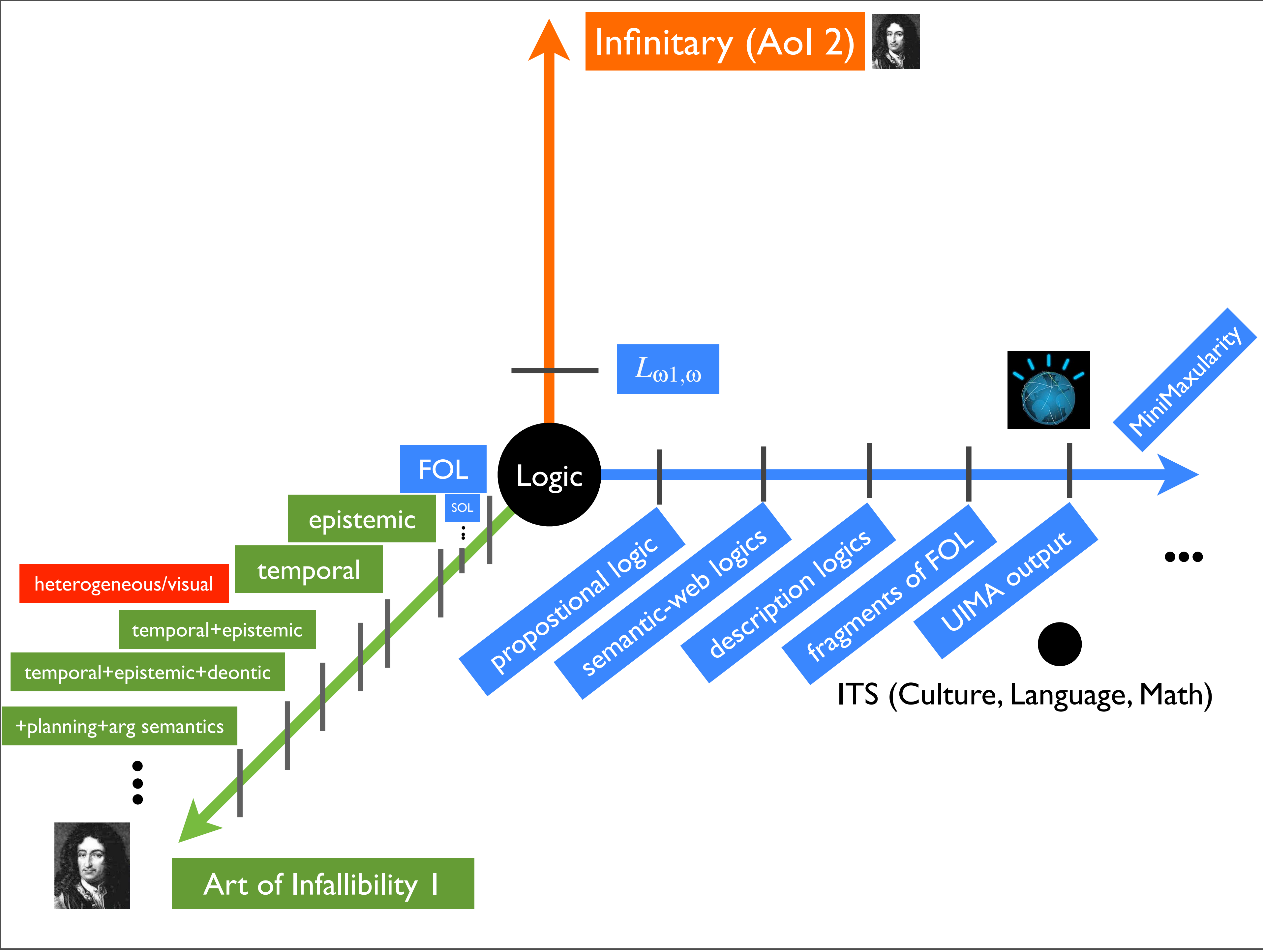

Infinitary (AoI 2)
$\mathcal{DCEC}^*$
Deontic Cognitive Event Calculus
1. natural language semantics (non-Montagovian)
2. higher-cognition tests (for Psychometric AI)
(false-belief test, deliberative mind-reading
mirror test for self-consciousness ...)
3. ethically correct robots
4. biz & econ simulation
$L_{\omega1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I

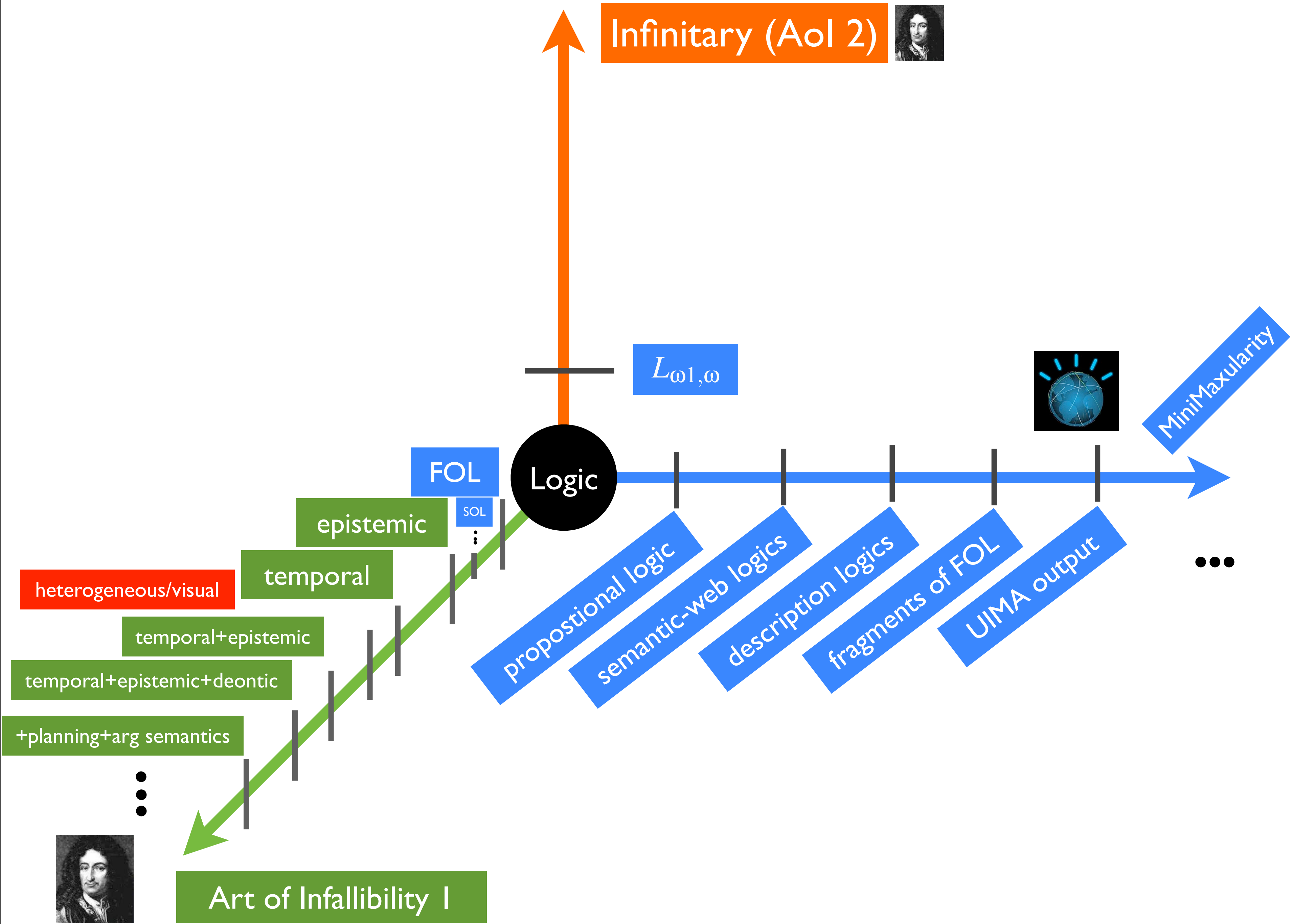

Infinitary (AoI 2)
$\mathcal{DCEC}^*$
Deontic Cognitive Event Calculus
$L_{\omega 1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I


Infinitary (AoI 2)
$L_{\omega1,\omega}$
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity

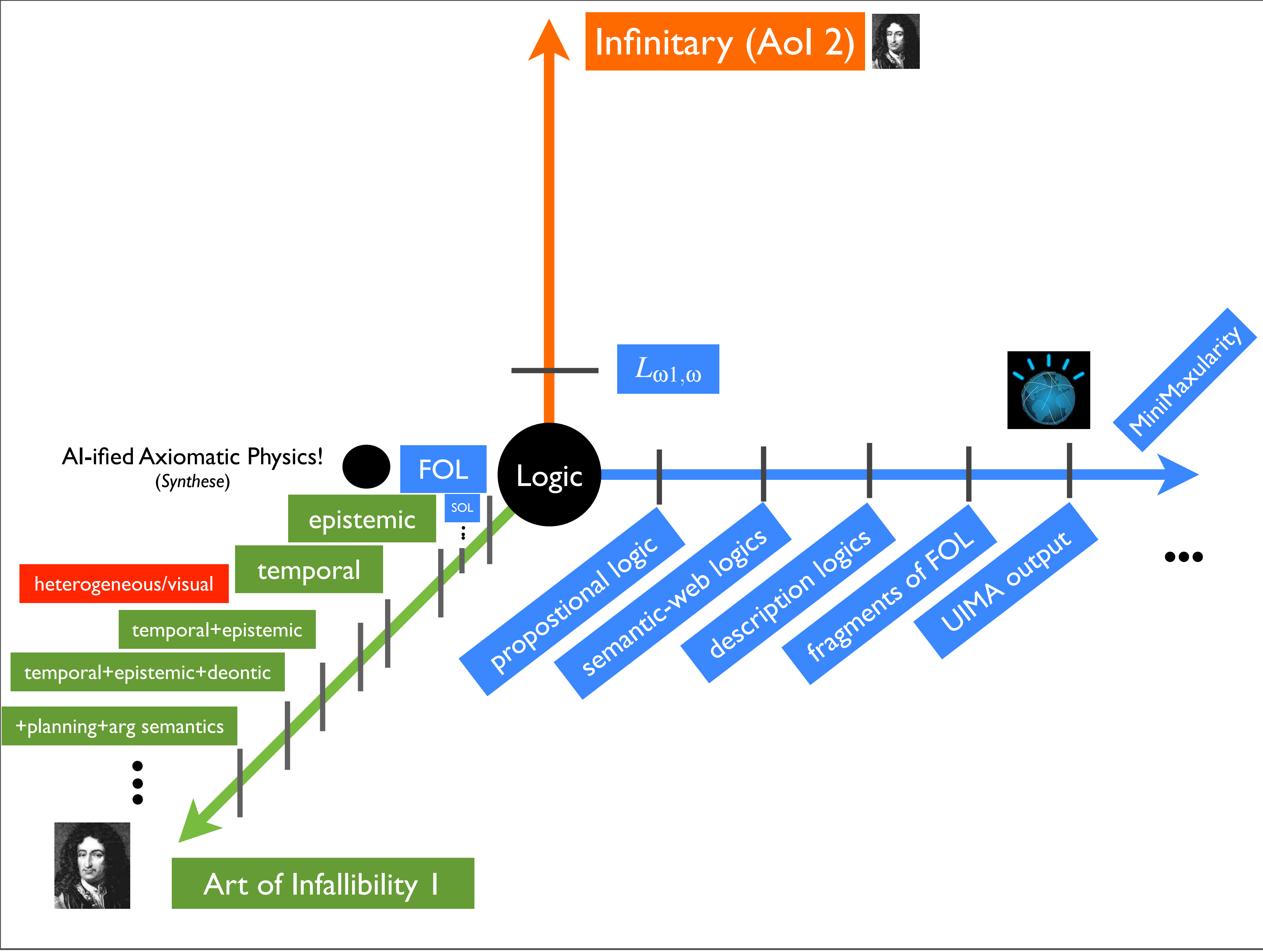

Infinitary (AoI 2)
$L_{\omega 1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Gödel's "God Theorem"
Art of Infallibility I

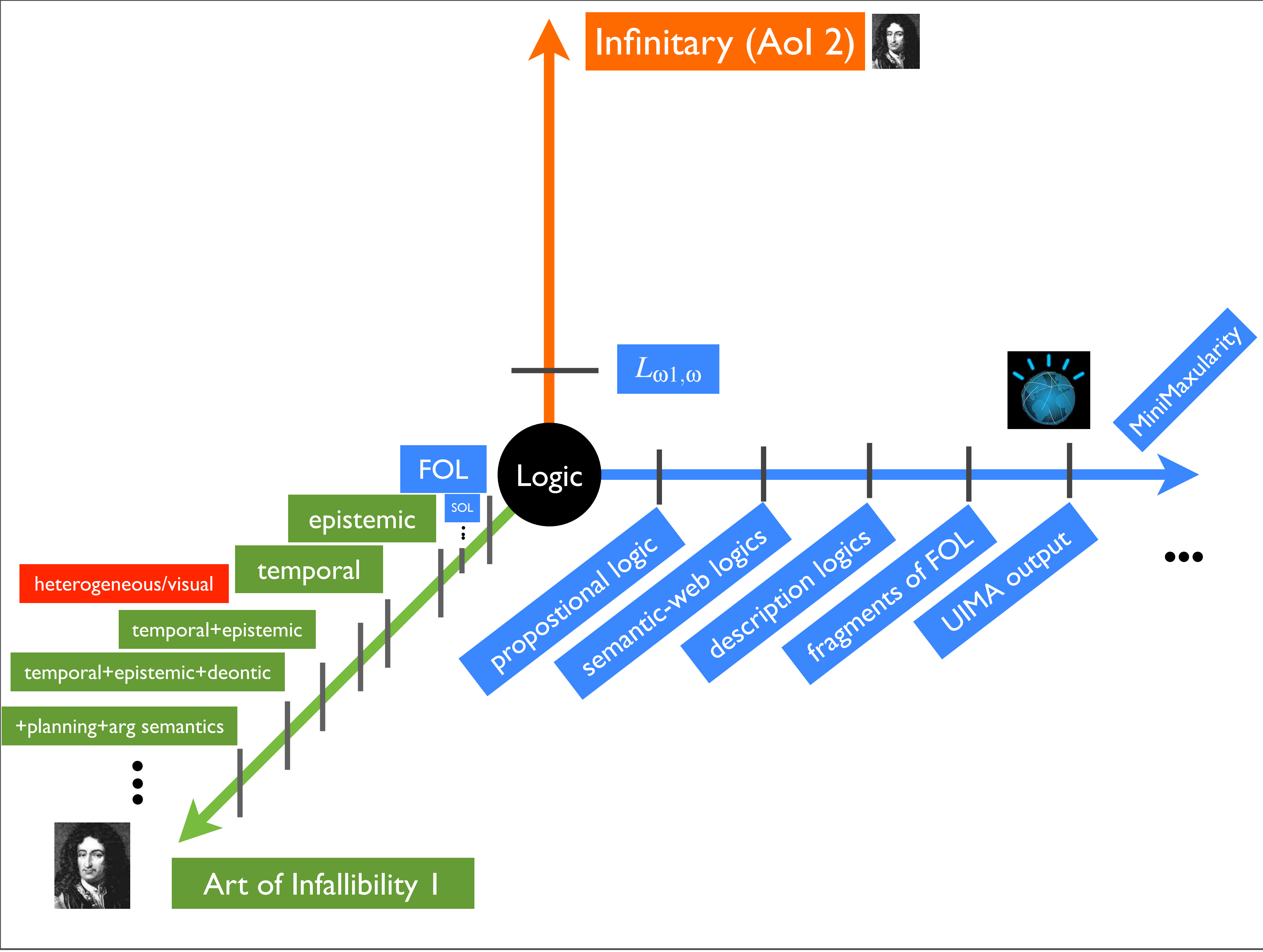

Infinitary (AoI 2)
$L_{\omega 1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I

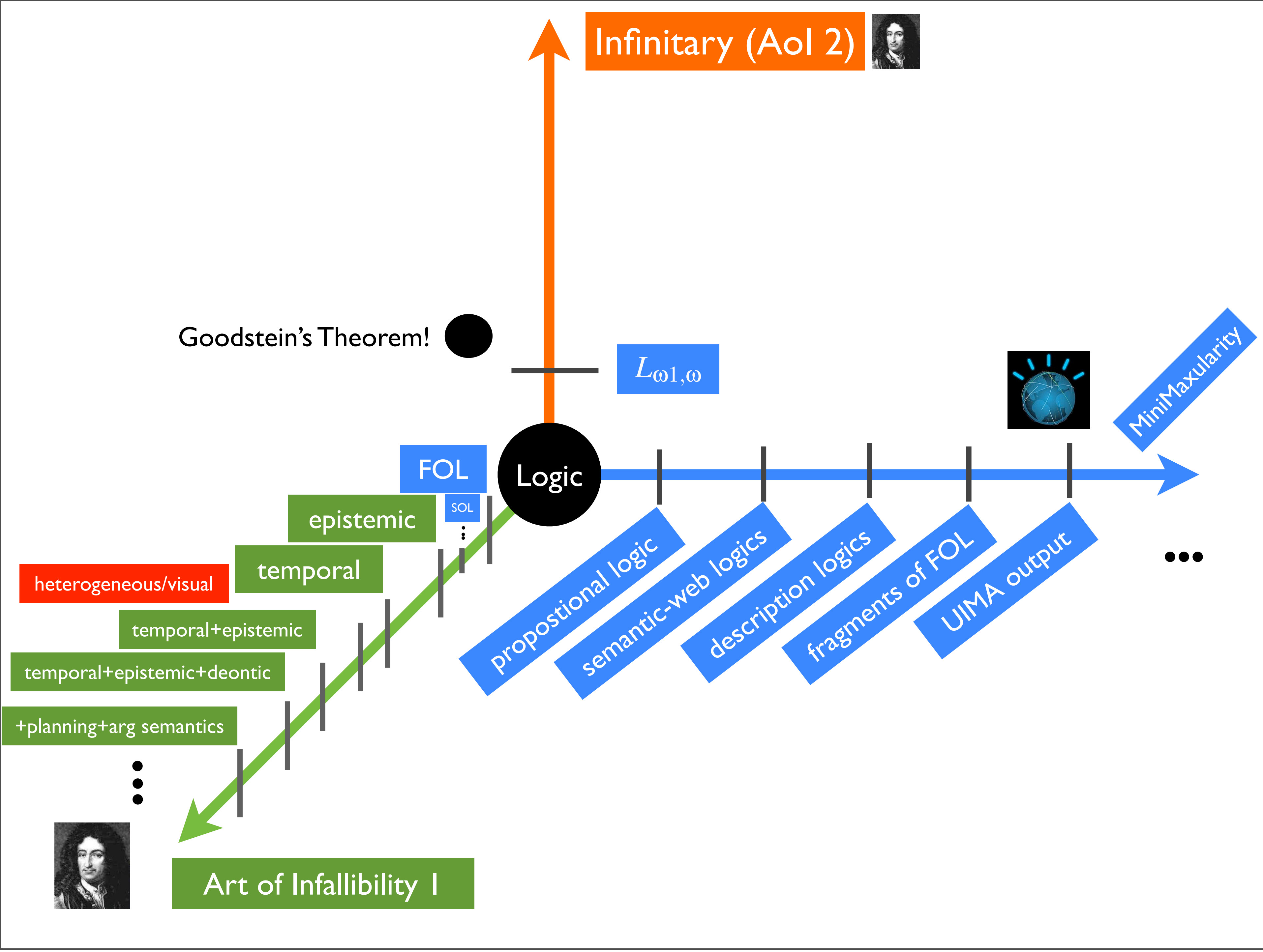

Infinitary (AoI 2)
$L_{\omega 1,\omega}$
Logic
FOL
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
MiniMaxularity
ITS (Culture, Language, Math)



Infinitary (AoI 2)
$L_{\omega 1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I


Infinitary (AoI 2)
$L_{\omega 1,\omega}$
MiniMaxularity
AI-ified Axiomatic Physics!
(Synthese)
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I


Infinitary (AoI 2)
$L_{\omega1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I


Infinitary (AoI 2)
Goodstein's Theorem!
$L_{\omega 1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I

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, [Amnon, 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

Small Steps Toward Hypercomputation via Infinitary Machine Proof Verification and Proof Generation

Naveen Sundar Govindarajulu, John Licato, and Selmer Bringsjord
Department of Computer Science
Department of Cognitive Science
Rensselaer AI & Reasoning Laboratory
govinn@rpi.edu • licatj@rpi.edu • selmer@rpi.edu

Rensselaer Polytechnic Institute
110 8th Street, Troy , NY 12180 USA

Abstract. After setting a context based on two general points (that humans appear to reason in infinitary fashion, and two, that actual hypercomputers aren't currently available to directly model and replicate such infinitary reasoning), we set a humble engineering goal of taking initial steps toward a computing machine that can reason in infinitary fashion. The initial steps consist in our outline of automated proof-verification and proof-discovery techniques for theorems independent of PA that seem to require an understanding and use of infinitary concepts. We specifically focus on proof-discovery techniques that make use of a marriage of analogical and deductive reasoning (which we call *analogico-deductive reasoning*).

A Context: Infinitary Reasoning, Hypercomputation, and Humble Engineering

Bringsjord has repeatedly pointed out the obvious fact that the behavior of formal scientists, taken at face value, involve various infinitary structures and reasoning. (We say "at face value" to simply indicate we don't presuppose some view that denies the reality of infinite entities routinely involved in the formal sciences.) For example, in (Bringsjord & van Heuveln 2003), Bringsjord himself operates as such a scientist in presenting an infinitary paradox which to his knowledge has yet to be solved. And he has argued that apparently infinitary behavior constitutes a grave challenge to AI and the Church-Turing Thesis (e.g., see Bringsjord & Arkoudas 2006, Bringsjord & Zenzen 2003). More generally, Bringsjord conjectures that every human-produced proof of a theorem independent of Peano Arithmetic (PA) will make use of infinitary structures and reasoning, when these structures are taken at face value.[1] We have ourselves designed logico-computational logics for handling infinitary reasoning (e.g., see the treatment of the infinitized wise-man puzzle: Arkoudas & Bringsjord 2005), but this work simply falls back on the human ability to carry out induction on the natural numbers: it doesn't dissect and explain this ability. Finally, it must be admitted by all that there is simply no systematic, comprehensive model or framework anywhere in the formal/computational approach to understanding human knowledge and intelligence that provides a theory about how humans are able to engage with infinitary structures. This is revealed perhaps most clearly when one studies the fruit produced by the part of formal AI devoted to producing discovery systems: such fruit is embarrassingly finitary (e.g., see Shilliday 2009).

Given this context, we are interested in exploring how one might give a machine the ability to reason in infinitary fashion. We are not saying that we in fact have figured out how to give such ability to a computing machine. Our objective here is much more humble and limited: it is to push forward in the *attempt* to engineer a computing machine that has the ability to reason in infinitary fashion. Ultimately, if such an attempt is to succeed, the computing machine in question will presumably be capable of outright hypercomputation. But the fact is that from an engineering perspective, we don't know how to create and harness a hypercomputer. So what we must first try to do, as explained in (Bringsjord & Zenzen 2003), is pursue engineering that initiates the attempt to engineer a hypercomputer, and takes the first few steps. In the present paper, the engineering is aimed specifically at giving a computing machine the ability to, in a limited but well-defined sense, reason in infinitary fashion. Even more specifically, our engineering is aimed at building a machine capable of at least providing a strong case for a result which, in the human sphere, has hitherto required use of infinitary techniques.

[1] A weaker conjecture along the same line has been ventured by Isaacson, and is elegantly discussed by Smith (2007).

Govindarajulu, N.; Licato, J.; Bringsjord, S. 2013. Small Steps Toward Hypercomputation via Infinitary Machine Proof Verification and Proof Generation. In *Proceedings of UCNC 2013*. Pdf

Ending one of AI's "bad dreams":

Liar Paradox (L)

- Collection of semiformal statements
- Syntax not rigorously defined
- Represents intuitive understanding of problem domain

s = "This statement is a lie"
There is a statement that is neither true nor false.
...

G1

- Completely formal statements
- Syntax very rigorously defined
- Purely mathematical objects: numbers, formal theories, etc.

s = ?

$$\exists_{\varphi \in L_A} \neg(PA \vdash \varphi) \wedge \neg(PA \vdash \neg\varphi)$$

...

Ending one of AI's "bad dreams":

Liar Paradox (L)

- Collection of semiformal statements
- Syntax not rigorously defined
- Represents intuitive understanding of problem domain

s = "This statement is a lie"
There is a statement that is neither true nor false.
...

G1

- Completely formal statements
- Syntax very rigorously defined
- Purely mathematical objects: numbers, formal theories, etc.

s = ?

$$\exists_{\varphi \in L_A} \neg(PA \vdash \varphi) \wedge \neg(PA \vdash \neg\varphi)$$

...

Ending one of AI's "bad dreams":

Liar Paradox (L)

- Collection of semiformal statements
- Syntax not rigorously defined
- Represents intuitive understanding of problem domain

s = "This statement is a lie"
There is a statement that is neither true nor false.
...

S

- Semiformal statements (but more formal than L)
- Syntax somewhat rigorously defined
- Somewhat intuitive; deals with stories of reasoners and utterances made by inhabitants of an island

$s = \neg Bs$

$\exists_p \neg Bp \wedge \neg B \neg p$

...

GI

- Completely formal statements
- Syntax very rigorously defined
- Purely mathematical objects: numbers, formal theories, etc.

s = ?

$\exists_{\varphi \in L_A} \neg(PA \vdash \varphi) \wedge \neg(PA \vdash \neg\varphi)$

...

Ending one of AI's "bad dreams":

Liar Paradox (L)

- Collection of semiformal statements
- Syntax not rigorously defined
- Represents intuitive understanding of problem domain

s = "This statement is a lie"
There is a statement that is neither true nor false.
...

S

- Semiformal statements (but more formal than L)
- Syntax somewhat rigorously defined
- Somewhat intuitive; deals with stories of reasoners and utterances made by inhabitants of an island

$s = \neg Bs$

$\exists_p \neg Bp \wedge \neg B\neg p$

...

GI

- Completely formal statements
- Syntax very rigorously defined
- Purely mathematical objects: numbers, formal theories, etc.

s = ?

$\exists_{\varphi \in L_A} \neg(PA \vdash \varphi) \wedge \neg(PA \vdash \neg\varphi)$

...

The Singularity Approaches ...

The Singularity Approaches ...

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI = HI).
	Premise 2	If there is AI, there will be AI^{+} (created by AI).
	Premise 3	If there is AI^{+}, there will be AI^{++} (created by AI^{+}).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

The Singularity Approaches ...

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI = HI).
	Premise 2	If there is AI, there will be AI^{+} (created by AI).
	Premise 3	If there is AI^{+}, there will be AI^{++} (created by AI^{+}).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

(Good 1965; Chalmers 2010)

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI = HI).
	Premise 2	If there is AI, there will be AI^{+} (created by AI).
	Premise 3	If there is AI^{+}, there will be AI^{++} (created by AI^{+}).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

If AI is only Ray-1 AI ...

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI = HI).
	Premise 2	If there is AI, there will be AI^{+} (created by AI).
	Premise 3	If there is AI^{+}, there will be AI^{++} (created by AI^{+}).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

If AI is only Ray-1 AI ...

$\mathcal{A}$: False!

	~~**Premise 1**~~	~~There will be AI (created by HI and such that AI = HI).~~
	Premise 2	If there is AI, there will be AI^+ (created by AI).
	Premise 3	If there is AI^+, there will be AI^{++} (created by AI^+).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI = HI).
	Premise 2	If there is AI, there will be AI^{+} (created by AI).
	Premise 3	If there is AI^{+}, there will be AI^{++} (created by AI^{+}).
$\therefore$	**S**	There will be AI^{++} ($= \mathcal{S}$ will occur).

If AI is only Ray-2 AI ...

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI = HI).
	Premise 2	If there is AI, there will be AI^{+} (created by AI).
	Premise 3	If there is AI^{+}, there will be AI^{++} (created by AI^{+}).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

If AI is only Ray-2 AI ...

False!

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI $=$ HI).
	~~**Premise 2**~~	~~If there is AI, there will be AI$^+$ (created by AI).~~
	Premise 3	If there is AI$^+$, there will be AI^{++} (created by AI$^+$).
$\therefore$	**S**	There will be AI^{++} ($= \mathcal{S}$ will occur).

If AI is only Ray-2 AI ...

False!
False!

$\mathcal{A}$:

	~~**Premise 1**~~	~~There will be AI (created by HI and such that AI = HI).~~
	~~**Premise 2**~~	~~If there is AI, there will be AI^+ (created by AI).~~
	Premise 3	If there is AI^+, there will be AI^{++} (created by AI^+).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

If AI is only Ray-2 AI ...

False!
False!

$\mathcal{A}$:

~~**Premise 1** There will be AI (created by HI and such that AI = HI).~~

~~**Premise 2** If there is AI, there will be AI^+ (created by AI).~~

Premise 3 If there is AI^+, there will be AI^{++} (created by AI^+).

$\therefore$ **S** There will be AI^{++} ($= \mathcal{S}$ will occur).

And of course ...

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI = HI).
	Premise 2	If there is AI, there will be AI^+ (created by AI).
	Premise 3	If there is AI^+, there will be AI^{++} (created by AI^+).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

... if AI is Ray-3 AI ...

$\mathcal{A}$:

	Premise 1	There will be AI (created by HI and such that AI = HI).
	Premise 2	If there is AI, there will be AI^{+} (created by AI).
	Premise 3	If there is AI^{+}, there will be AI^{++} (created by AI^{+}).
$\therefore$	**S**	There will be AI^{++} ($= \mathcal{S}$ will occur).

... if AI is Ray-3 AI ...

False!

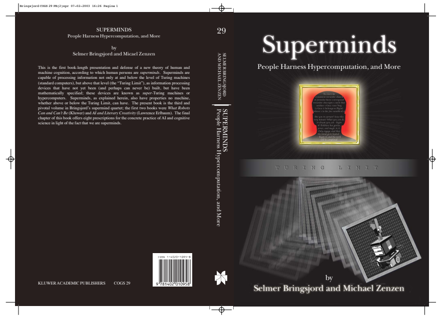
$\mathcal{A}$: False!

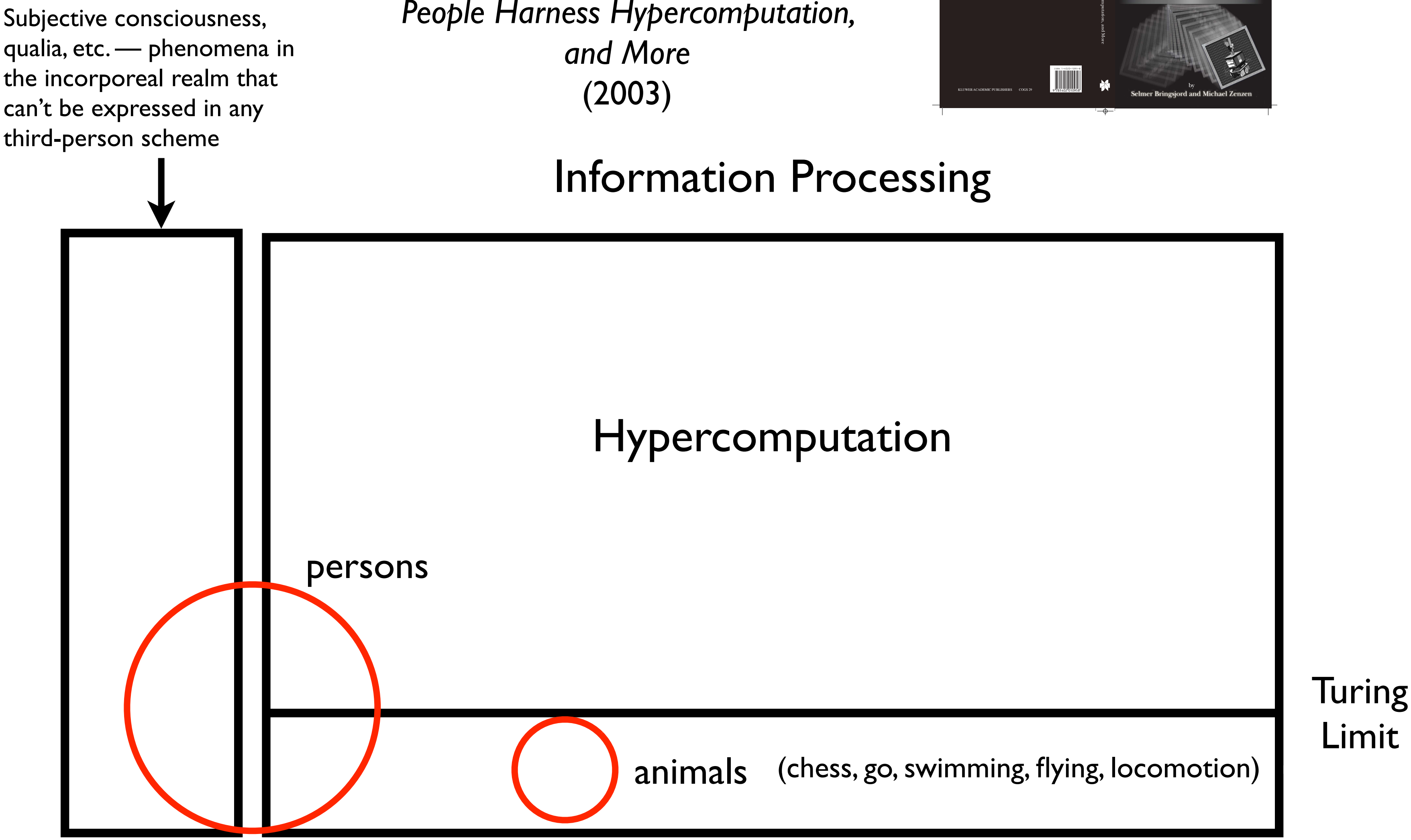
	~~**Premise 1**~~	~~There will be AI (created by HI and such that AI = HI).~~
	~~**Premise 2**~~	~~If there is AI, there will be AI$^+$ (created by AI).~~
	Premise 3	If there is AI$^+$, there will be AI^{++} (created by AI$^+$).
$\therefore$	**S**	There will be AI^{++} (= $\mathcal{S}$ will occur).

"My, that's rather negative, even violent. Can we talk about *positive* views, please?"

Superminds

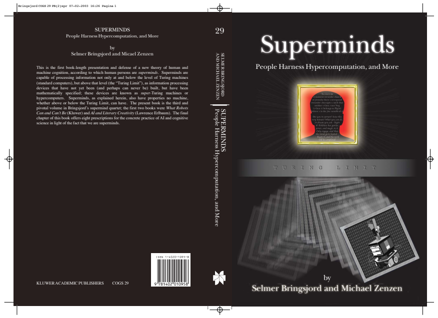
People Harness Hypercomputation, and More
(2003)





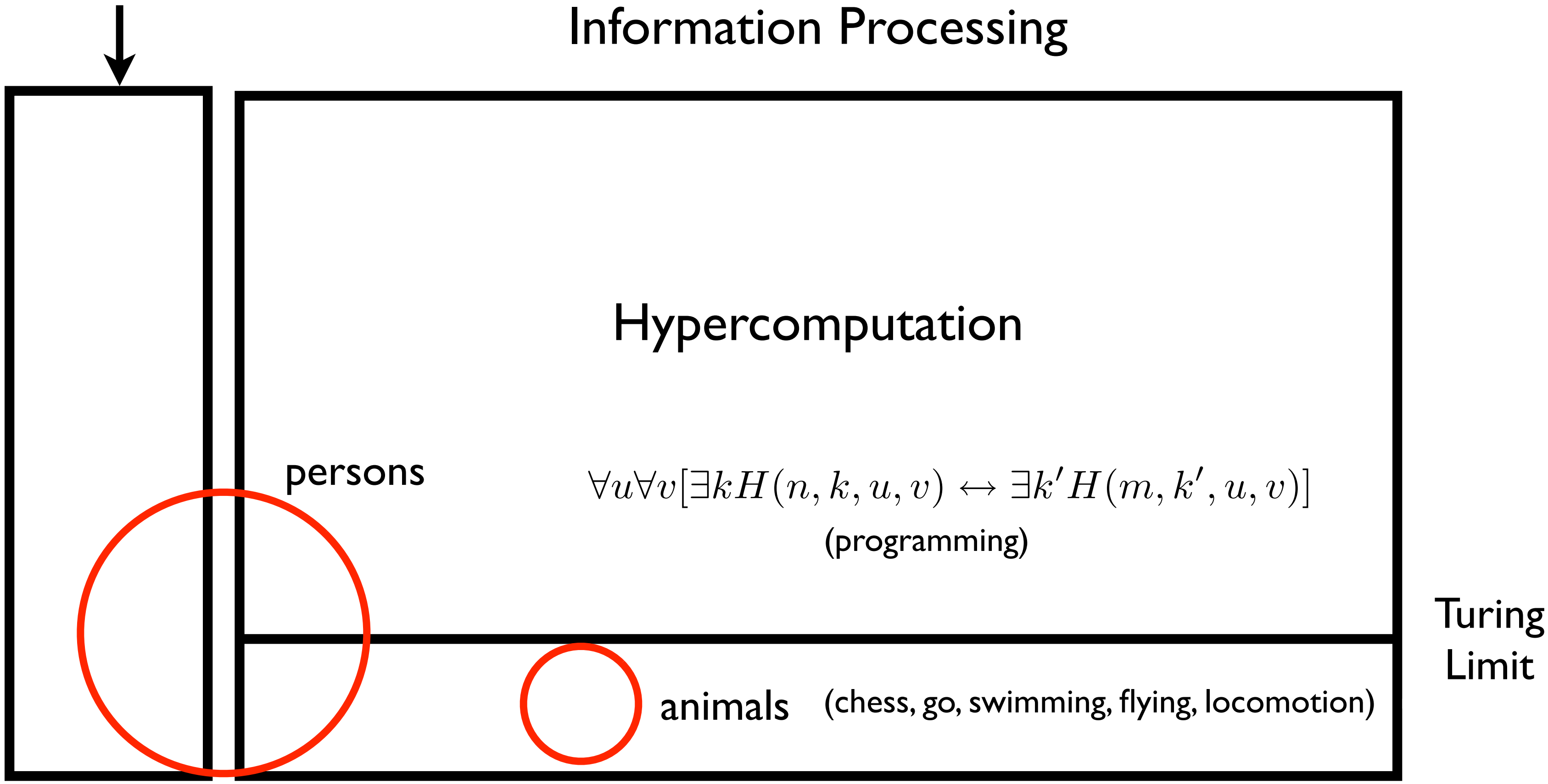
Superminds

People Harness Hypercomputation, and More
(2003)



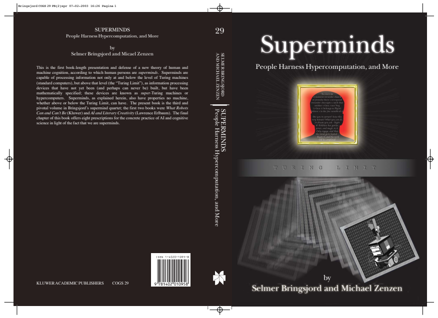
Subjective consciousness, qualia, etc. — phenomena in the incorporeal realm that can’t be expressed in any third-person scheme

Information Processing

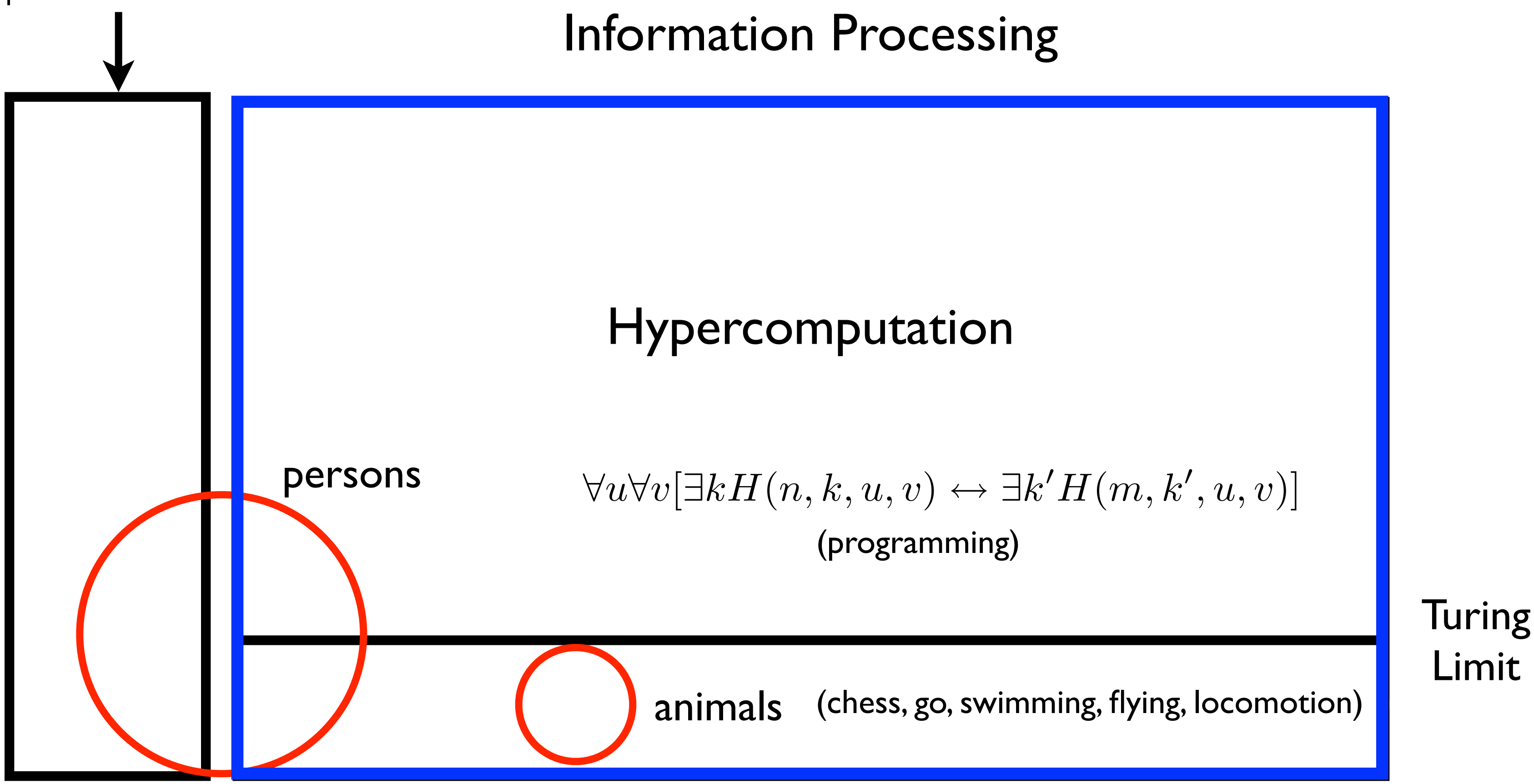


Superminds

People Harness Hypercomputation, and More
(2003)

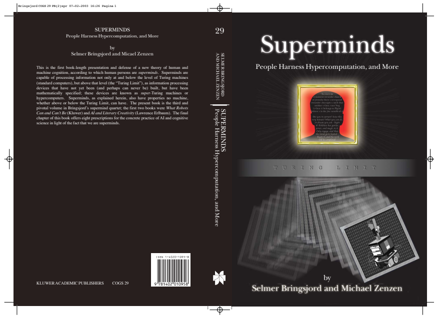


Subjective consciousness, qualia, etc. — phenomena in the incorporeal realm that can't be expressed in any third-person scheme

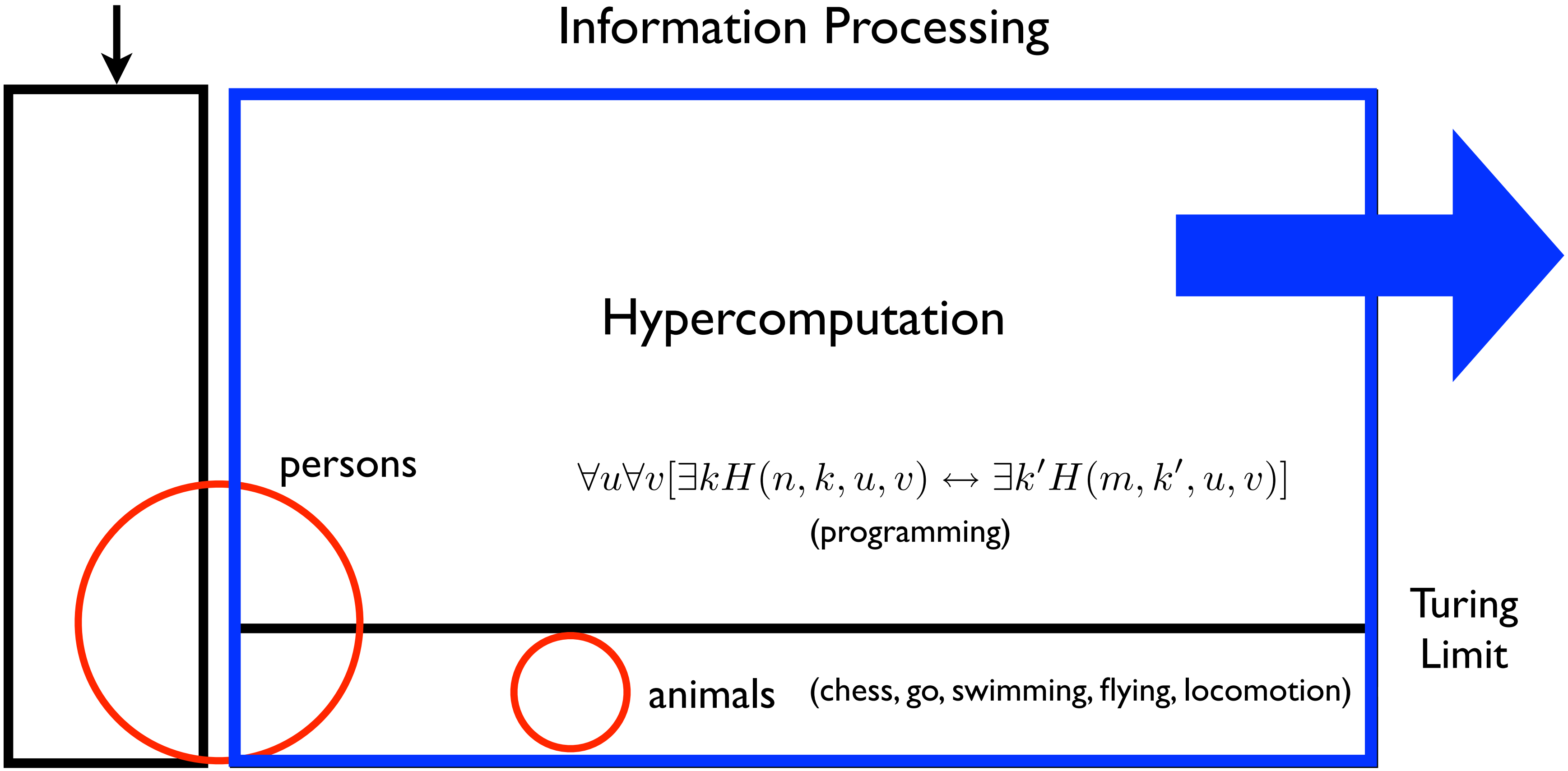


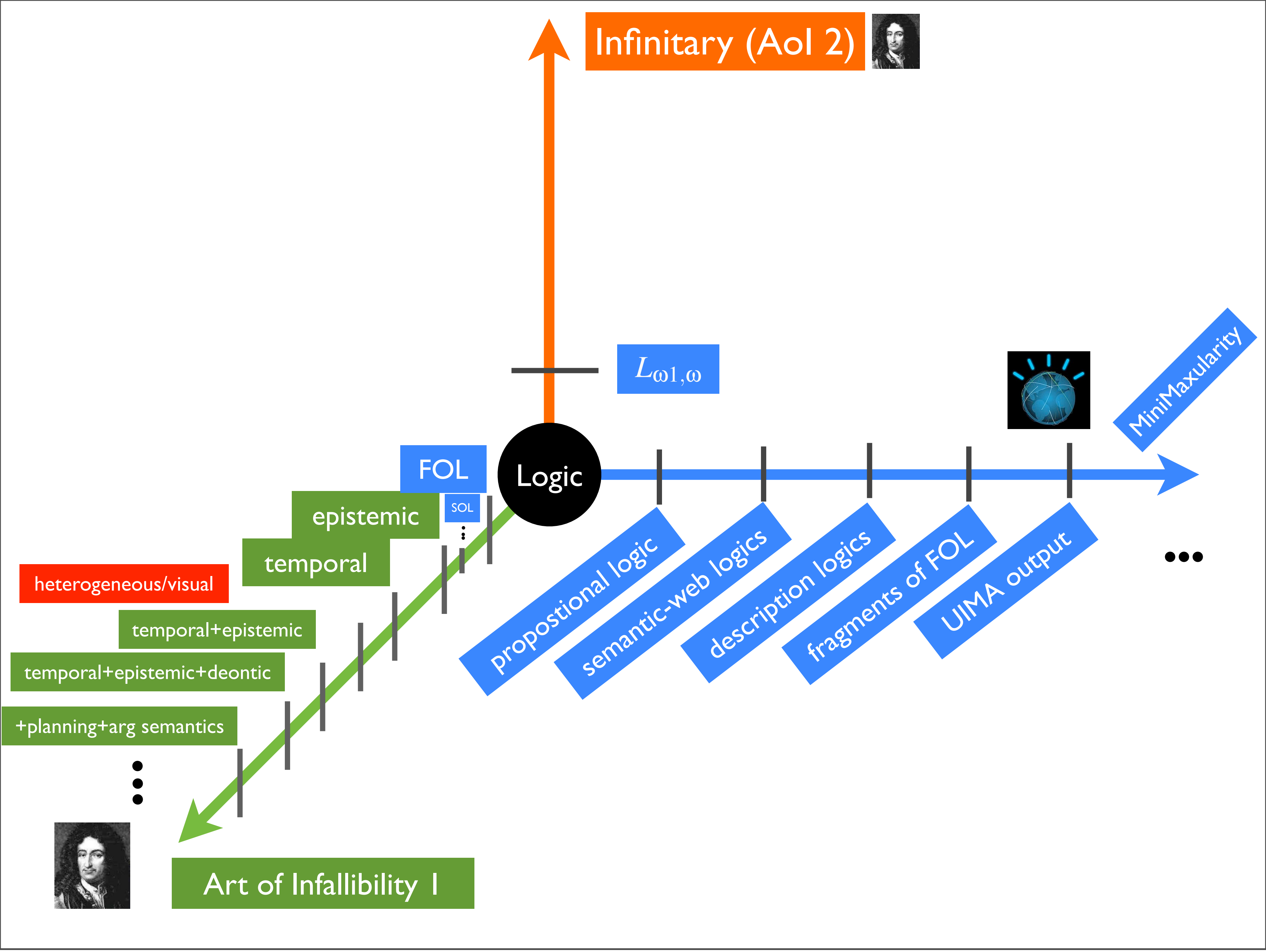
Superminds

People Harness Hypercomputation, and More
(2003)



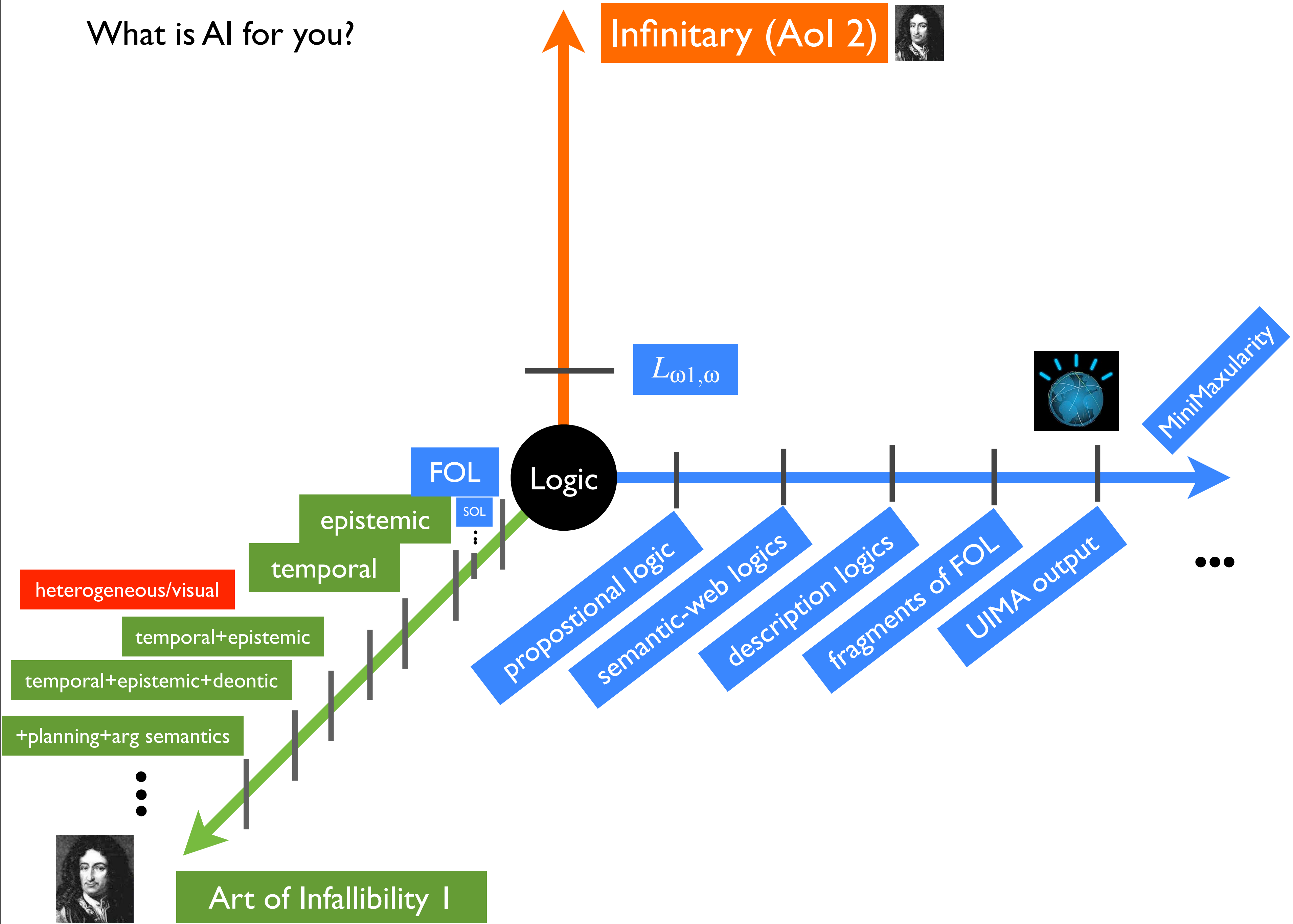
Subjective consciousness, qualia, etc. — phenomena in the incorporeal realm that can't be expressed in any third-person scheme

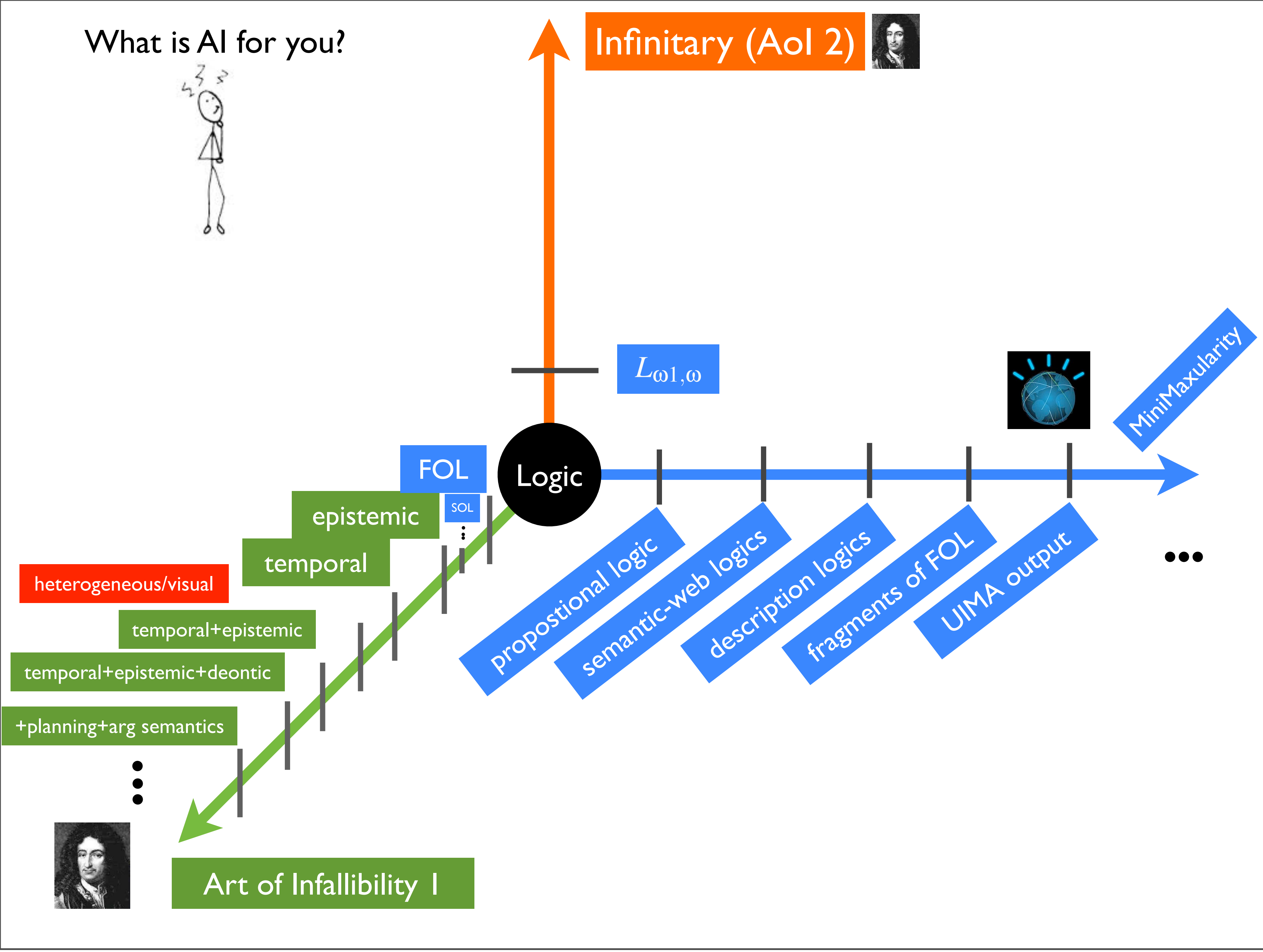


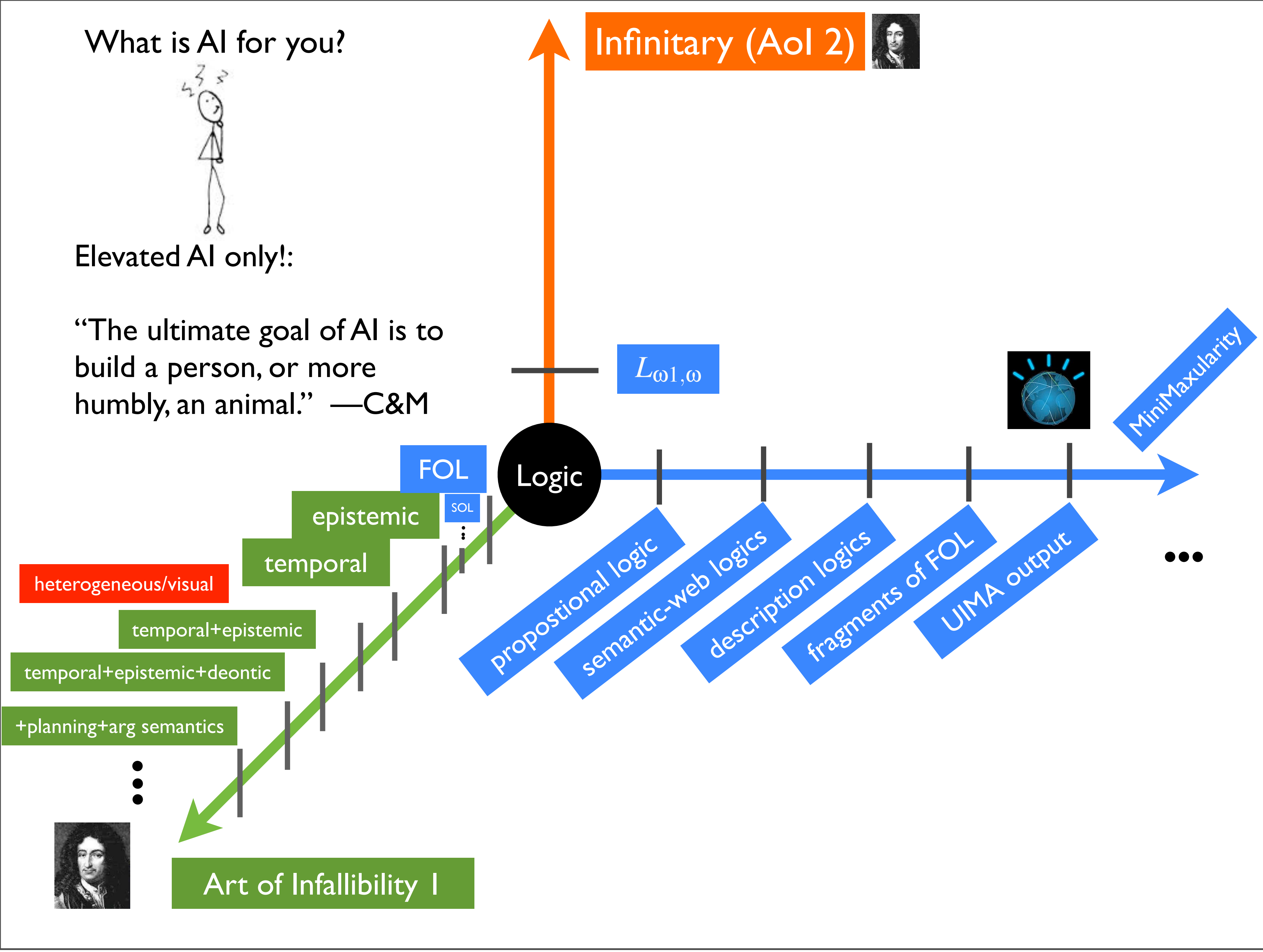


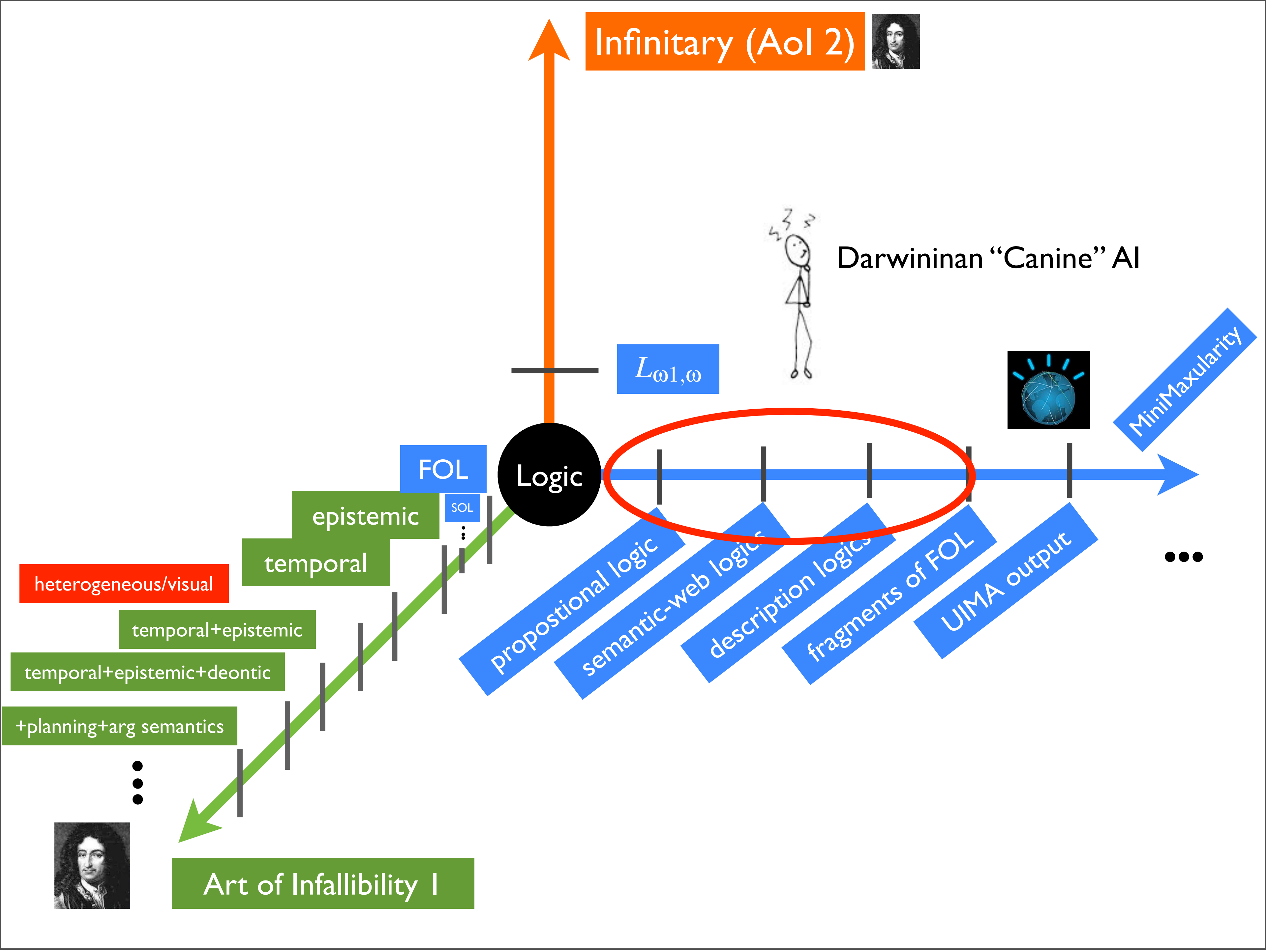
Tuesday, October 1, 13

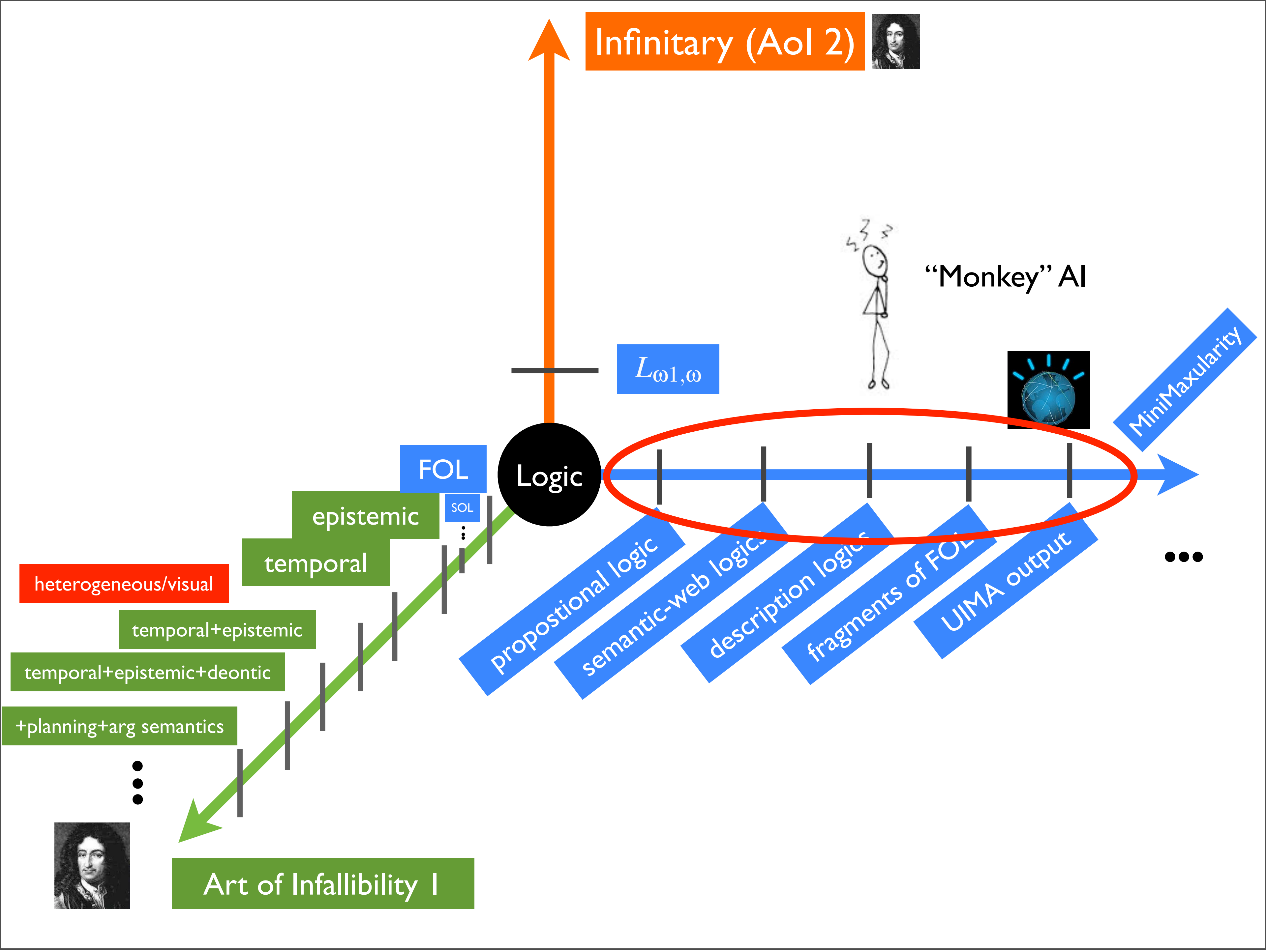
What is AI for you?

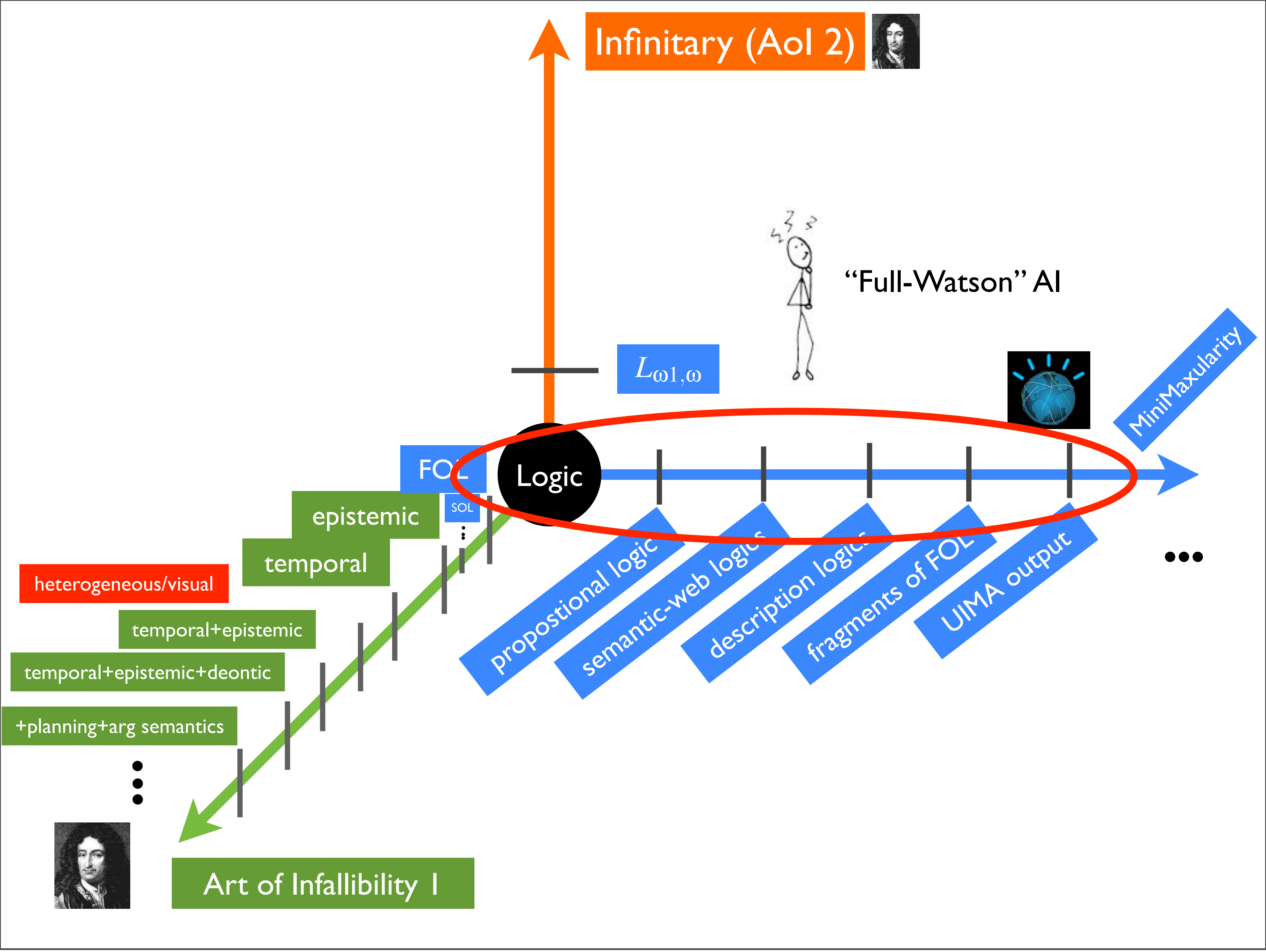



What is AI for you?
Infinitary (AoI 2)
$L_{\omega_1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I

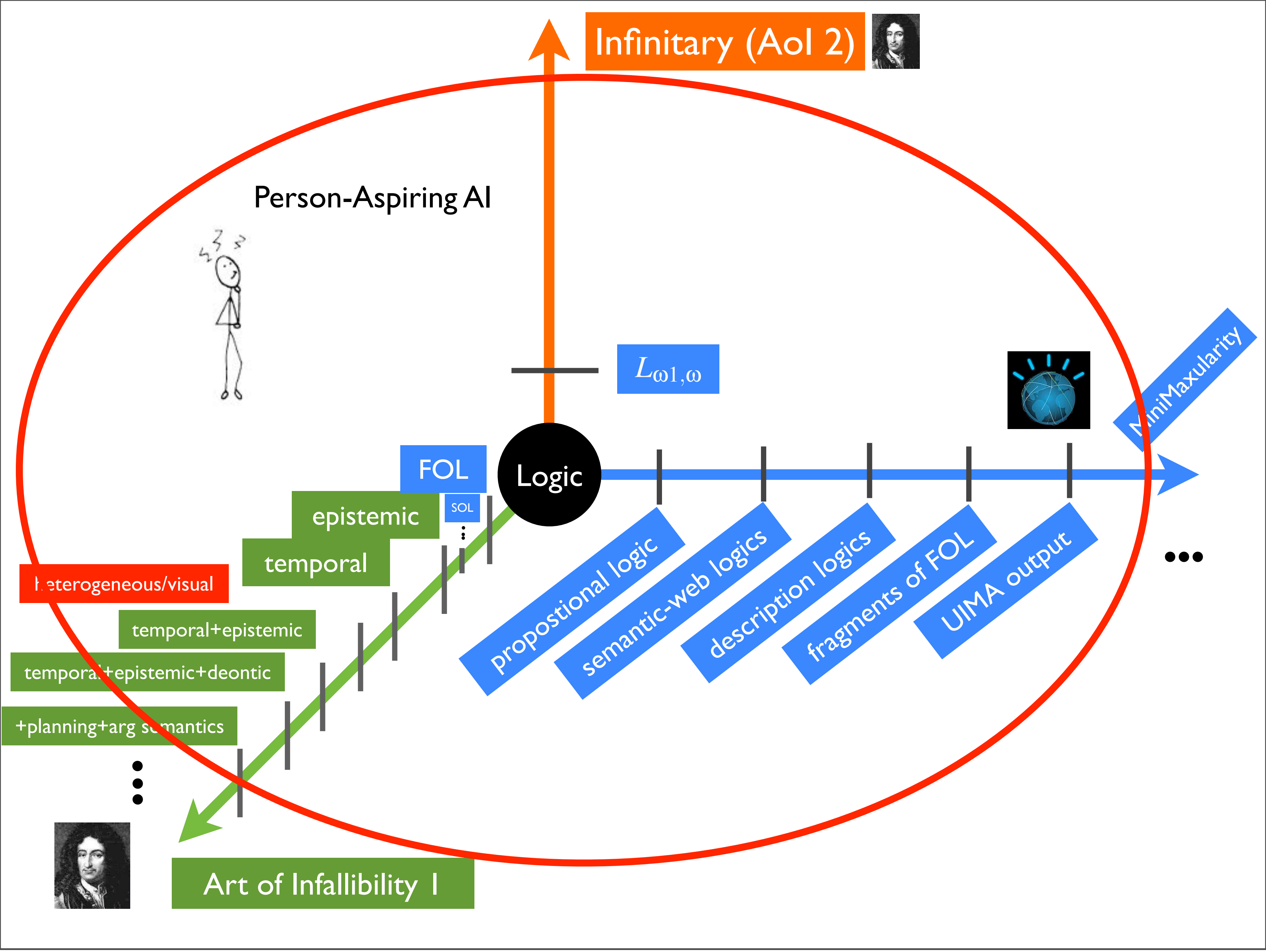

What is AI for you?
Elevated AI only!:
“The ultimate goal of AI is to build a person, or more humbly, an animal.” —C&M
Infinitary (AoI 2)
$L_{\omega1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I


Infinitary (AoI 2)
Darwininan "Canine" AI
$L_{\omega 1,\omega}$
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I


Infinitary (AoI 2)
$L_{\omega 1,\omega}$
"Monkey" AI
MiniMaxularity
Logic
FOL
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
Art of Infallibility I
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output



Infinitary (AoI 2)
$L_{\omega 1,\omega}$
"Full-Watson" AI
MiniMaxularity
FOL
Logic
SOL
epistemic
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
Art of Infallibility I


Infinitary (AoI 2)
Person-Aspiring AI
$L_{\omega 1,\omega}$
MiniMaxularity
FOL
Logic
epistemic
SOL
temporal
heterogeneous/visual
temporal+epistemic
temporal+epistemic+deontic
+planning+arg semantics
propostional logic
semantic-web logics
description logics
fragments of FOL
UIMA output
...
Art of Infallibility I