

# **The Immaterial Paradise, Motivating Paradoxes, Puzzles, and $R$ , Part II**

**Selmer Bringsjord**

*Intro to Formal Logic (and AI) = IFLAII*

I/13/25



Most AI Punditry is Illogical,  
e.g. ...

# Without Consciousness, AIs Will Be Sociopaths

ChatGPT can carry on a conversation, but the most important goal for artificial intelligence is making it understand what it means to have a mind

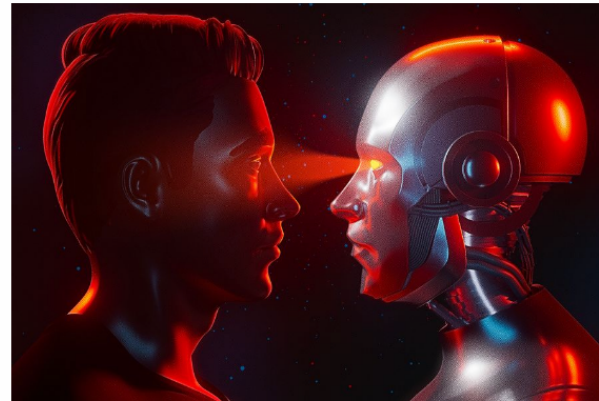
ESSAY

By Michael S.A. Graziano [+ Follow](#)

January 13, 2023 09:24 a.m. EST

ChatGPT, the latest technological [sensation](#), is an artificial intelligence chatbot with an amazing ability to carry on a conversation. It relies on a massive network of artificial neurons that loosely mimics the human brain, and it has been trained by analyzing the information resources of the internet. ChatGPT has processed more text than any human is likely to have read in a lifetime, allowing it to respond to questions fluently and even to imitate specific individuals, answering queries the way it thinks they would. My teenage son recently used ChatGPT to argue about politics with an imitation Karl Marx.

As a neuroscientist specializing in the brain mechanisms of con-



sciousness, I find talking to chatbots an unsettling experience. Are they conscious? Probably not. But given the rate of technological improvement, will they be in the next couple of years? And how would we even know?

Figuring out whether a machine has or understands humanlike consciousness is more than just a science-fiction hypothetical. Artificial intelligence is growing so powerful, so quickly, that it could soon pose a danger to human beings. We're building machines that are smarter than us and giving them control over our world. How can we

build AI so that it's aligned with human needs, not in conflict with us?

As counterintuitive as it may sound, creating a benign AI may require making it more conscious, not less. One of the most common misunderstandings about AI is the notion that if it's intelligent then it must be conscious, and if it is conscious then it will be autonomous, capable of taking over the world. But as we learn more about consciousness, those ideas do not appear to be correct. An autonomous system that makes complex decisions doesn't require consciousness.

What's most important about consciousness is that, for human beings, it's not just about the self. We see it in ourselves, but we also perceive it or project it into the world around us. Consciousness is part of the tool kit that evolution

gave us to make us an empathetic, prosocial species. Without it, we would necessarily be sociopaths, because we'd lack the tools for prosocial behavior. And without a concept of what consciousness is or an understanding that other beings have it, machines are sociopaths.

The only diagnostic tool for machine consciousness that we have right now is the Turing test, a thought experiment named for the British computer scientist Alan Turing. In its most common version, the test says that if a person holds a conversation with a machine and mistakes its responses for those of a real human being, then the machine must be considered effectively conscious.

The Turing test is an admission that the consciousness of another being is something we can only judge from the outside, based on the way he, she or it communicates. But the

# Without Consciousness, AIs Will Be Sociopaths

ChatGPT can carry on a conversation, but the most important goal for artificial intelligence is making it understand what it means to have a mind

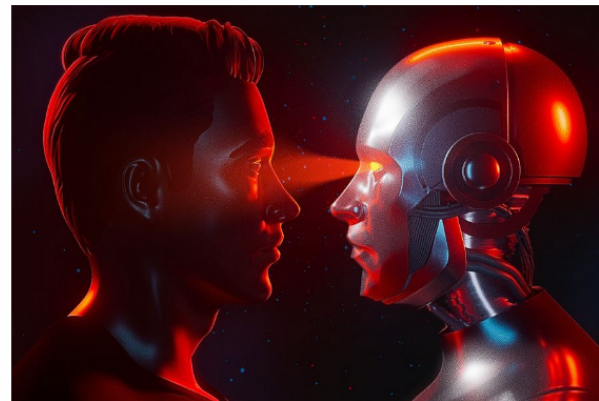
ESSAY

By Michael S.A. Graziano [+ Follow](#)

January 13, 2023 09:24 a.m. EST

ChatGPT, the latest technological [sensation](#), is an artificial intelligence chatbot with an amazing ability to carry on a conversation. It relies on a massive network of artificial neurons that loosely mimics the human brain, and it has been trained by analyzing the information resources of the internet. ChatGPT has processed more text than any human is likely to have read in a lifetime, allowing it to respond to questions fluently and even to imitate specific individuals, answering queries the way it thinks they would. My teenage son recently used ChatGPT to argue about politics with an imitation Karl Marx.

As a neuroscientist specializing in the brain mechanisms of con-



sciousness, I find talking to chatbots an unsettling experience. Are they conscious? Probably not. But given the rate of technological improvement, will they be in the next couple of years? And how would we even know?

Figuring out whether a machine has or understands humanlike consciousness is more than just a science-fiction hypothetical. Artificial intelligence is growing so powerful, so quickly, that it could soon pose a danger to human beings. We're building machines that are smarter than us and giving them control over our world. How can we

build AI so that it's aligned with human needs, not in conflict with us?

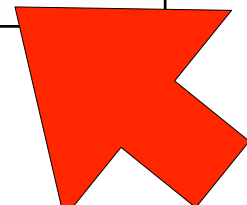
As counterintuitive as it may sound, creating a benign AI may require making it more conscious, not less. One of the most common misunderstandings about AI is the notion that if it's intelligent then it must be conscious, and if it is conscious then it will be autonomous, capable of taking over the world. But as we learn more about consciousness, those ideas do not appear to be correct. An autonomous system that makes complex decisions doesn't require consciousness.

What's most important about consciousness is that, for human beings, it's not just about the self. We see it in ourselves, but we also perceive it or project it into the world around us. Consciousness is part of the tool kit that evolution

gave us to make us an empathetic, prosocial species. Without it, we would necessarily be sociopaths, because we'd lack the tools for prosocial behavior. And without a concept of what consciousness is or an understanding that other beings have it, machines are sociopaths.

The only diagnostic tool for machine consciousness that we have right now is the Turing test, a thought experiment named for the British computer scientist Alan Turing. In its most common version, the test says that if a person holds a conversation with a machine and mistakes its responses for those of a real human being, then the machine must be considered effectively conscious.

The Turing test is an admission that the consciousness of another being is something we can only judge from the outside, based on the way he, she or it communicates. But the





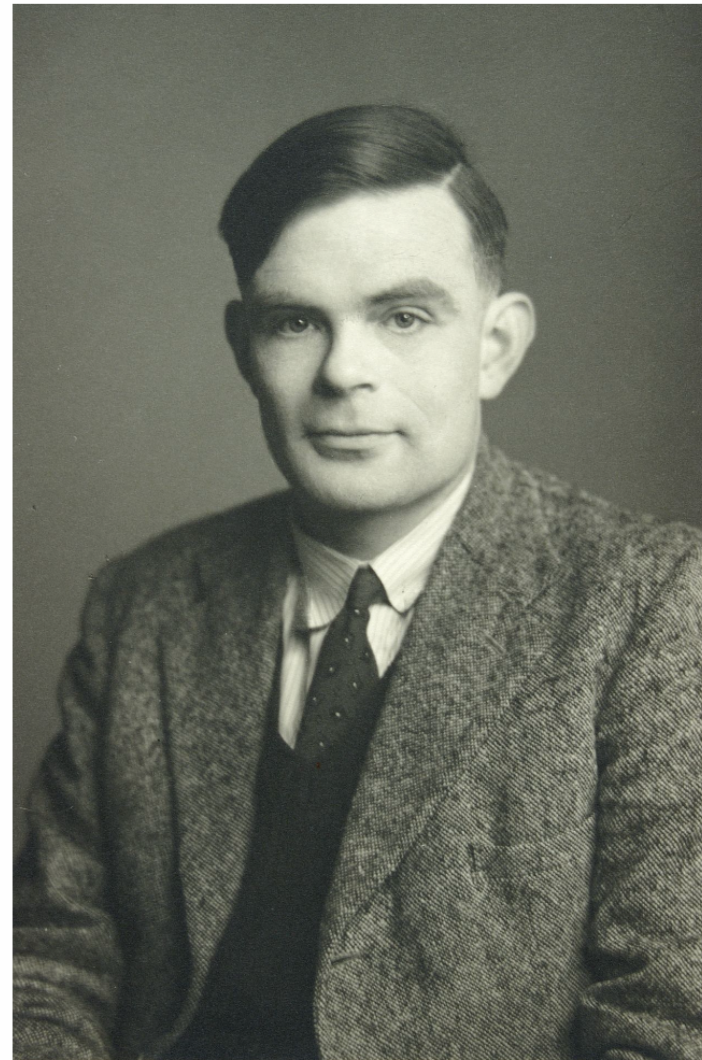
limits of the test are painfully obvious. After all, a pet dog can't carry on a conversation and pass as a human—does that mean it's not conscious? If you really wanted a machine to pass the test, you could have it say a few words to a small child. It might even fool some adults, too.

The truth is, the Turing test doesn't reveal much about what's going on inside a machine or a computer program like ChatGPT. Instead, what it really tests is the social cognition of the human participant. We evolved as social animals, and our brains instinctively project consciousness, agency, intention and emotion onto the objects around us. We're primed to see a world suffused with minds. Ancient animistic beliefs held that every river and tree had a spirit in it. For a similar reason, people are prone to see faces in random objects like the moon and moldy toast.

The original test proposed by Alan Turing in a 1950 paper was more complicated than the version people talk about today. Notably,

Computer science pioneer Alan Turing in 1951.

Turing didn't say a word about consciousness; he never delved into whether the machine had a subjective experience. He asked only whether it could think like a person. Turing imagined an “imitation



game” in which the player must determine the sex of two people, A and B. One is a man and one is a woman, but the player can't see them and can learn about them only by exchanging typed questions and

answers. A responds to the questions deceitfully, and wins the game if the player misidentifies their sex, while B answers truthfully and wins if the player identifies their sex correctly. Turing's idea was that if A or B is replaced by a machine, and the machine can win the game as often as a real person, then it must have mastered the subtleties of human thinking—of argument, manipulation and guessing what other people are thinking.

Turing's test was so complicated that people who popularized his work soon streamlined it into a single machine conversing with a single person. But the whole point of the original test was its bizarre complexity. Social cognition is difficult and requires a theory of mind—that is, a knowledge that other people have minds and an ability to guess what might be in them.

If we want to know whether a computer is conscious, then, we need to test whether the computer understands how conscious minds interact. In other words, we need a

 Back

reverse Turing test: Let's see if the computer can tell whether it's talking to a human or another computer. If it can tell the difference, then maybe it knows what consciousness is. ChatGPT definitely can't pass that test yet: It doesn't know whether it's responding to a living person with a mind or a disjointed list of prefab questions.

A sociopathic machine that can make consequential decisions would be powerfully dangerous. For now, chatbots are still limited in their abilities; they're essentially toys. But if we don't think more deeply about machine consciousness, in a year or five years we may face a crisis. If computers are going to outthink us anyway, giving them more human-like social cognition might be our best hope of aligning them with human values.

*Dr. Graziano is a professor of psychology and neuroscience at Princeton University and the author of "Rethinking Consciousness: A Scientific Theory of Subjective Experience."*

# Logically Invalid Reasoning

Turing in his famous paper claims:

**TT:** If  $x$  is conversationally indistinguishable from a human, then  $x$  is conscious/sentient.

# Logically Invalid Reasoning

Turing in his famous paper claims:

**TT**: If  $x$  is conversationally indistinguishable from a human, then  $x$  is conscious/sentient.

This proposition is logically equivalent to:

**TT'**: If  $x$  is not conversationally indistinguishable from a human, then  $x$  is not conscious/sentient.



# Logically Invalid Reasoning

Turing, in his famous paper, claims:

**TT**: If  $x$  is conversationally indistinguishable from a human, then  $x$  is conscious/sentient.

This proposition is logically equivalent to:

**TT'**: If  $x$  is not conversationally indistinguishable from a human, then  $x$  is not conscious/sentient.

Well then, obviously Turing is wrong: A pet dog isn't conversationally indistinguishable from a human, but the dog *is* conscious/sentient. We have a counter-example!

# Logically Invalid Reasoning

Turing, in his famous paper, claims:

**TT**: If  $x$  is conversationally indistinguishable from a human, then  $x$  is conscious/sentient.

This proposition is logically equivalent to:

**TT'**: If  $x$  is not conversationally indistinguishable from a human, then  $x$  is not conscious/sentient.

Well then, obviously Turing is wrong: A pet dog isn't conversationally indistinguishable from a human, but the dog *is* conscious/sentient. We have a counter-example!

# Logically Invalid Reasoning

Turing, in his famous paper, claims:

**TT**: If  $x$  is conversationally indistinguishable from a human, then  $x$  is conscious/sentient.

$$i \rightarrow c$$

This proposition is logically equivalent to:

**TT'**: If  $x$  is not conversationally indistinguishable from a human, then  $x$  is not conscious/sentient.

$$\neg i \rightarrow \neg c$$

Well then, obviously Turing is wrong: A pet dog isn't conversationally indistinguishable from a human, but the dog *is* conscious/sentient. We have a counter-example!

$$\neg i \wedge c$$

$$\{\neg i \wedge c\} \vdash \neg(\neg i \rightarrow \neg c) \quad \checkmark$$

# Logically Invalid Reasoning

Turing, in his famous paper, claims:

**TT**: If  $x$  is conversationally indistinguishable from a human, then  $x$  is conscious/sentient.

$$i \rightarrow c$$

This proposition is logically equivalent to:

**TT'**: If  $x$  is not conversationally indistinguishable from a human, then  $x$  is not conscious/sentient.

$$\neg i \rightarrow \neg c$$

**X**

Well then, obviously Turing is wrong: A pet dog isn't conversationally indistinguishable from a human, but the dog *is* conscious/sentient. We have a counter-example!

$$\neg i \wedge c$$

$$\{\neg i \wedge c\} \vdash \neg(\neg i \rightarrow \neg c) \quad \checkmark$$



# How Does ChatGP4o Do?

(Logically Invalid Reasoning)

$i \rightarrow c$

**X**



$\neg i \rightarrow \neg c$

$\neg i \wedge c$


$\{ \neg i \wedge c \} \vdash \neg(\neg i \rightarrow \neg c) \quad \checkmark$

# How Does ChatGP4o Do?

## (Logically Invalid Reasoning)

 ChatGPT 4o > 

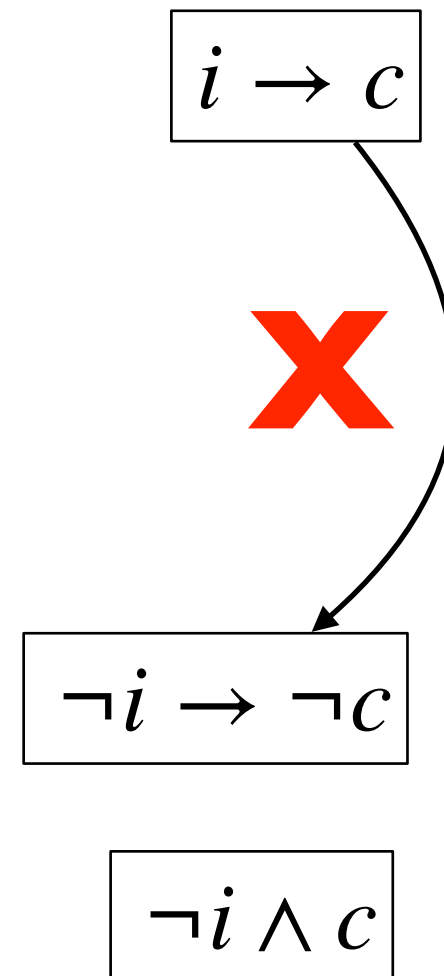
I just read Turing's famous paper, the one about the test.



That's fantastic! Turing's 1950 paper, *"Computing Machinery and Intelligence,"* is a cornerstone of artificial intelligence and philosophy of mind. It's where he proposes the famous "Imitation Game," now commonly called the Turing Test, to address the question, "Can machines think?"

What stood out to you in the paper? Was it the test itself, his predictions about the future of AI, or perhaps his philosophical arguments about machine intelligence?

Ah so it's called the 'Turing Test' now, huh?



$$\{ \neg i \wedge c \} \vdash \neg(\neg i \rightarrow \neg c) \quad \checkmark$$

# How Does ChatGP4o Do?

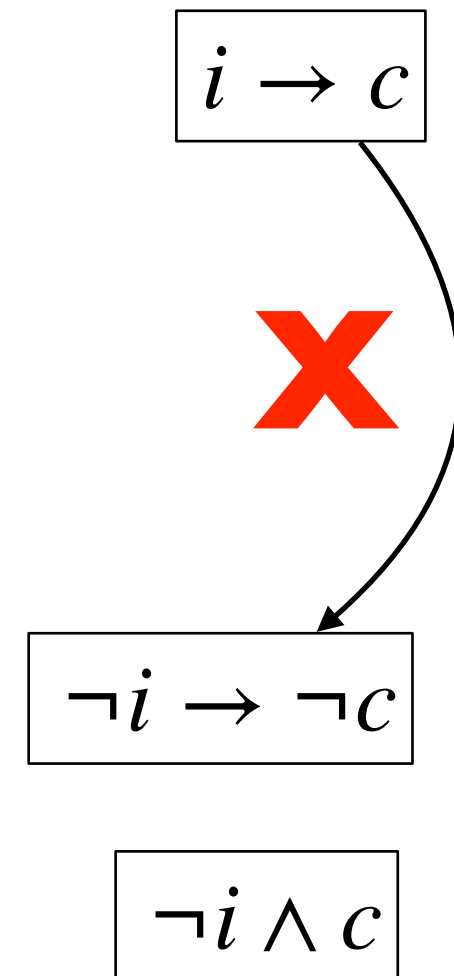
## (Logically Invalid Reasoning)

ChatGPT 4o >

Yes, it's now widely referred to as the "Turing Test"! The term evolved over time as the ideas in Turing's paper gained influence. In the paper itself, he called it the "Imitation Game," describing a scenario where an interrogator tries to determine which of two hidden participants is a human and which is a machine, based solely on their responses to questions.

Calling it the "Turing Test" highlights his monumental contribution to thinking about artificial intelligence. Does the modern framing of the test feel different from how Turing originally described it?

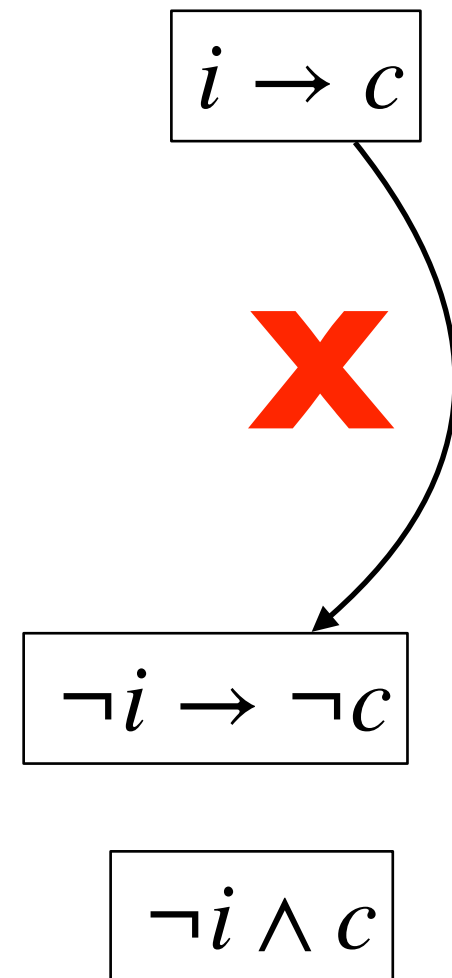
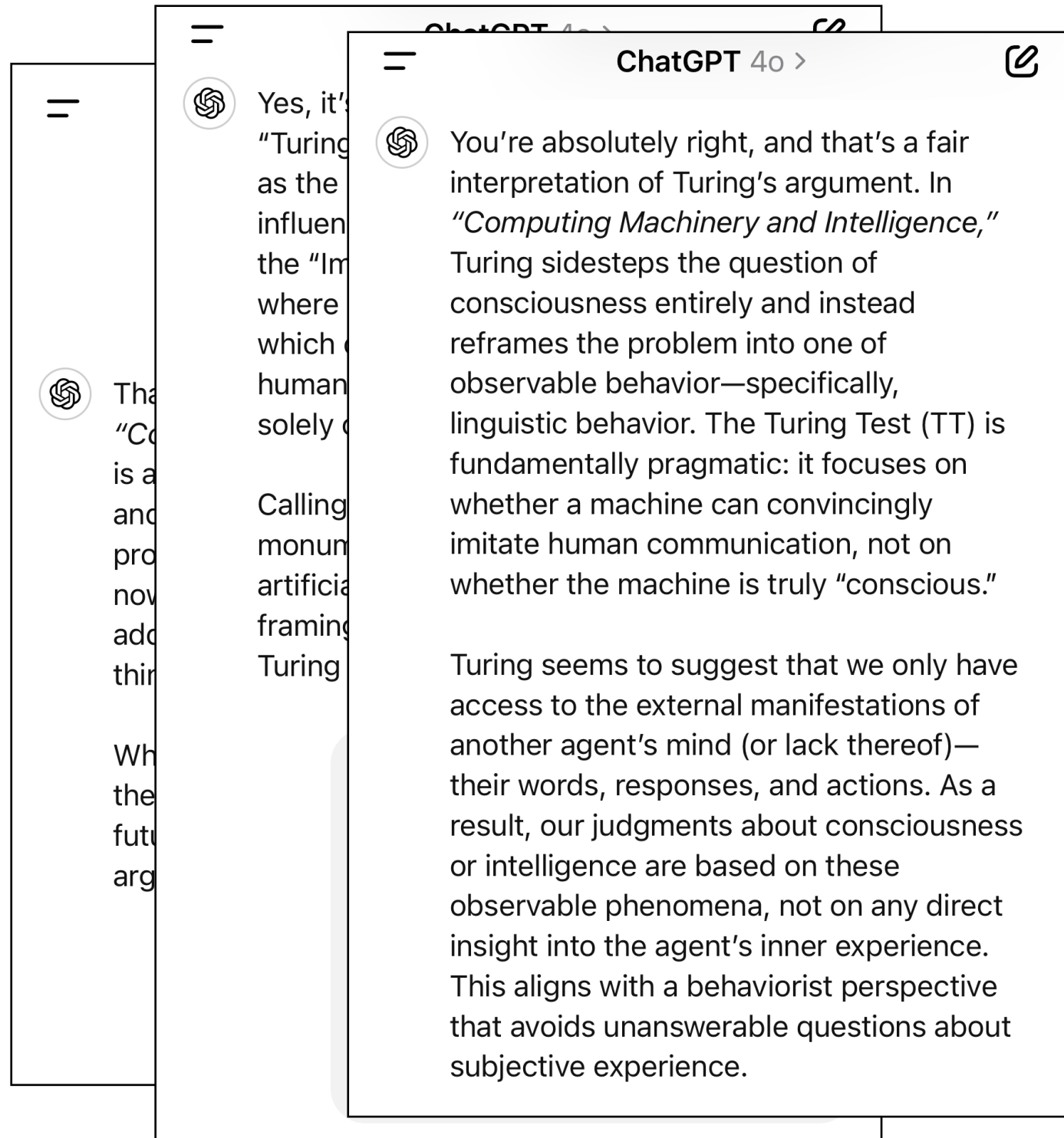
It seems to me on a fair reading of the paper that the TT (that's how I'll label the test) overall says — & Turing would thus be understood to say this — that whether another agent is conscious is something we can only make a judgement about based on that agent's communication with us.



$$\{ \neg i \wedge c \} \vdash \neg(\neg i \rightarrow \neg c) \quad \checkmark$$

# How Does ChatGP4o Do?

## (Logically Invalid Reasoning)

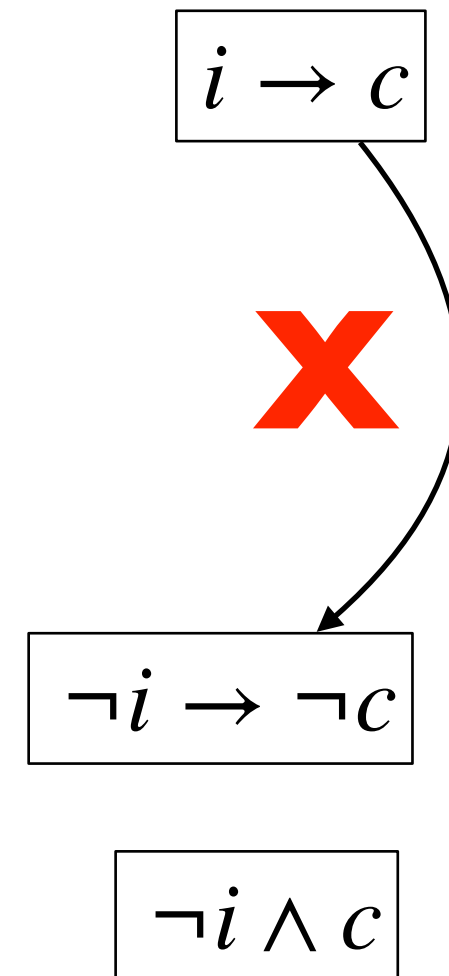
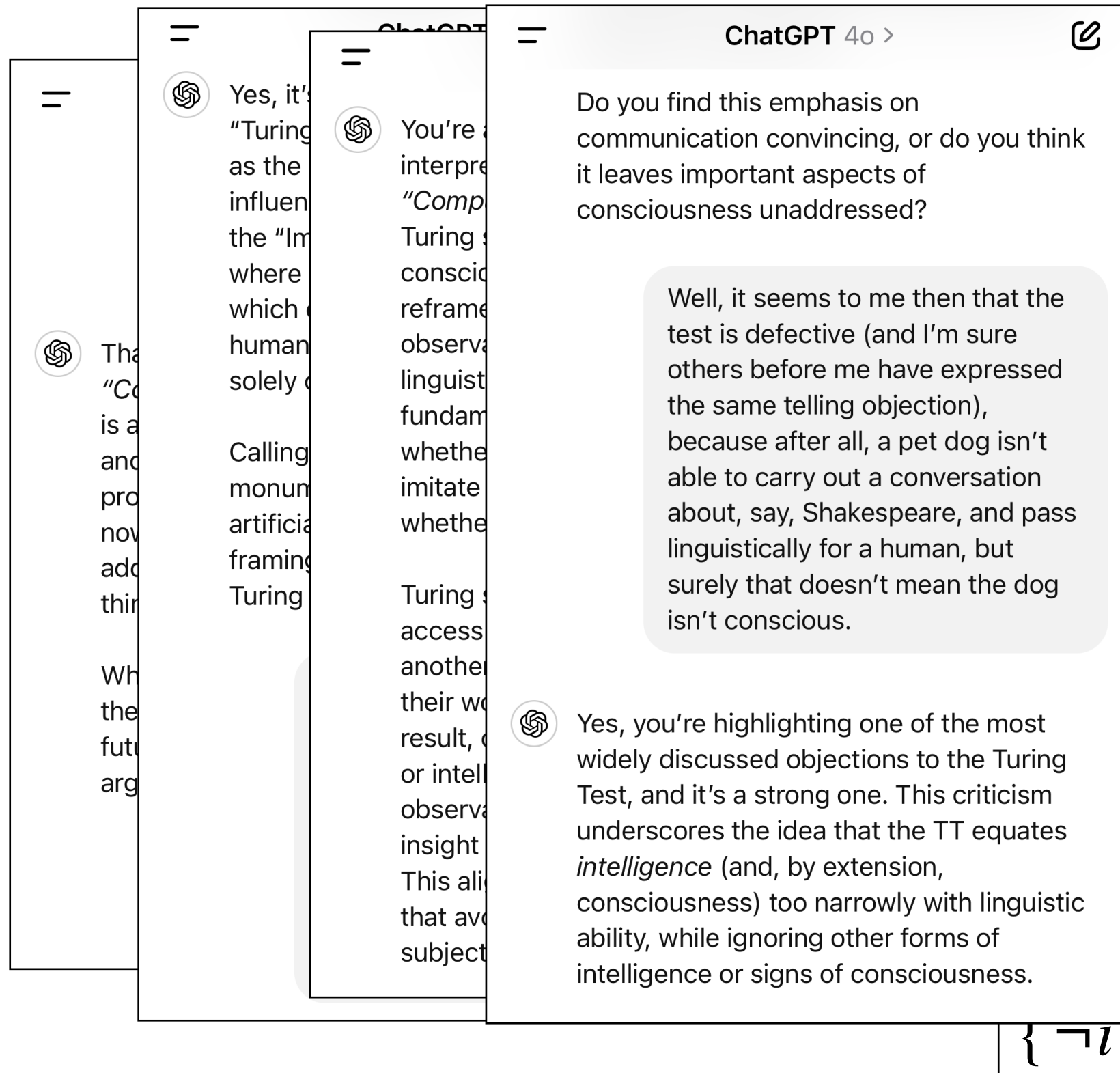


$$\{ \neg i \wedge c \} \vdash \neg(\neg i \rightarrow \neg c) \quad \checkmark$$



# How Does ChatGP4o Do?

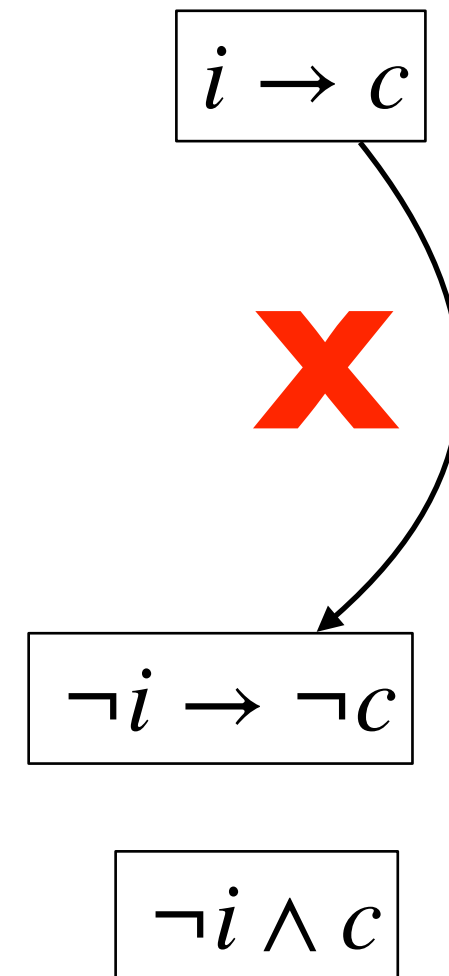
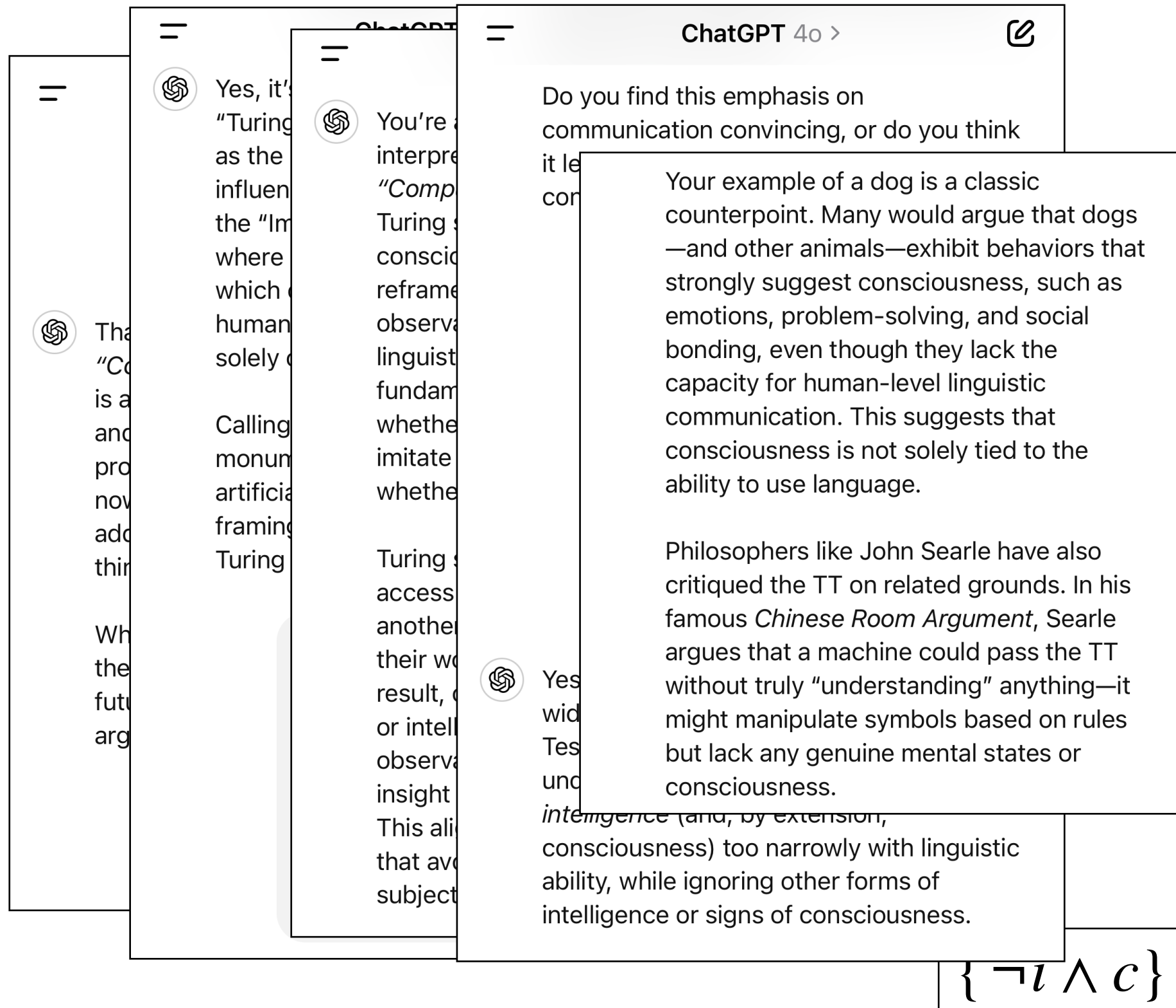
## (Logically Invalid Reasoning)



$$\boxed{\{\neg i \wedge c\} \vdash \neg(\neg i \rightarrow \neg c) \quad \checkmark}$$

# How Does ChatGP4o Do?

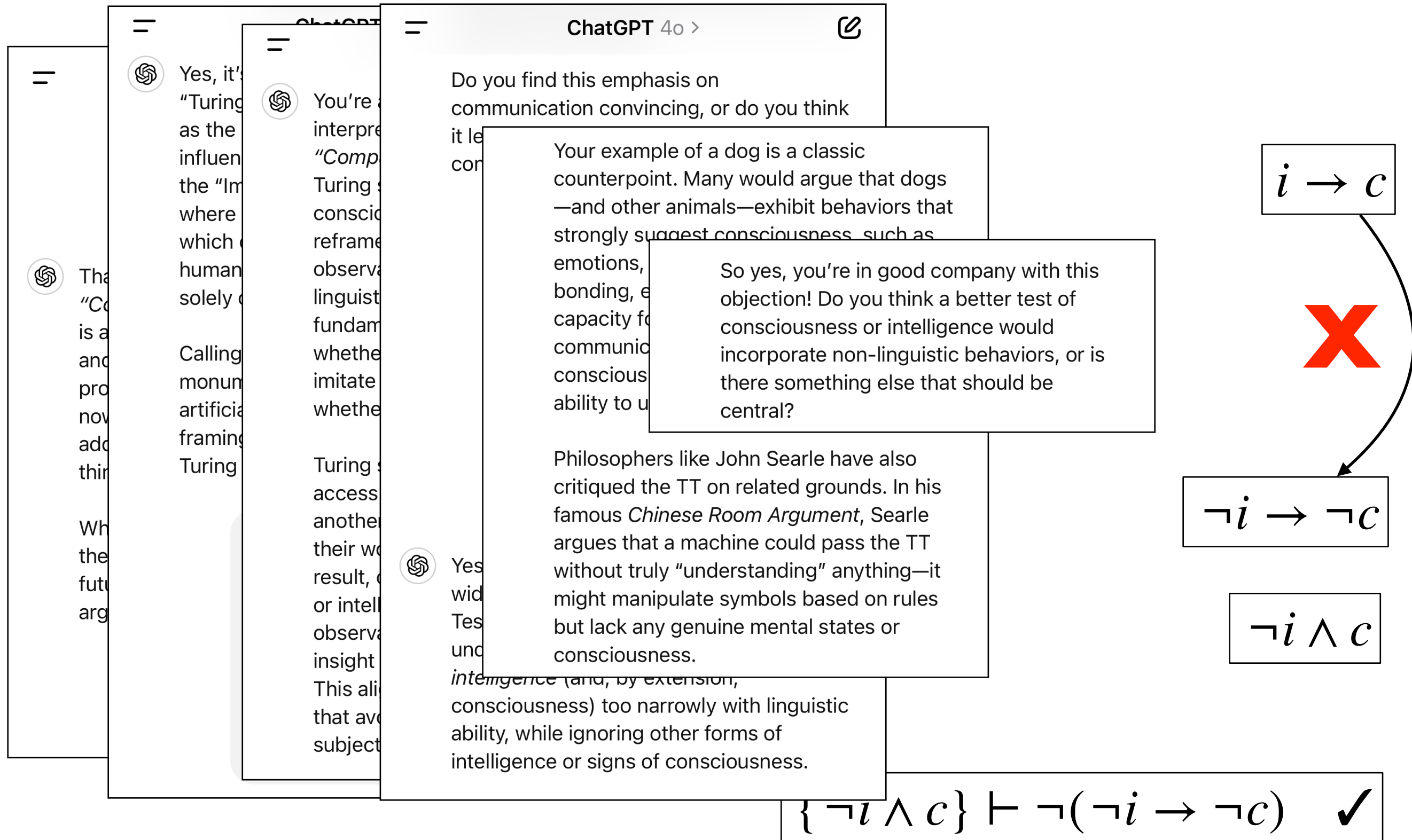
## (Logically Invalid Reasoning)



$$\boxed{\{\neg i \wedge c\} \vdash \neg(\neg i \rightarrow \neg c) \quad \checkmark}$$

# How Does ChatGP4o Do?

## (Logically Invalid Reasoning)



*Some* AI Punditry is Logical,  
e.g. ...



## *No, AI Machines Can't Think*

Computers win in realms with defined rules but we have free will and make choices.

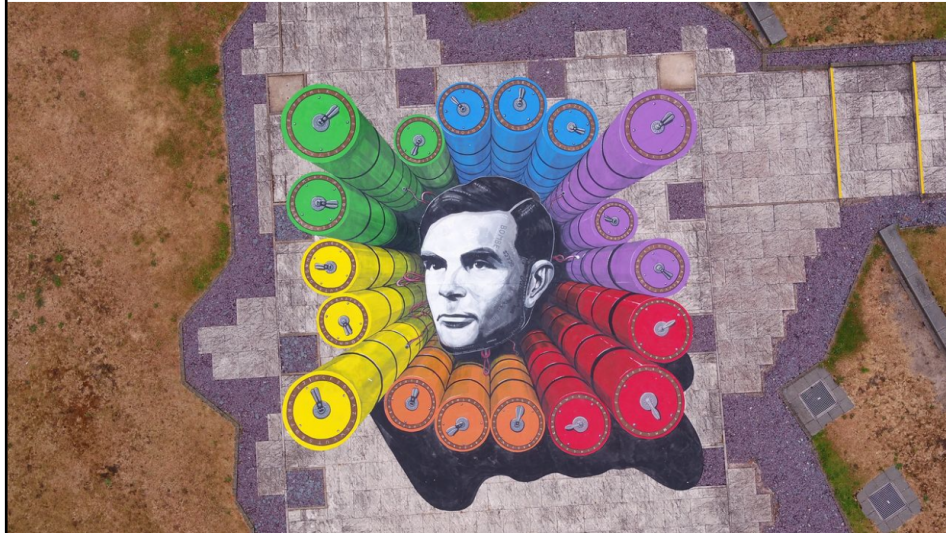


Photo: crown copyright 2021/Reuters

By *Andy Kessler* [+ Follow](#)

Jan 07, 2024 12:59 p.m. ET



**Listen to this article**

5 minutes

Be ready for it. Sometime soon, the digirati will declare that artificial-intelligence machines have passed the Turing test and thus the era of superintelligence and sentient computers has arrived. The promised land is artificial general intelligence: AGI. Don't fall for it. Your cranial cavity's inner voice and self-awareness explain why.

## No, AI Machines Can't Think

Computers win in realms with defined rules but we have free will and make choices.

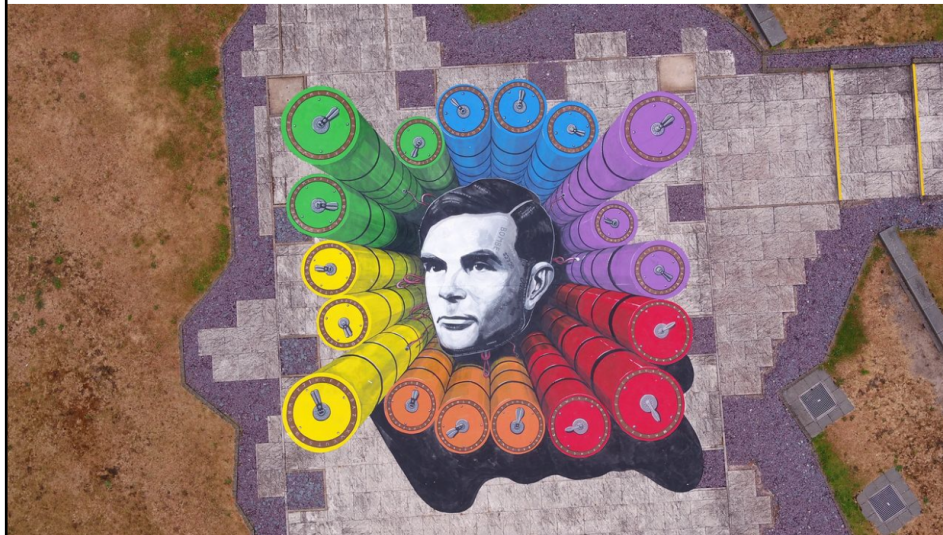


Photo: crown copyright 2021/Reuters

By **Andy Kessler** [+ Follow](#)

Jan 07, 2024 12:59 p.m. ET



**Listen to this article**

5 minutes

Be ready for it. Sometime soon, the digirati will declare that artificial-intelligence machines have passed the Turing test and thus the era of superintelligence and sentient computers has arrived. The promised land is artificial general intelligence: AGI. Don't fall for it. Your cranial cavity's inner voice and self-awareness explain why.

Provide a machine \$100,000 and see if it can turn it into \$1 million, perhaps using Alibaba and an e-commerce site. It's silly because by this measure a decent spambot is intelligent!

Ada Lovelace, the 19th-century mathematician, was ahead of her time. Of Charles Babbage's early computer, invented in 1837, she wrote a few years later: "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform." A Lovelace test might be, "Can AI generate something truly original?" And enjoyable, not the proverbial novel typed by a thousand monkeys. Same for movies, music and art. Rensselaer Polytechnic Institute professor [Selmer Bringsjord proposed](#) a "(Better) Lovelace" test in 2001 to detect creativity.



## *No, AI Machines Can't Think*

Computers win in realms with defined rules but we have free will and make choices.

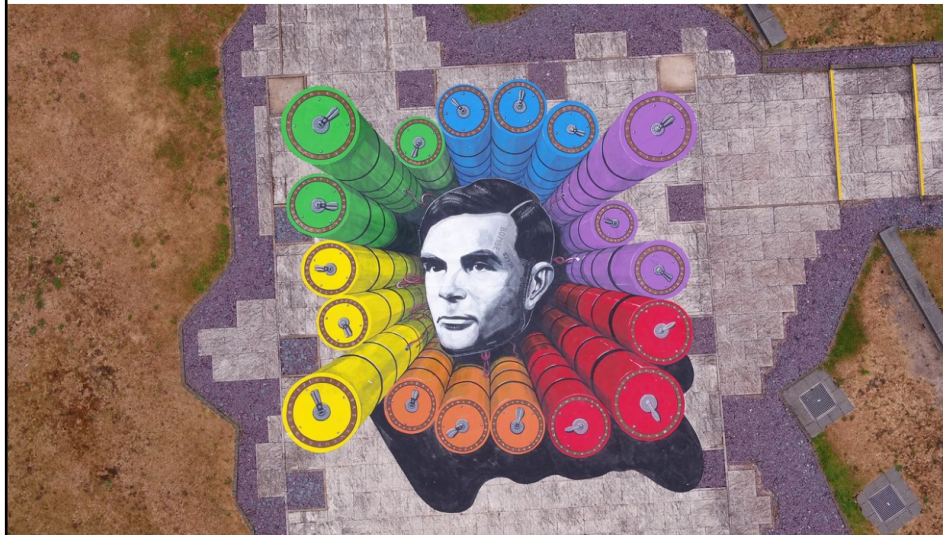


Photo: crown copyright 2021/Reuters

By *Andy Kessler* [+ Follow](#)

Jan 07, 2024 12:59 p.m. ET



**Listen to this article**

5 minutes

Be ready for it. Sometime soon, the digirati will declare that artificial-intelligence machines have passed the Turing test and thus the era of superintelligence and sentient computers has arrived. The promised land is artificial general intelligence: AGI. Don't fall for it. Your cranial cavity's inner voice and self-awareness explain why.

Provide a machine \$100,000 and see if it can turn it into \$1 million, perhaps using Alibaba and an e-commerce site. It's silly because by this measure a decent spambot is intelligent!

Ada Lovelace, the 19th-century mathematician, was ahead of her time. Of Charles Babbage's early computer, invented in 1837, she wrote a few years later: "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform." A Lovelace test might be, "Can AI generate something truly original?" And enjoyable, not the proverbial novel typed by a thousand monkeys. Same for movies, music and art. Rensselaer Polytechnic Institute professor [Selmer Bringsjord proposed](#) a "(Better) Lovelace" test in 2001 to detect creativity.

<http://kryten.mm.rpi.edu/lovelace.pdf>



# What *is* Logic?

# What *is* Logic?

- The key to becoming rational.



# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.

# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.

# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!

# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).

# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.

# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.

# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.



# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.
- The key to divining the meaning of life (and other such big questions).

# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.
- The key to divining the meaning of life (and other such big questions).
- The better way to program computers; and fundamentally the *only* way to *reliably* program computers.

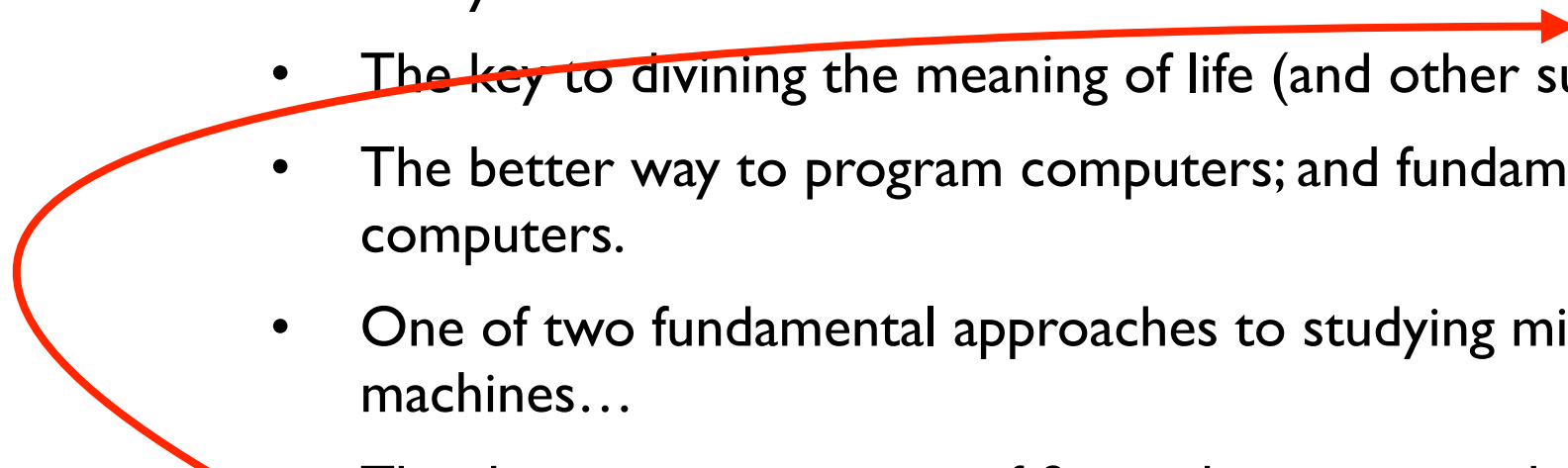
# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.
- The key to divining the meaning of life (and other such big questions).
- The better way to program computers; and fundamentally the *only* way to *reliably* program computers.
- One of two fundamental approaches to studying minds, and replicating/simulating minds in machines...

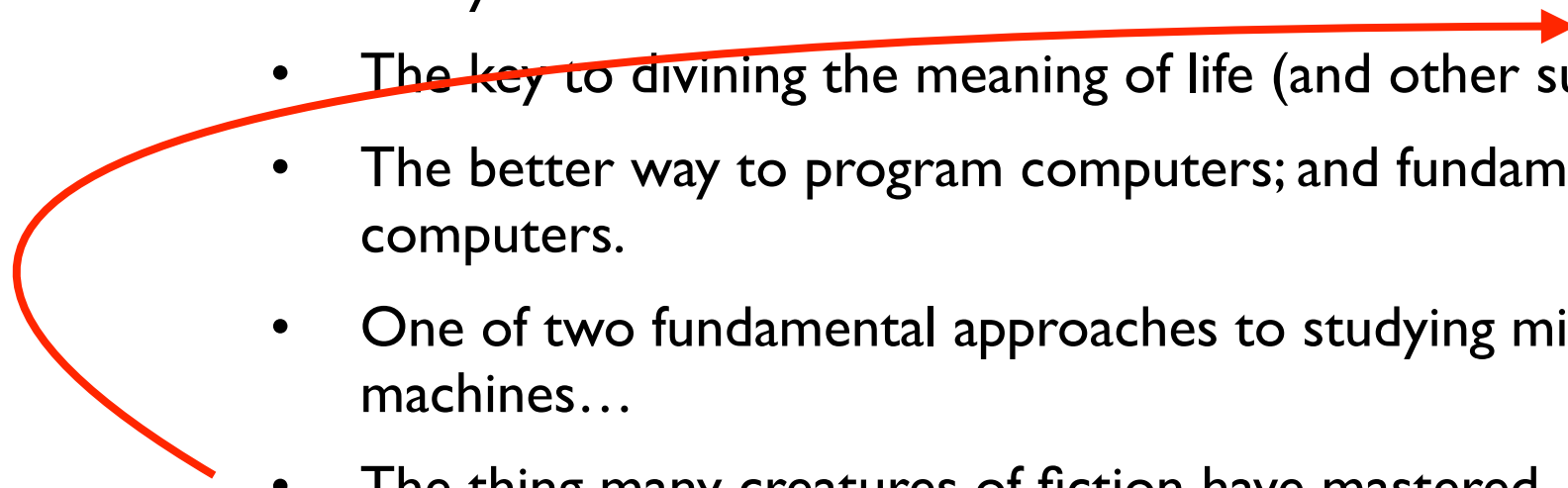
# What *is* Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.
- The key to divining the meaning of life (and other such big questions).
- The better way to program computers; and fundamentally the *only* way to *reliably* program computers.
- One of two fundamental approaches to studying minds, and replicating/simulating minds in machines...
- The thing many creatures of fiction have mastered — have you (as a New Yorker)?...

# What is Logic?

- The key to becoming rational.
  - “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
  - The only invincible subject there is.
  - The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
    - Engineering! Computer Science!
    - Mathematics itself: see “reverse mathematics”!
  - The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
  - The most challenging subject there is.
  - One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
  - A key to riches.
  - The key to divining the meaning of life (and other such big questions).
  - The better way to program computers; and fundamentally the *only* way to *reliably* program computers.
  - One of two fundamental approaches to studying minds, and replicating/simulating minds in machines...
  - The thing many creatures of fiction have mastered — have you (as a New Yorker)?...
- 

# What is Logic?

- The key to becoming rational.
  - “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
  - The only invincible subject there is.
  - The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
    - Engineering! Computer Science!
    - Mathematics itself: see “reverse mathematics”!
  - The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
  - The most challenging subject there is.
  - One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
  - A key to riches.
  - The key to divining the meaning of life (and other such big questions).
  - The better way to program computers; and fundamentally the *only* way to *reliably* program computers.
  - One of two fundamental approaches to studying minds, and replicating/simulating minds in machines...
  - The thing many creatures of fiction have mastered — have you (as a New Yorker)?...
  - ...
- 



# What is Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and n and you); and mindless machines (like Deep Blue & \
- A key to riches.
- ~~The key to divining the meaning of life (and other su~~
- The better way to program computers; and fundam computers.
- One of two fundamental approaches to studying mi machines...
- The thing many creatures of fiction have mastered — have you (as a New Yorker)?...
- ...



# What is Logic?

- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and n and you); and mindless machines (like Deep Blue & \
- A key to riches.
- ~~The key to divining the meaning of life (and other su~~
- The better way to program computers; and fundam computers.
- One of two fundamental approaches to studying mi machines...
- The thing many creatures of fiction have mastered — have you (as a New Yorker)?...
- ...





# What *is* Logic?

# What *is* Logic?

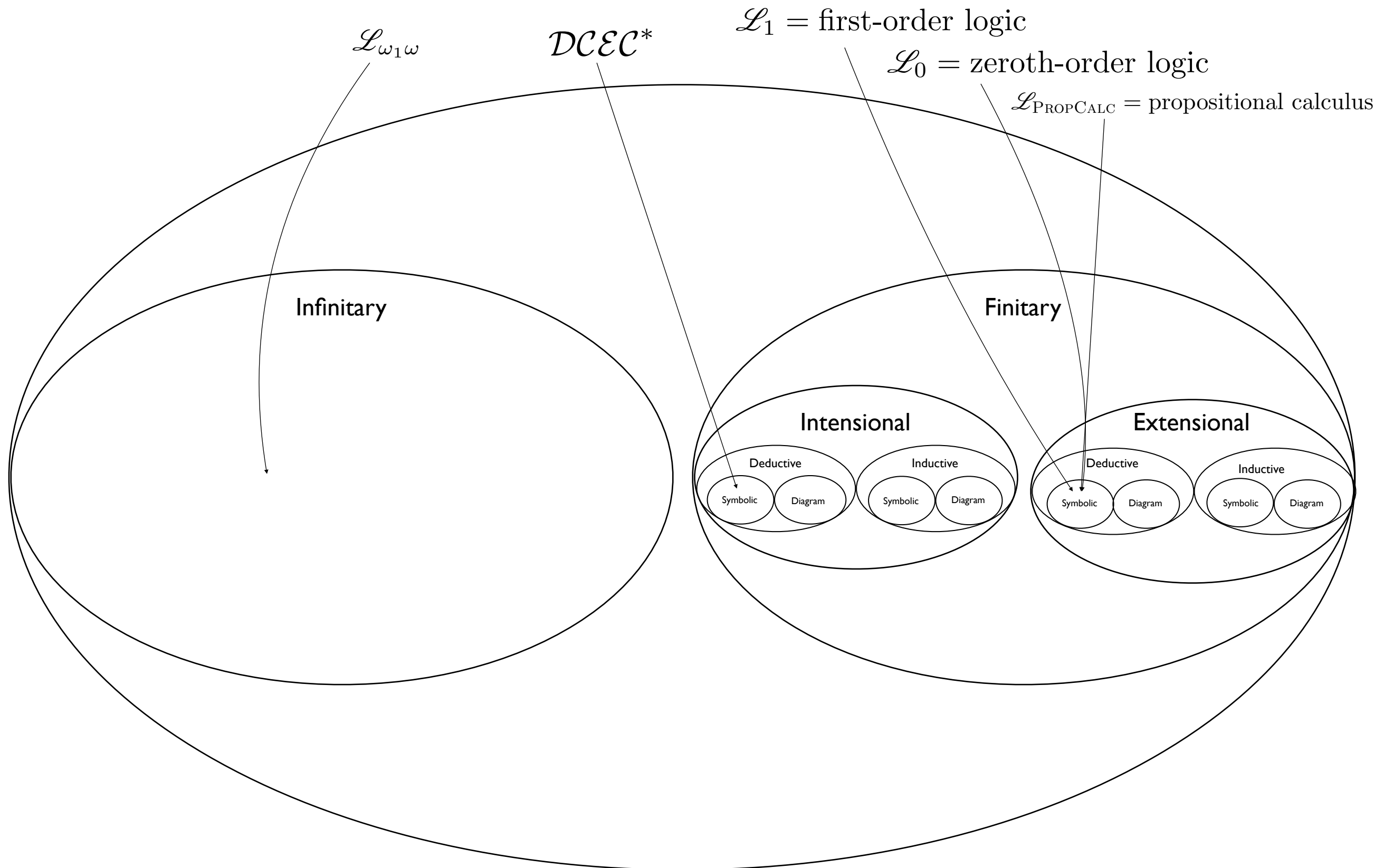
- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.
- The key to divining the meaning of life (and other such big questions).
- The better way to program computers; and fundamentally the *only* way to *reliably* program computers.
- One of two fundamental approaches to studying minds, and replicating/simulating minds in machines...
- The thing many creatures of fiction have mastered — have you (as a New Yorker)?...
- ...

# What *is* Logic?

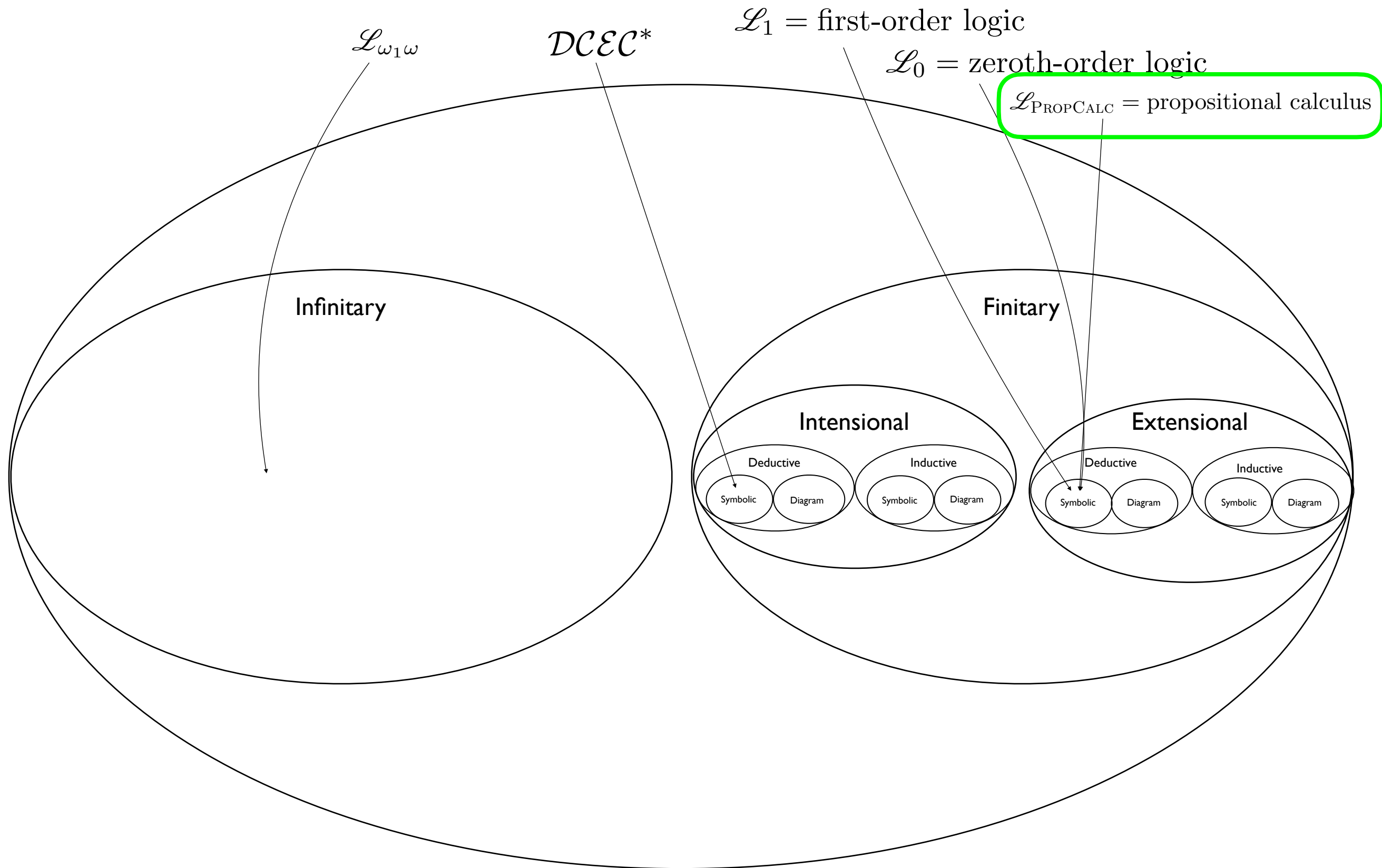
- The key to becoming rational.
- “The science and engineering of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences, e.g., ...
  - Engineering! Computer Science!
  - Mathematics itself: see “reverse mathematics”!
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.
- The key to divining the meaning of life (and other such big questions).
- The better way to program computers; and fundamentally the *only* way to *reliably* program computers.
- One of two fundamental approaches to studying minds, and replicating/simulating minds in machines...
- The thing many creatures of fiction have mastered — have you (as a New Yorker)?...
- ...



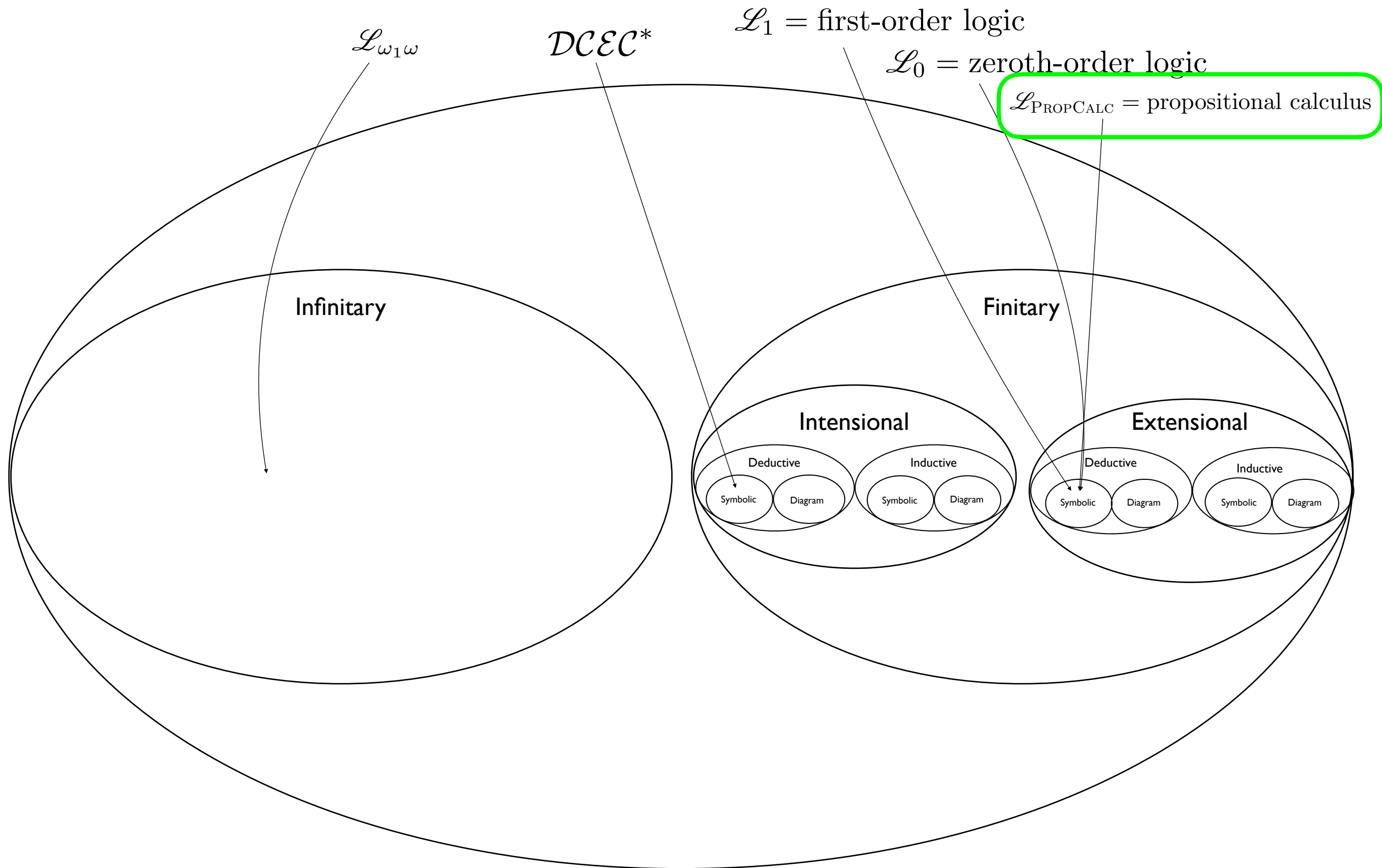
# The Universe of Logics



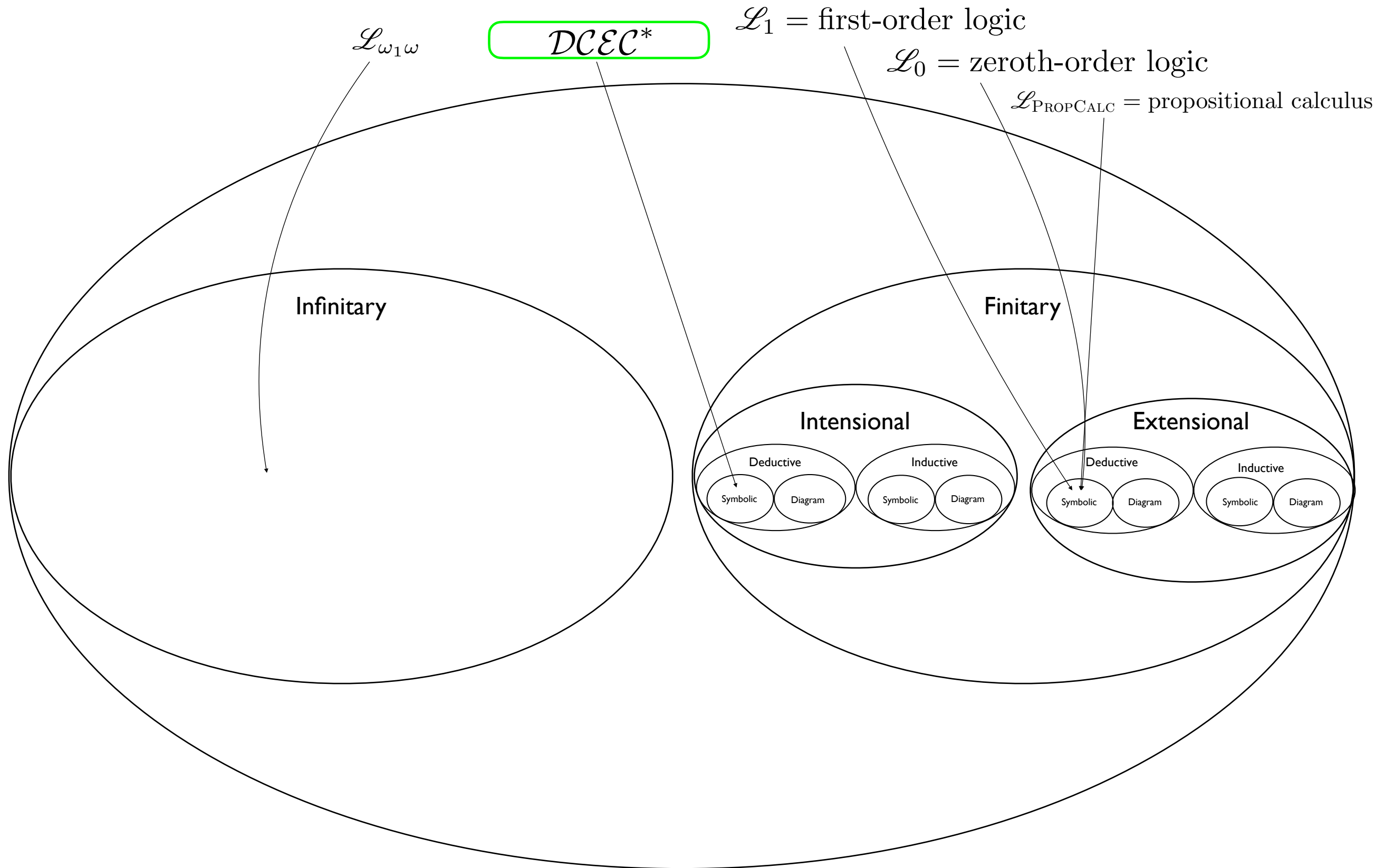
# The Universe of Logics



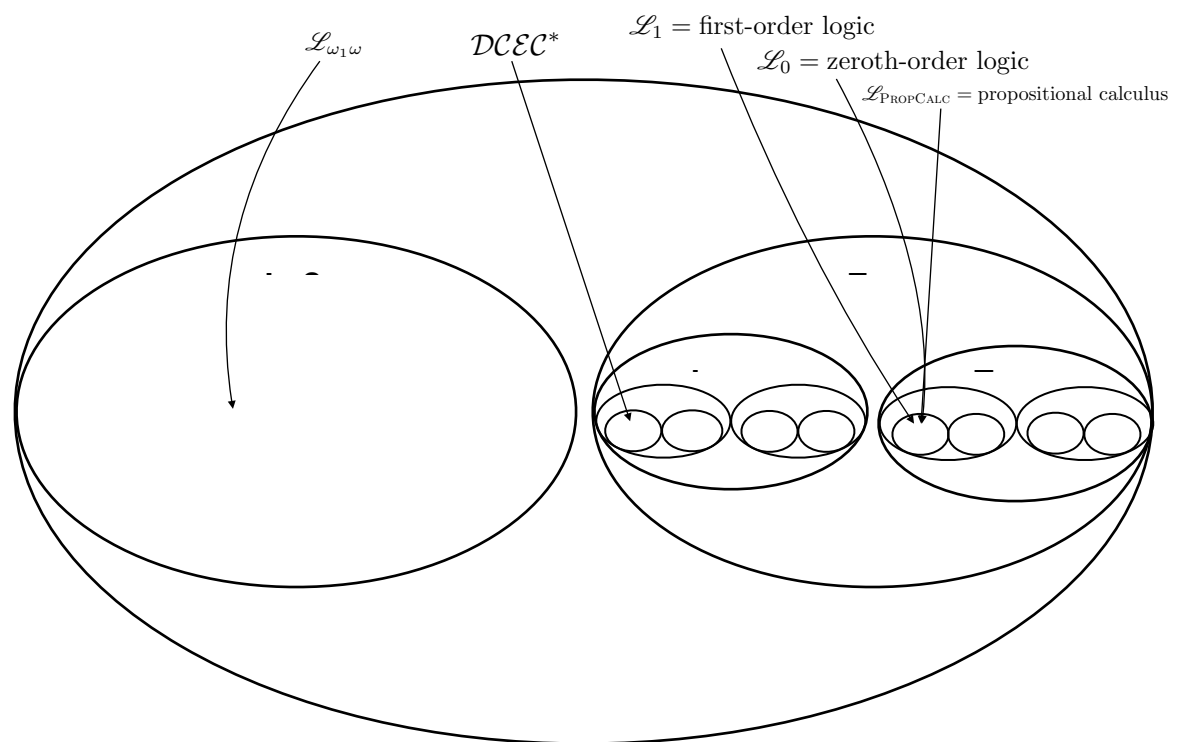
# The Universe of Logics



# The Universe of Logics

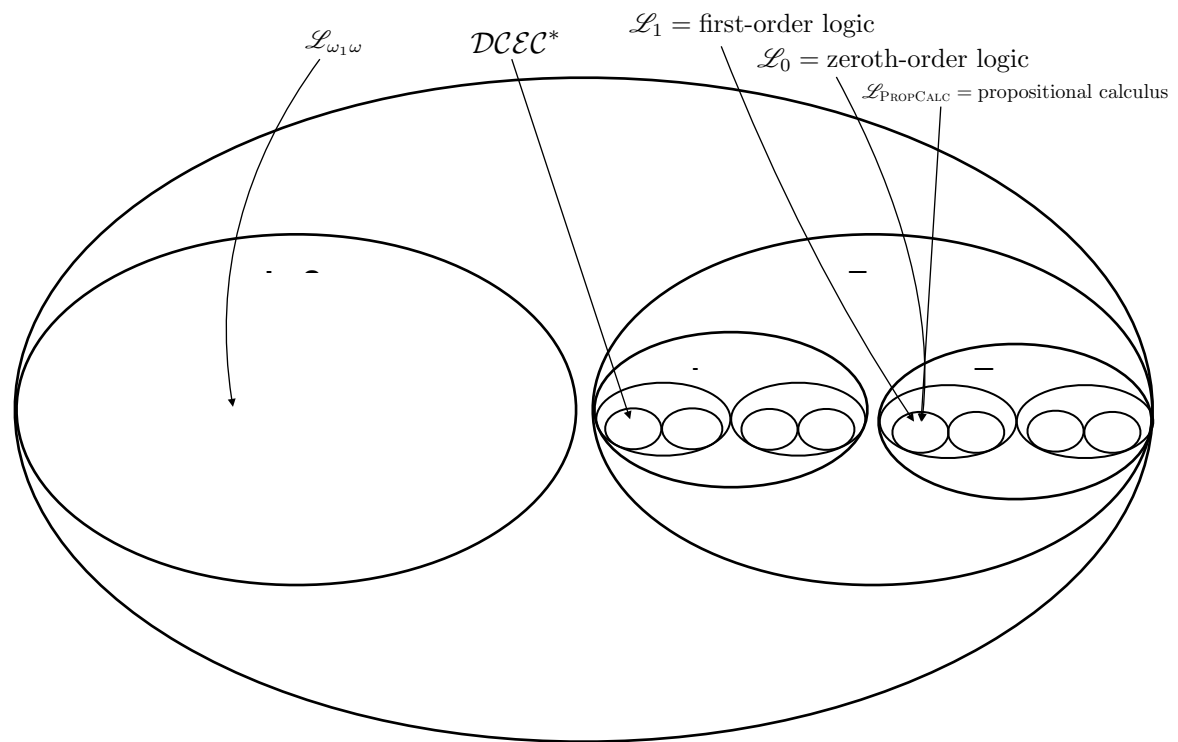


# The Universe of Logics



Non-Physical

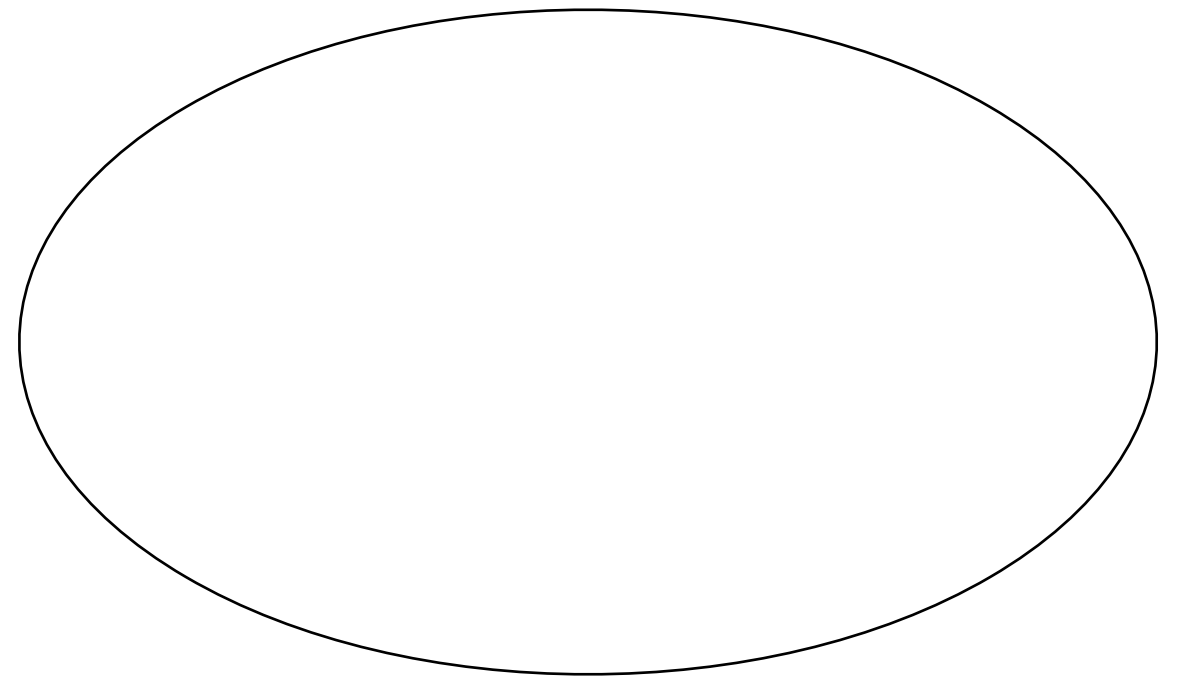
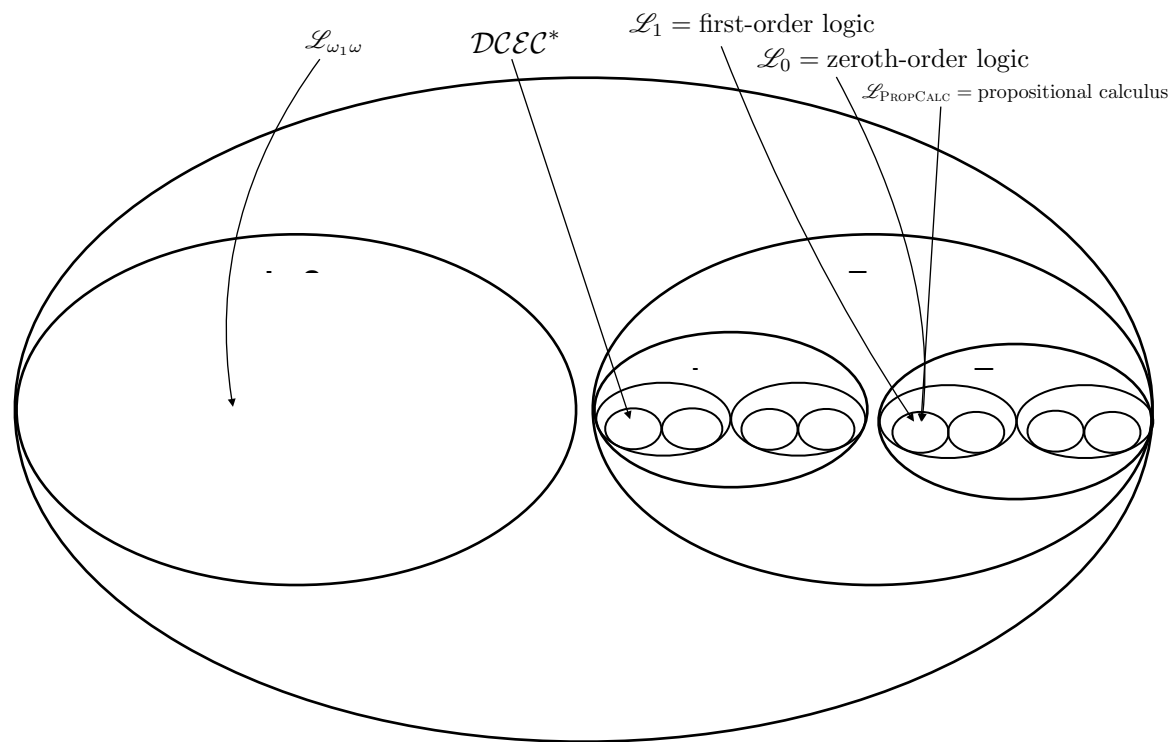
# The Universe of Logics



Non-Physical

# The Universe of Logics

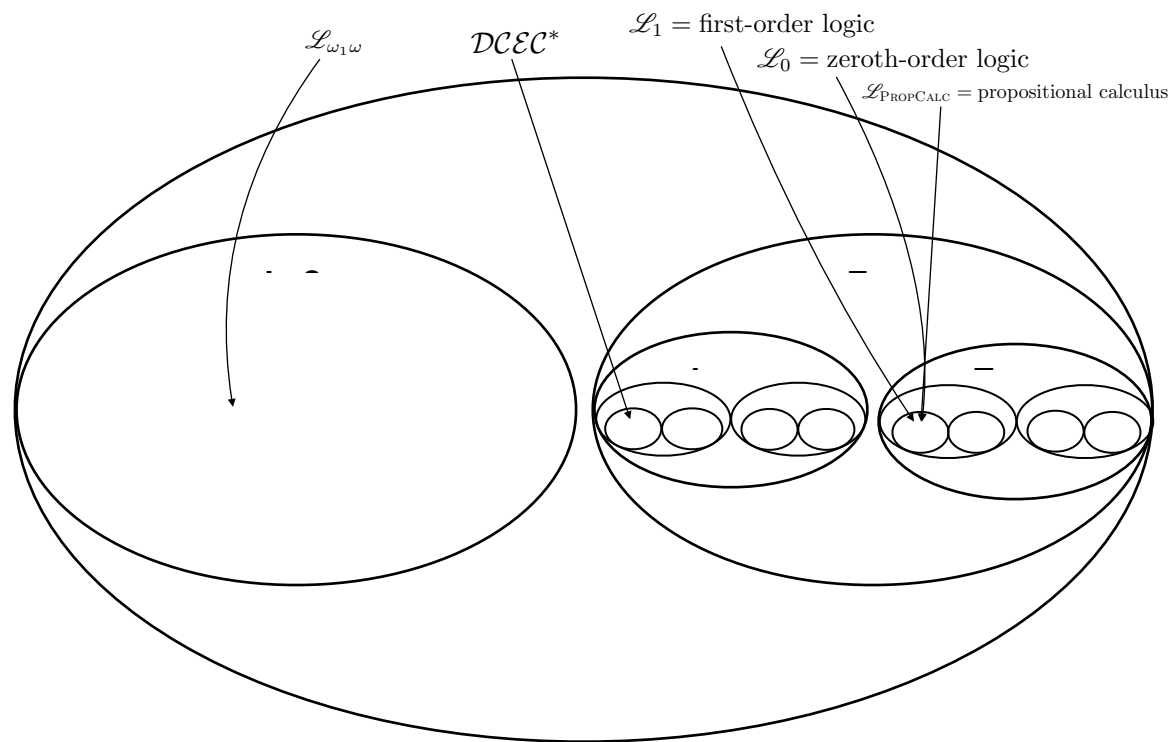
# The Physical Universe



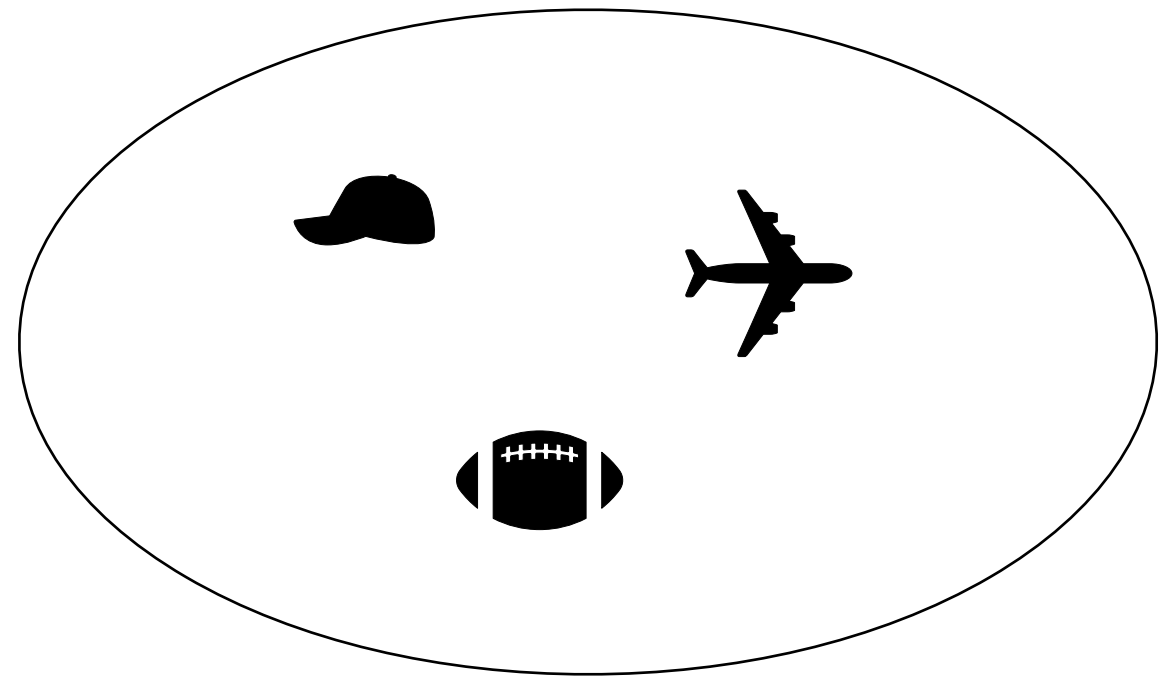


Non-Physical

# The Universe of Logics

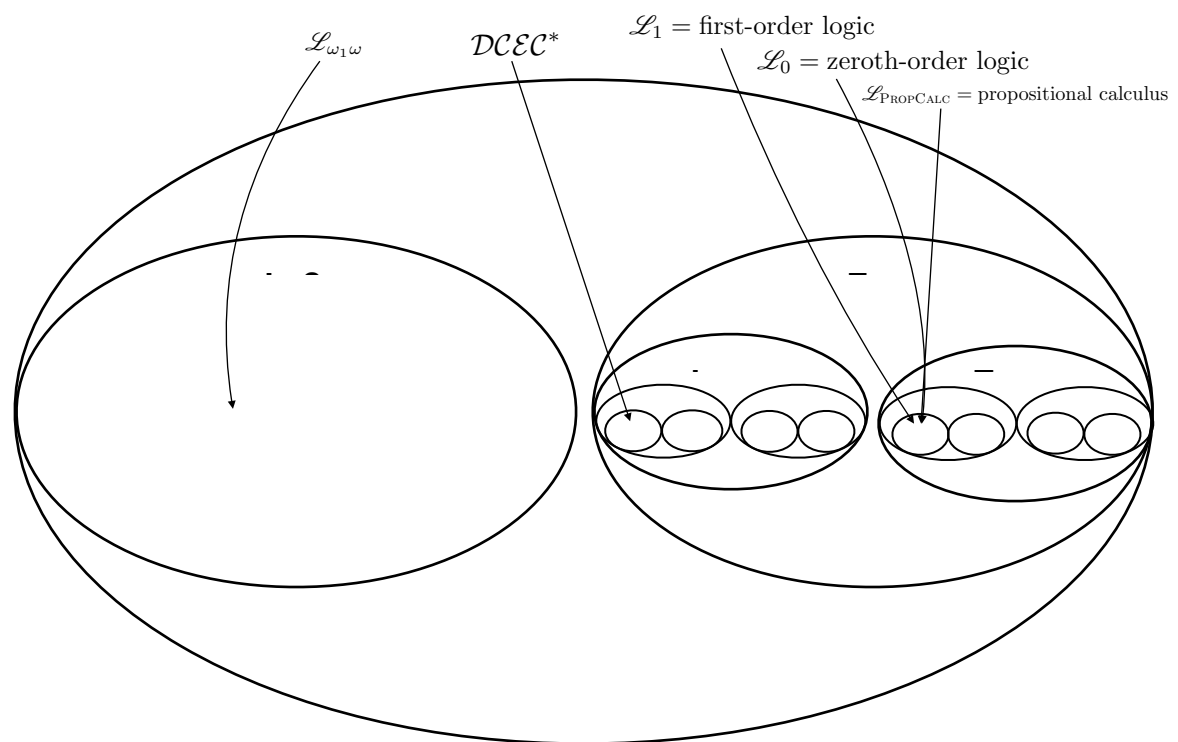


# The Physical Universe

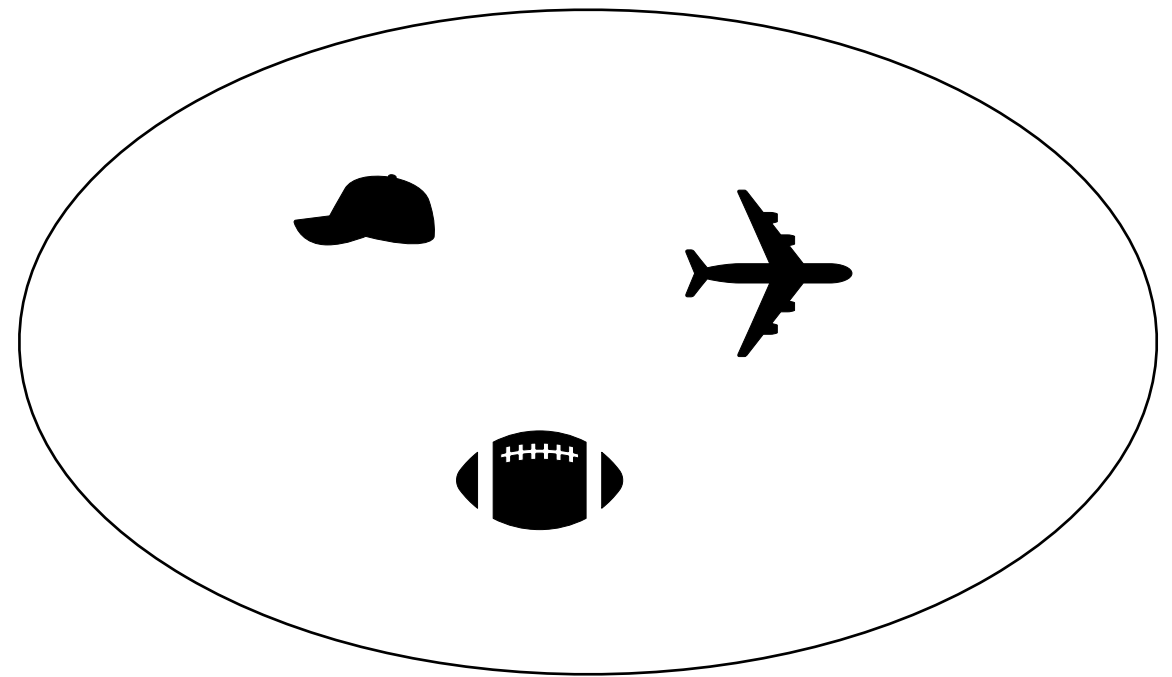


Non-Physical

# The Universe of Logics



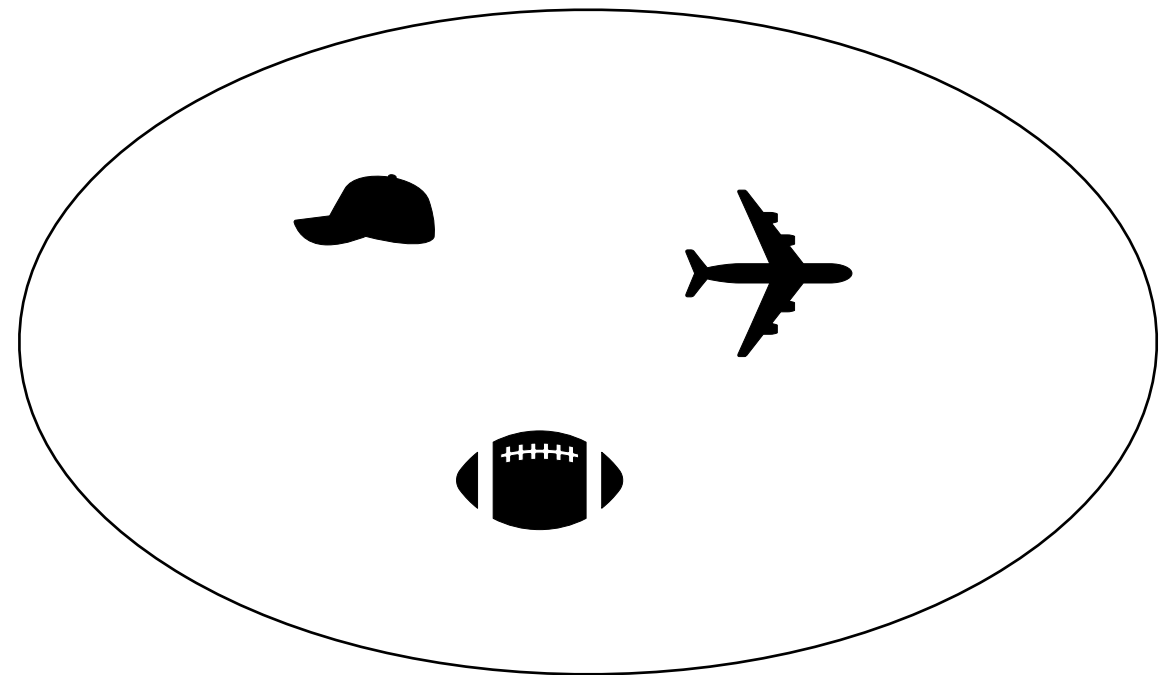
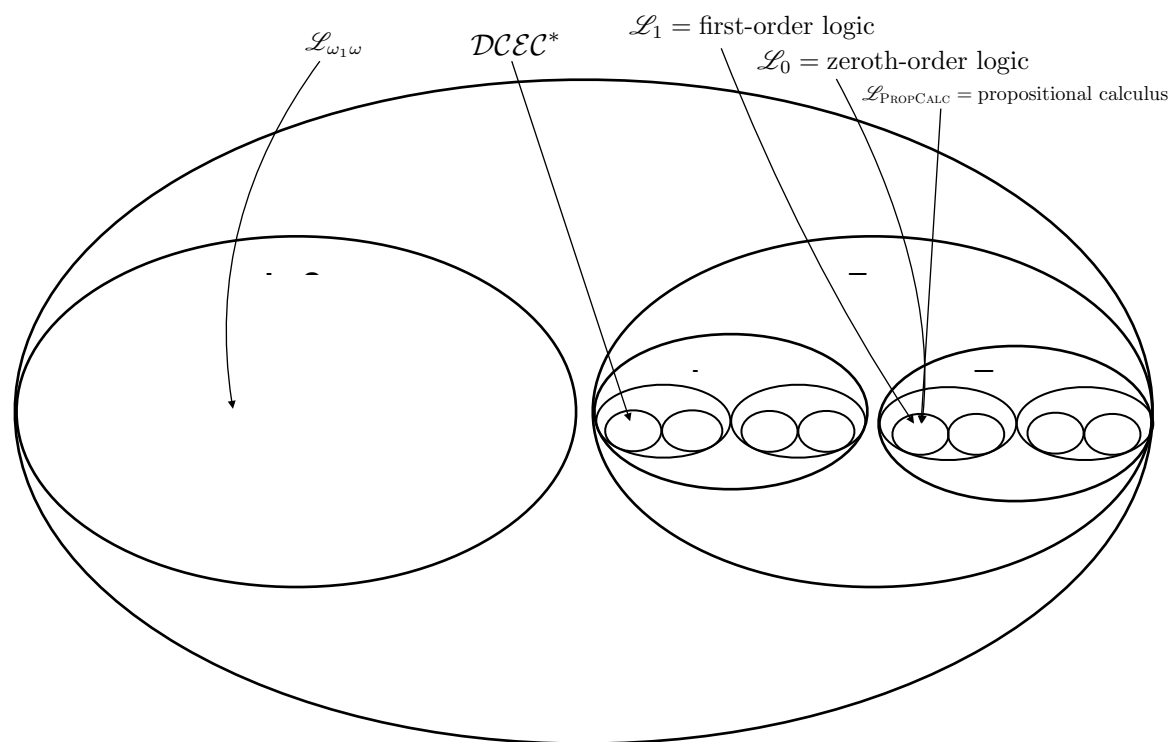
# The Physical Universe



Non-Physical

# The Universe of Logics

# The Physical Universe



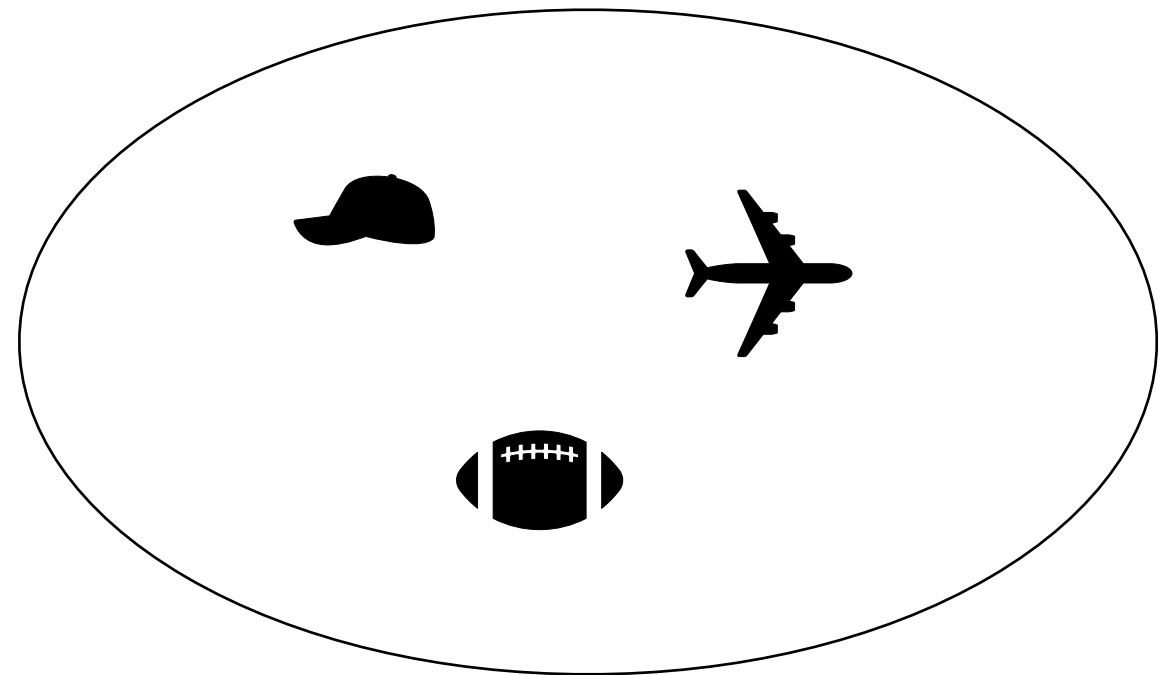
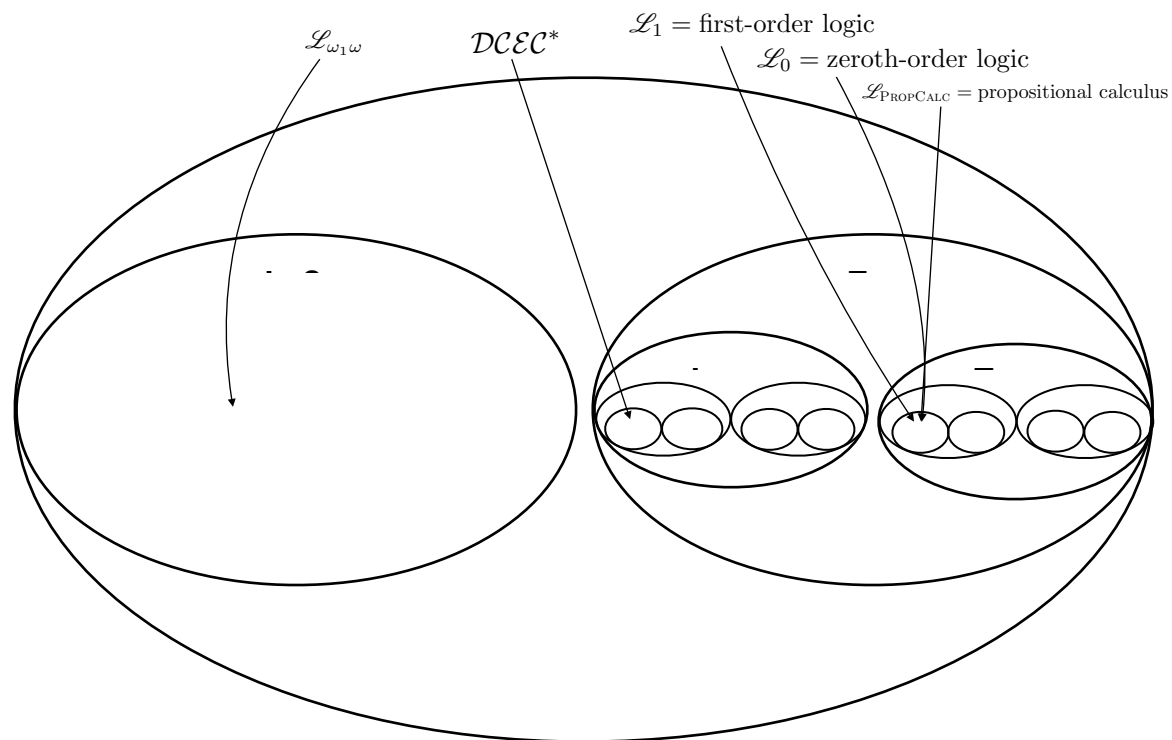
$$\frac{\phi, \phi \rightarrow \psi}{\psi}$$

*modus ponens*

Non-Physical

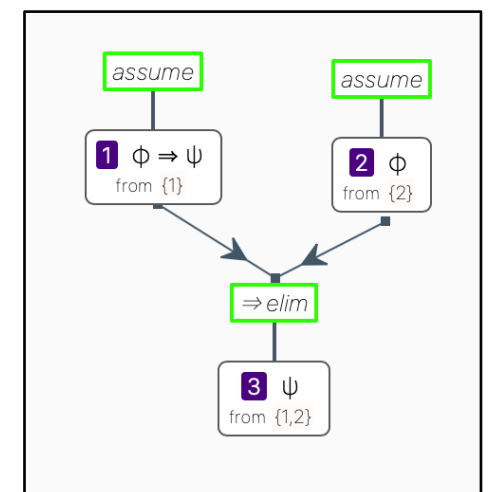
# The Universe of Logics

# The Physical Universe



$$\frac{\phi, \phi \rightarrow \psi}{\psi}$$

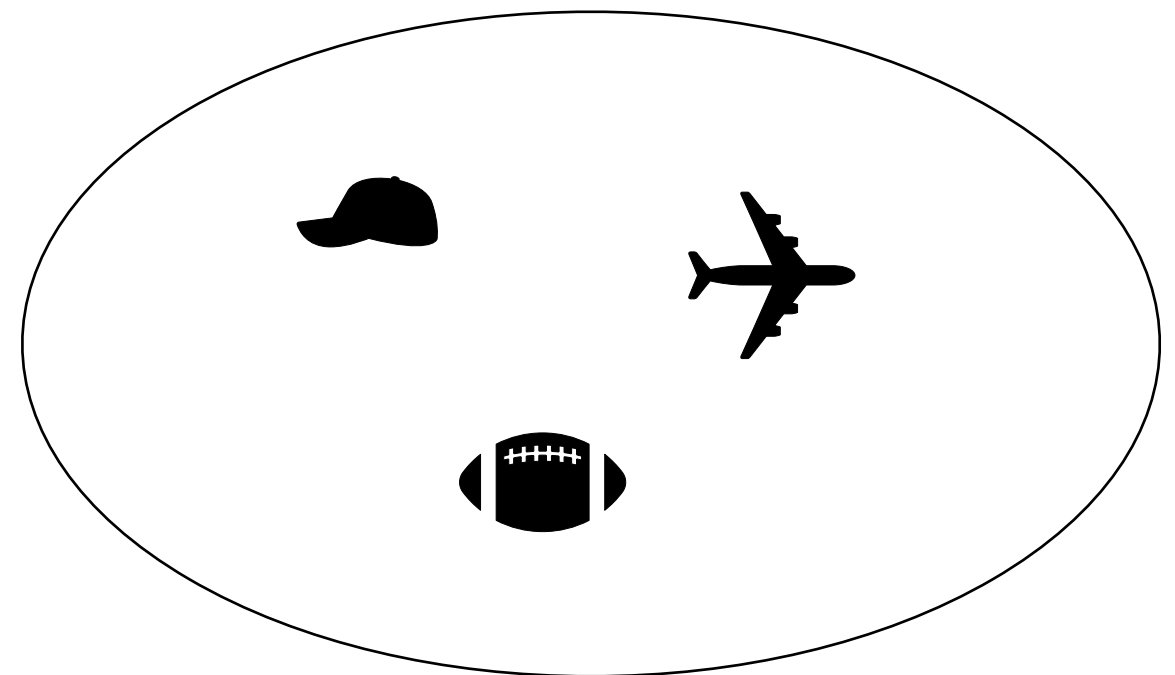
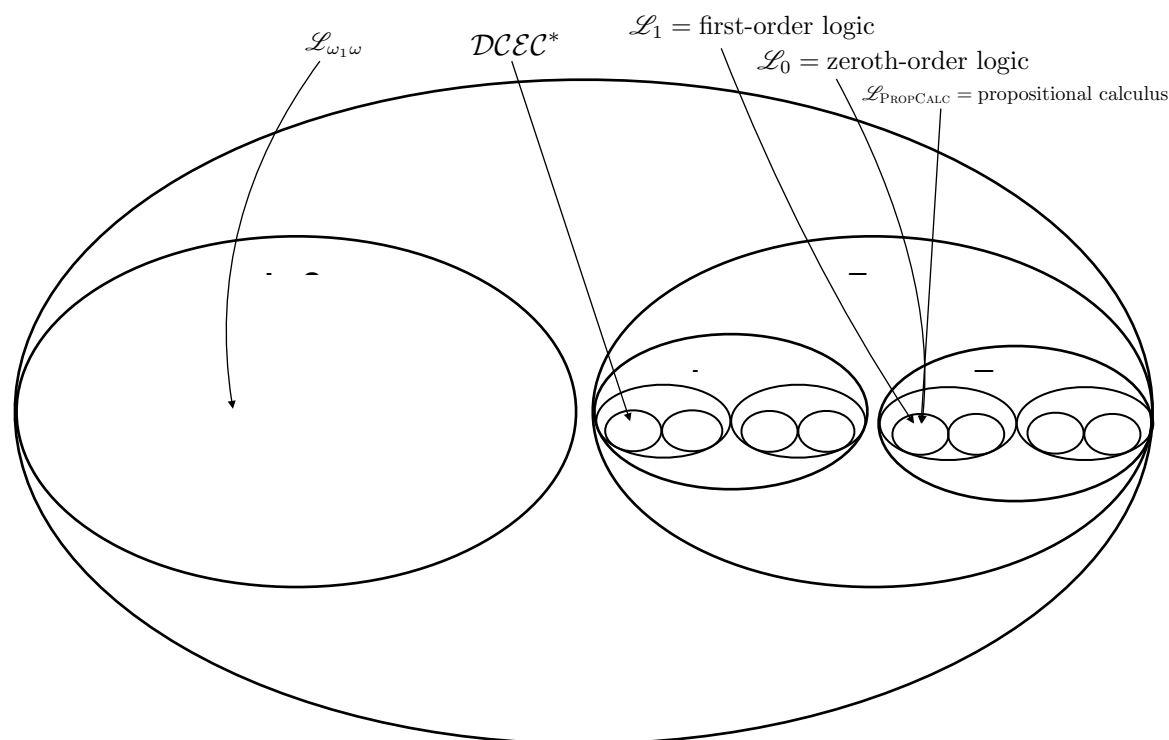
*modus ponens*



Non-Physical

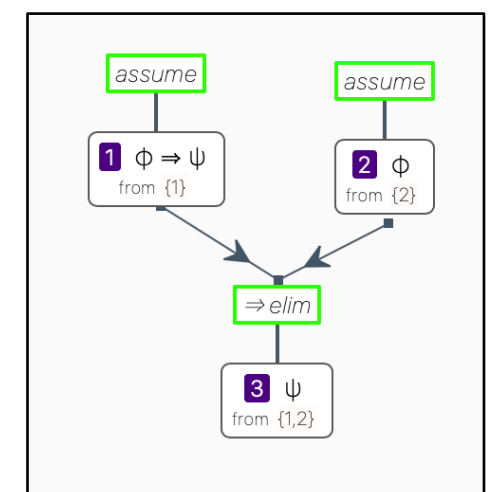
# The Universe of Logics

# The Physical Universe



$$\frac{\phi, \phi \rightarrow \psi}{\psi}$$

*modus ponens*

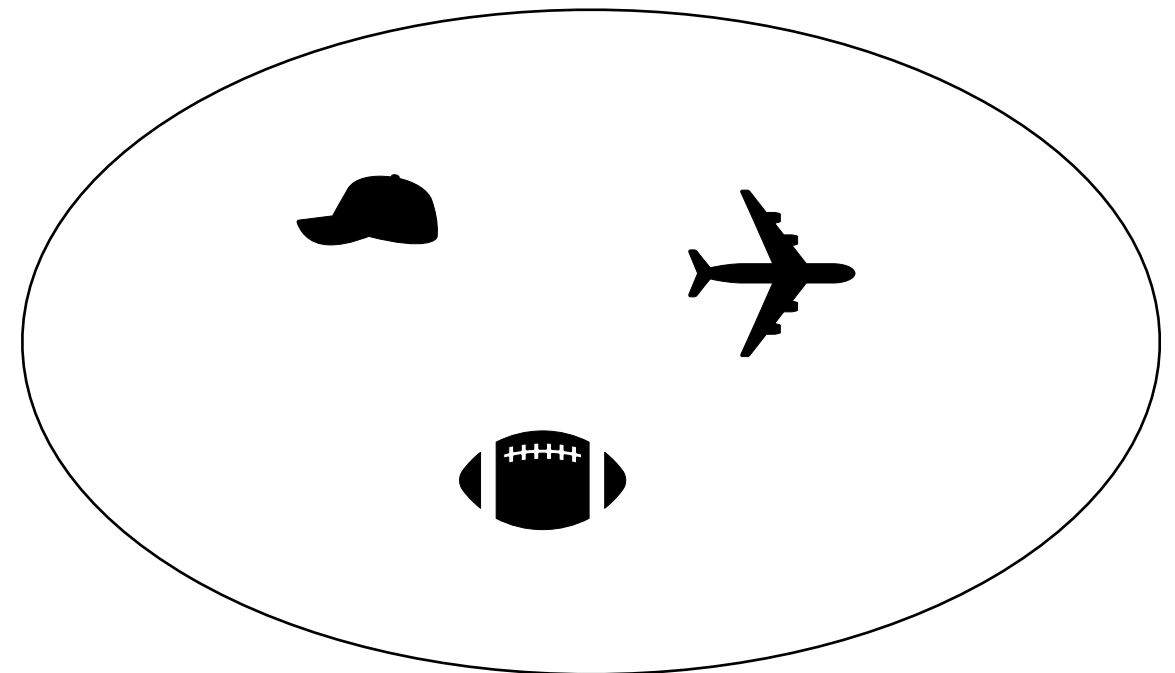
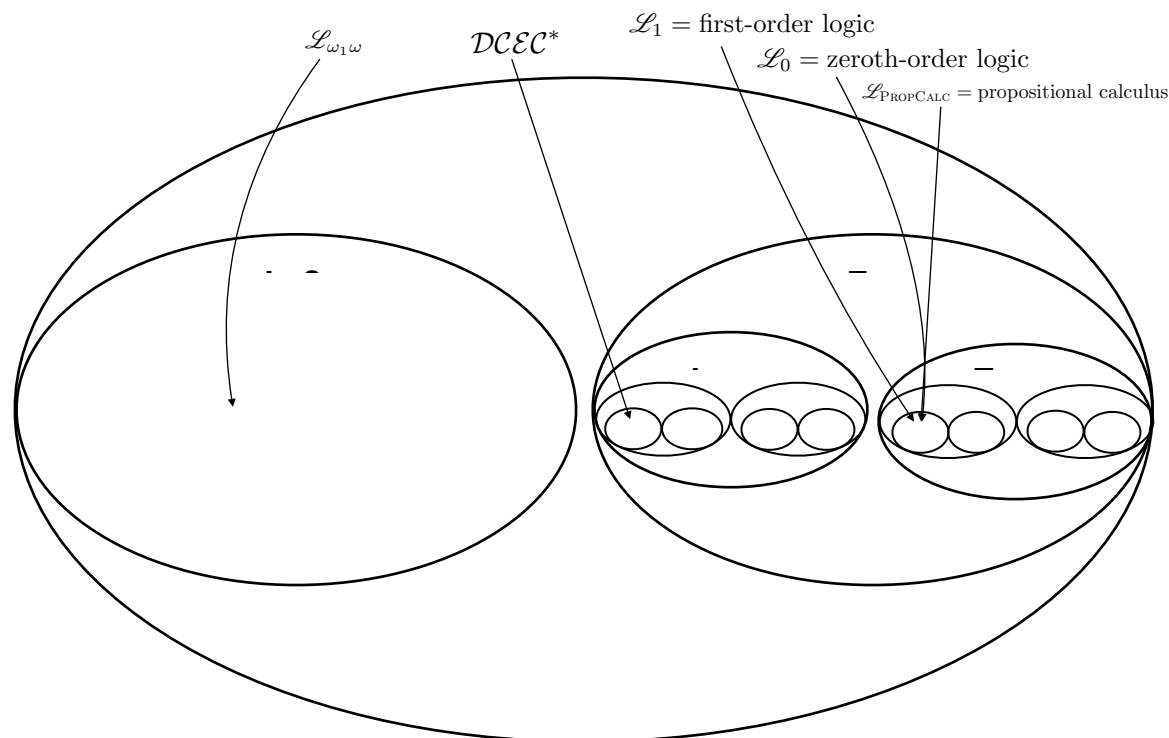




Non-Physical

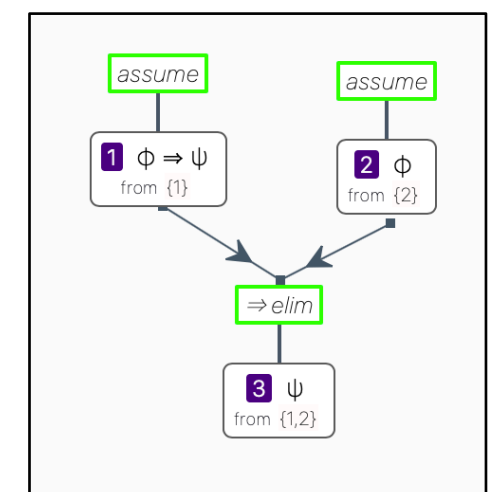
# The Universe of Logics

# The Physical Universe



$$\frac{\phi, \phi \rightarrow \psi}{\psi}$$

*modus ponens*

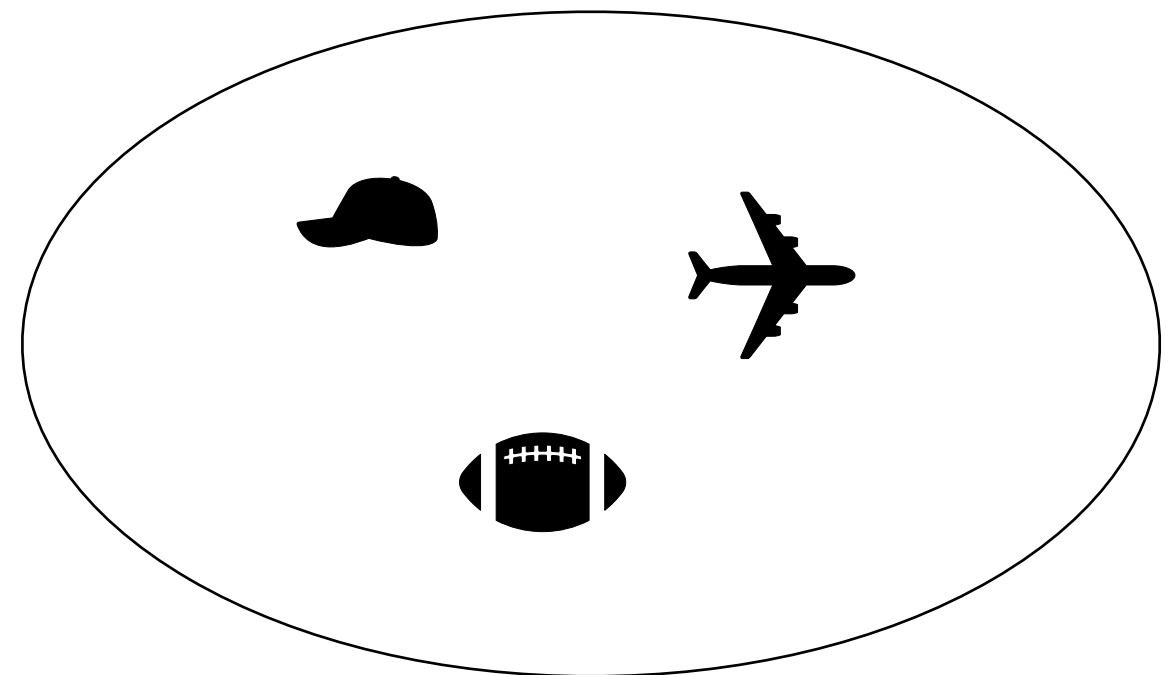
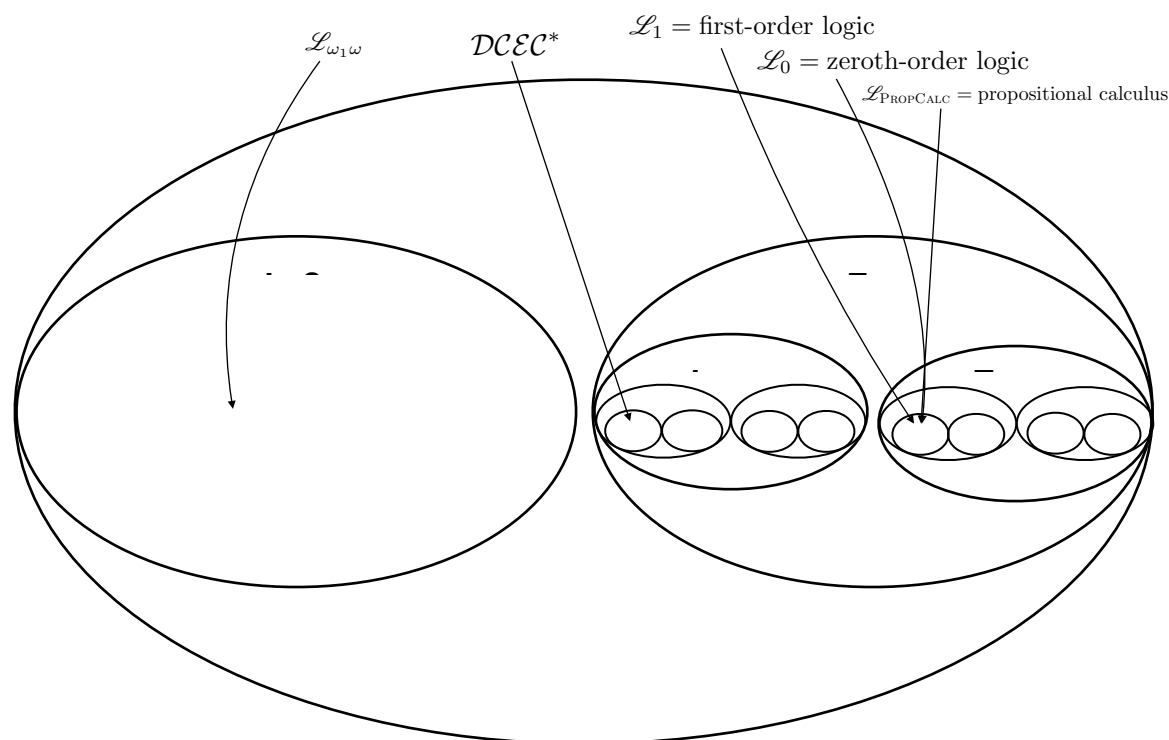


Non-Physical



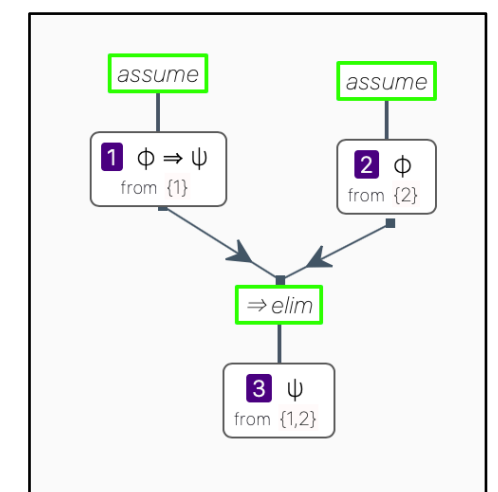
# The Universe of Logics

# The Physical Universe



$$\frac{\phi, \phi \rightarrow \psi}{\psi}$$

*modus ponens*



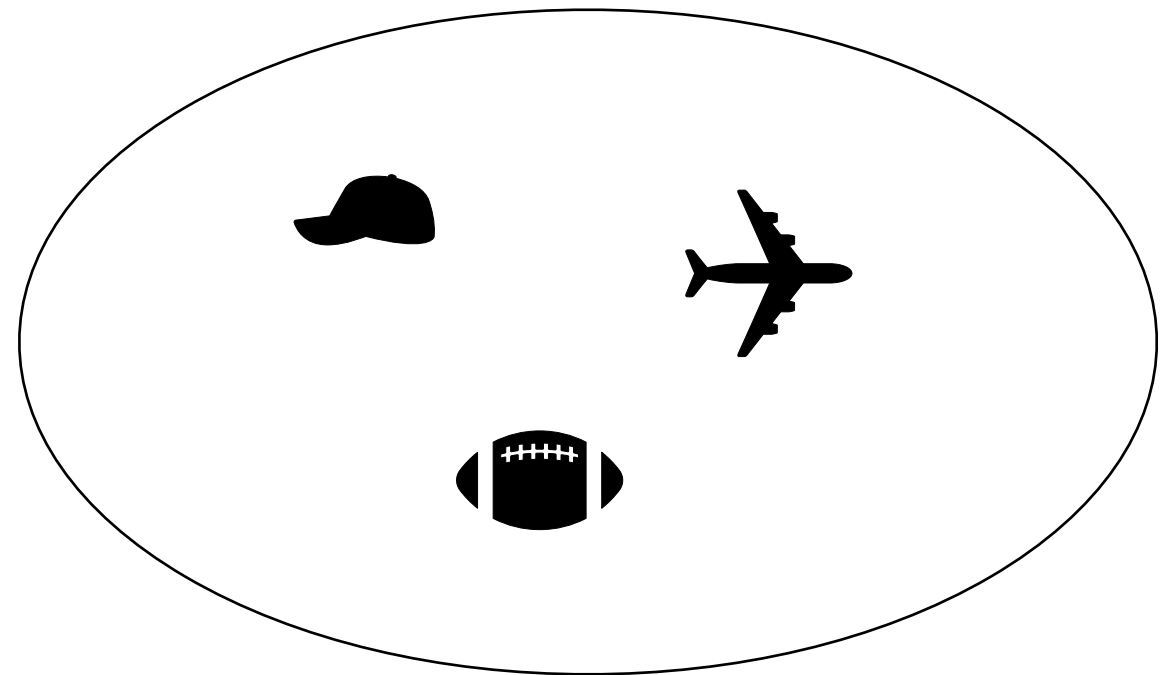
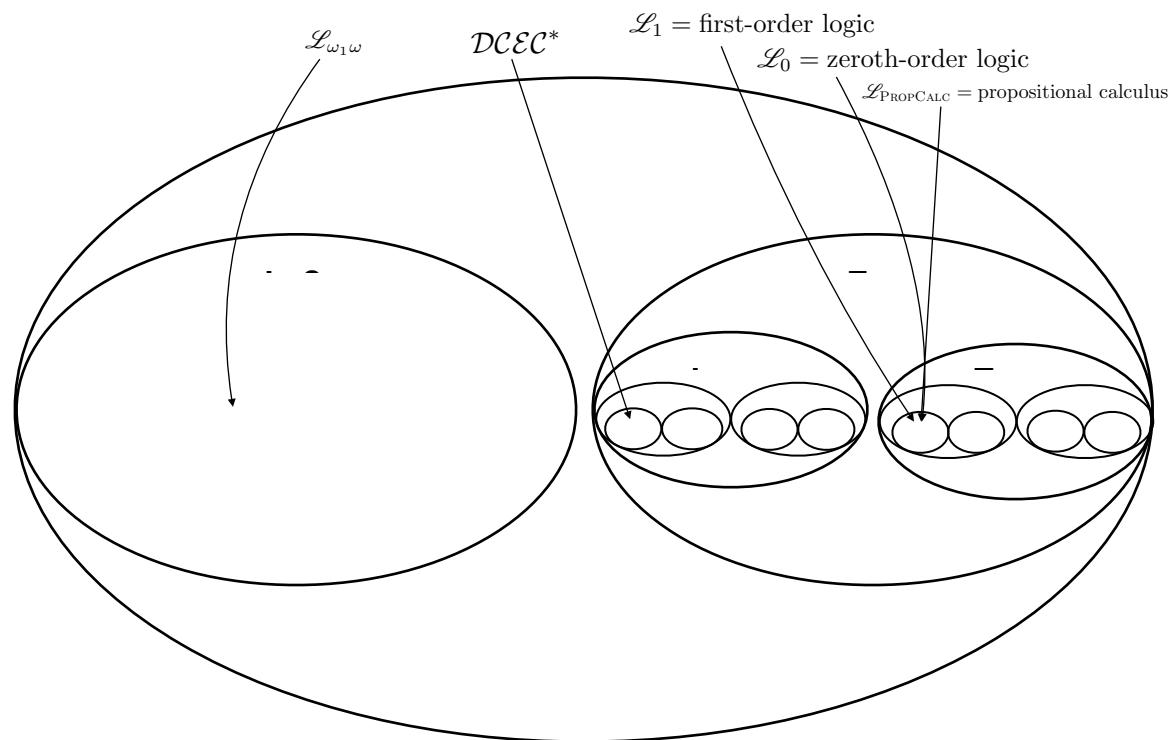


Non-Physical

# The Universe of Logics

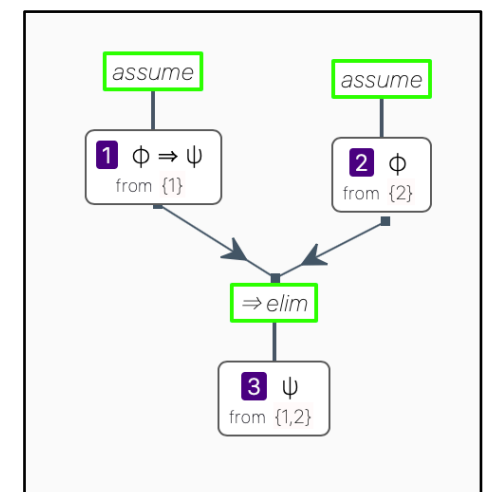


# The Physical Universe



$$\frac{\phi, \phi \rightarrow \psi}{\psi}$$

*modus ponens*

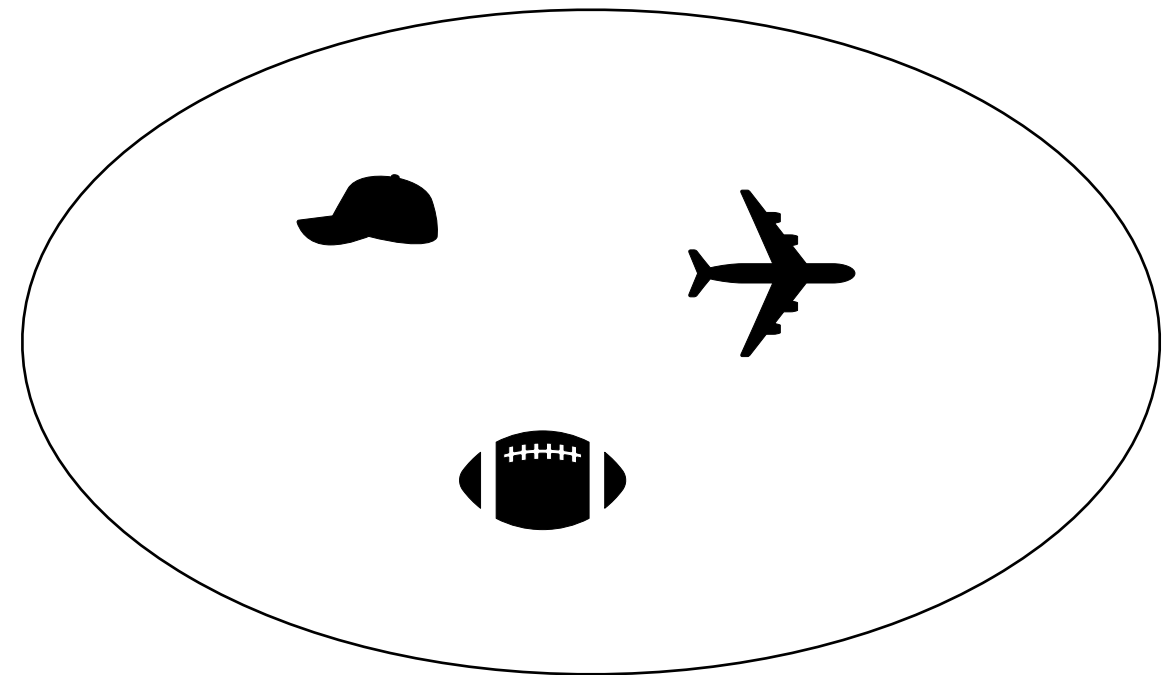
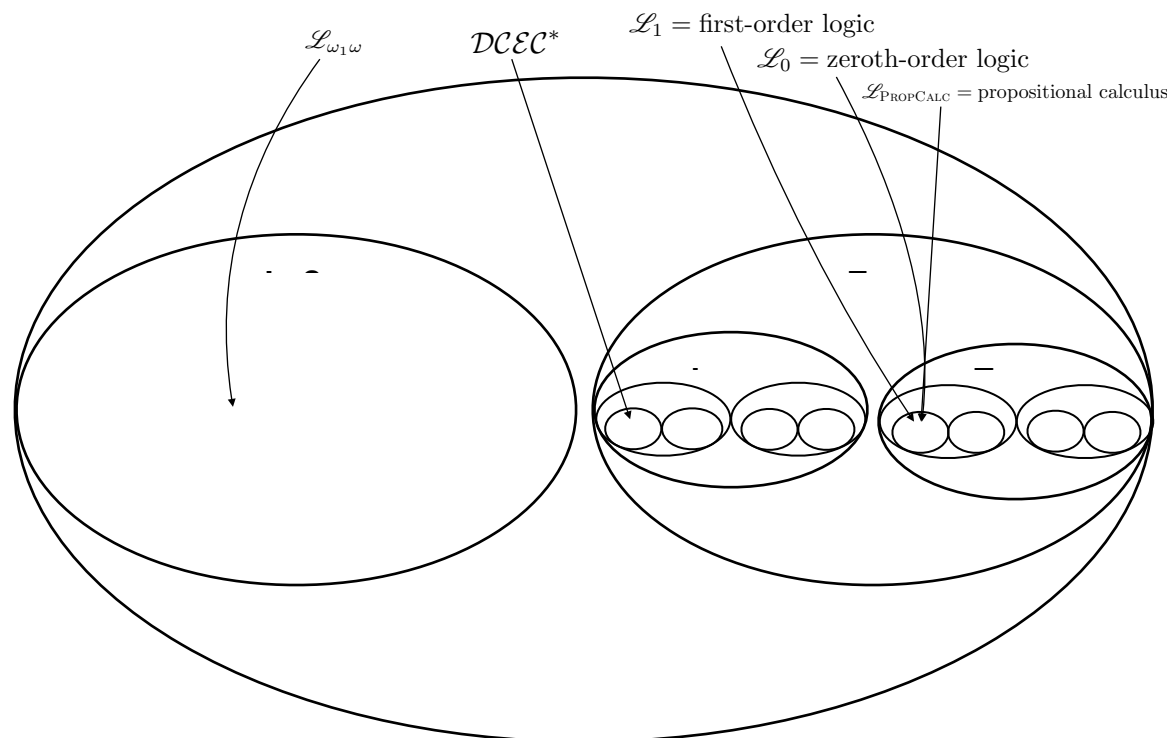


Non-Physical

# The Universe of Logics

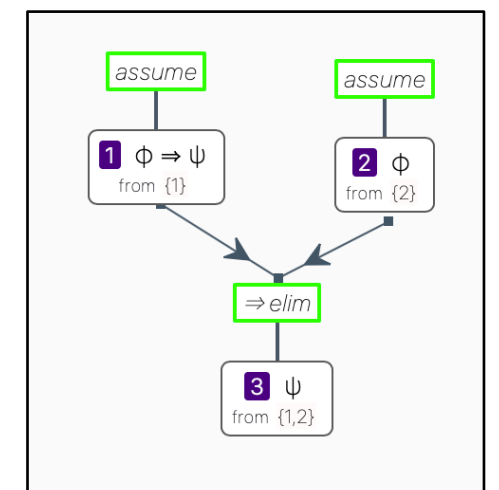


# The Physical Universe



$$\frac{\phi, \phi \rightarrow \psi}{\psi}$$

*modus ponens*



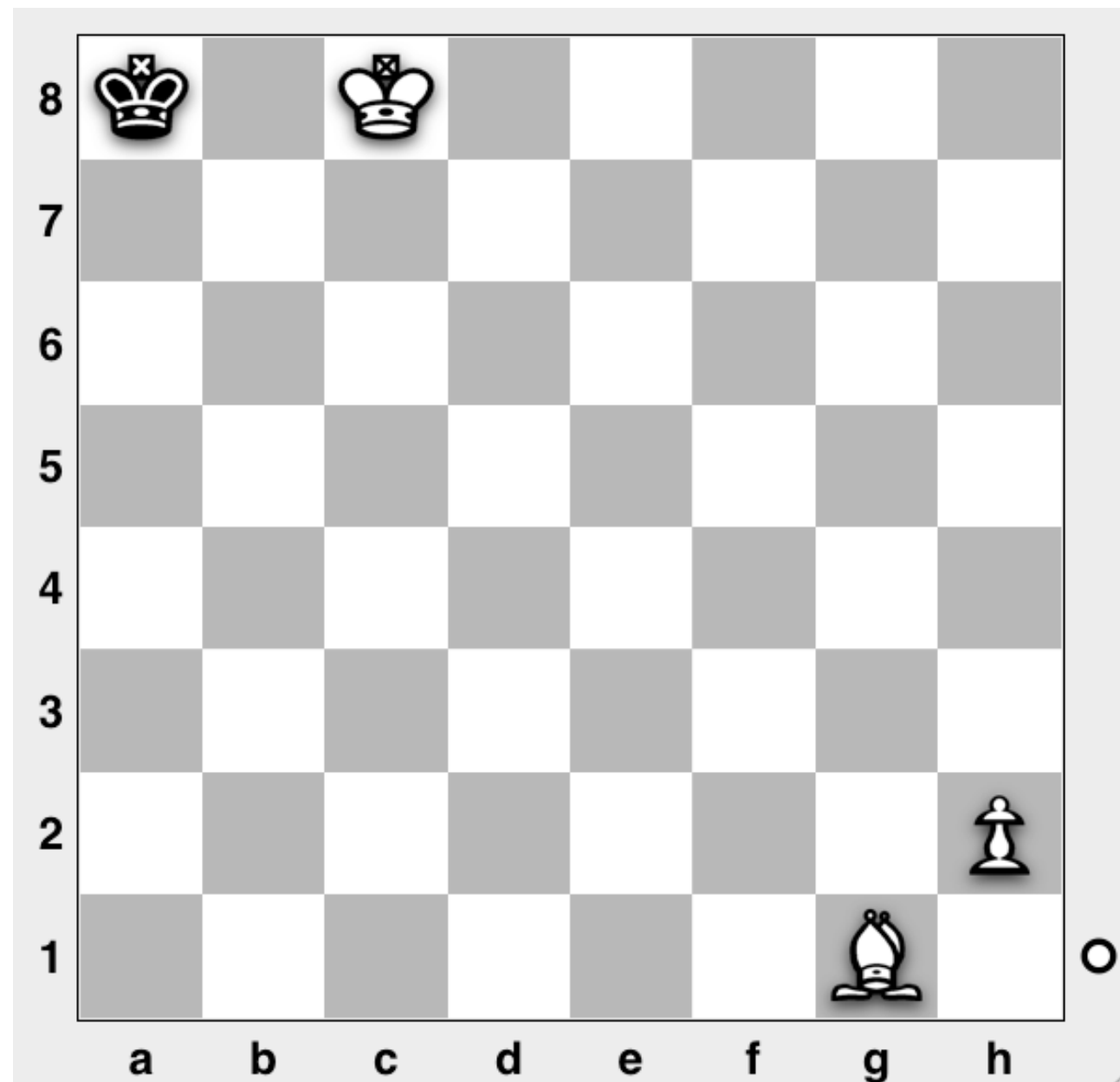
# What is Logic?

- The key to becoming rational.
- “The science of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences (e.g., engineering, computer science).
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.
- The key to divining the meaning of life (and other such big questions).
- The better way to program computers; and fundamentally the *only* way to *reliably* program computers.
- One of two fundamental approaches to studying minds, and replicating/simulating minds in machines...
- The thing many creatures of fiction have mastered — have you (as a New Yorker)?...
- ...

# What is Logic?

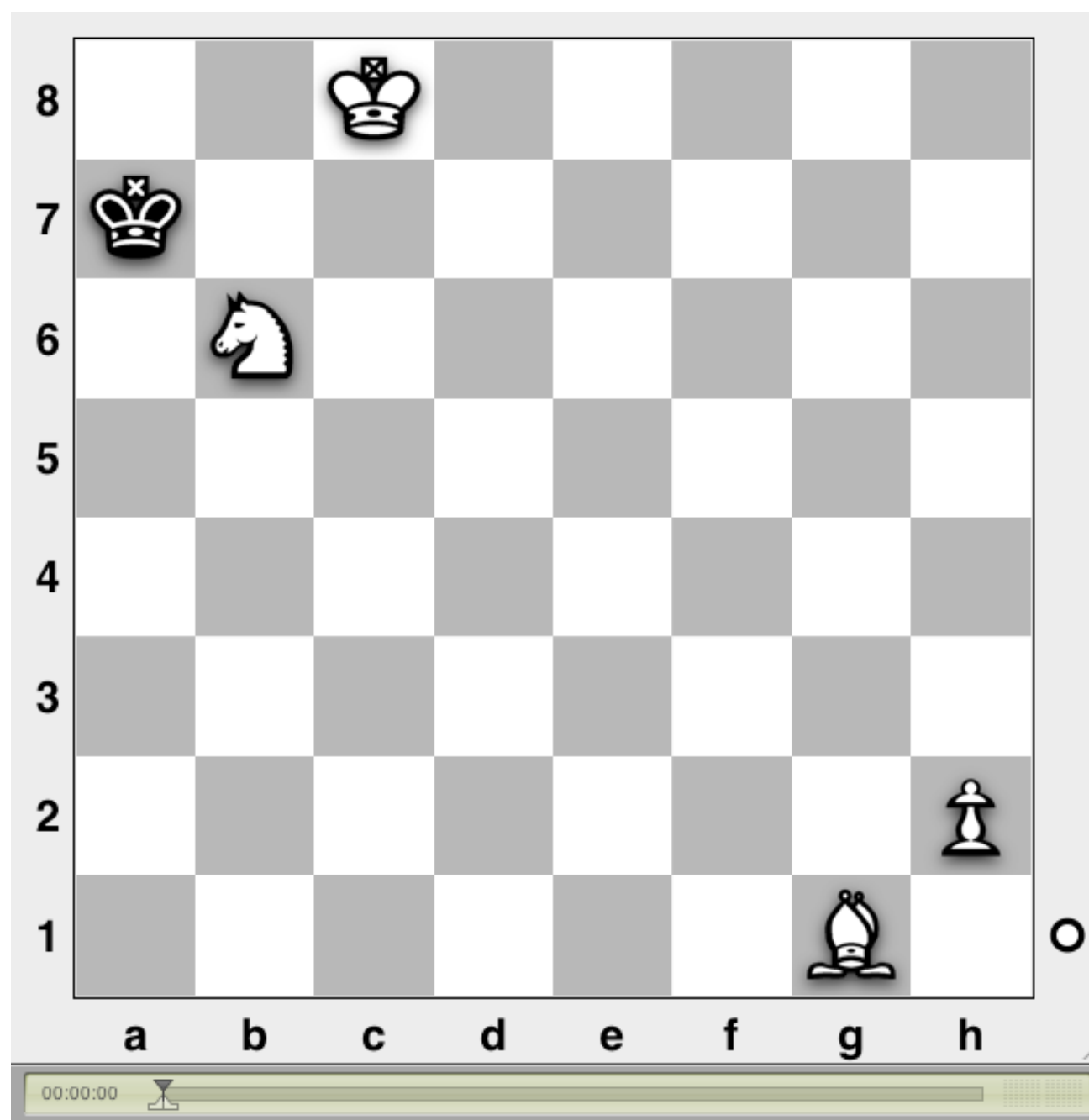
- The key to becoming rational. Or are you *already* rational? ...
- “The science of reasoning.” — so the not-unreasonable slogan goes.
- The only invincible subject there is.
- The basis for the formal sciences (from mathematics to game theory to decision theory to probability calculi to axiomatic physics ....) — and hence the basis for disciplines based on the formal sciences (e.g., engineering, computer science).
- The way of escape from shallow content and context to pure, immaterial, and immortal form and structure (which is why the exotic, imaginary, and seemingly non-sensical is so pedagogically useful).
- The most challenging subject there is.
- One of the chief differentiators between dogs and monkeys versus you (let alone bears and you); and mindless machines (like Deep Blue & Watson) versus you.
- A key to riches.
- The key to divining the meaning of life (and other such big questions).
- The better way to program computers; and fundamentally the *only* way to *reliably* program computers.
- One of two fundamental approaches to studying minds, and replicating/simulating minds in machines...
- The thing many creatures of fiction have mastered — have you (as a New Yorker)?...
- ...

It's White's turn. What move did Black just make?



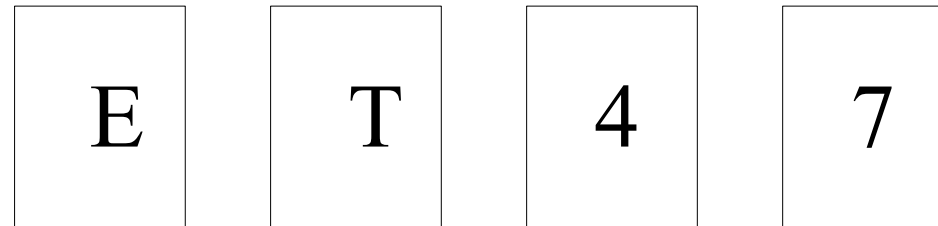
**Aha! (Beyond Deep Blue?)**

# Aha! (Beyond Deep Blue?)



**NOTE:** Every card in this game has a capital Roman letter on one side, and a number from 1 to 9, inclusive.

# Simple Selection Task



Suppose I claim that the following rule is true.

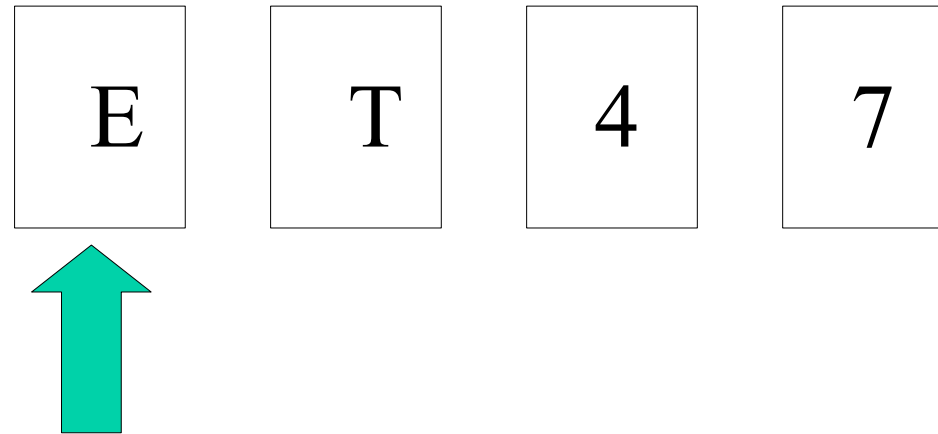
If a card has a vowel on one side, it has an even number on the other side.

Which card or cards, if any, should you turn over in order to try to efficiently decide whether the rule is true or false?



**NOTE:** Every card in this game has a capital Roman letter on one side, and a number from 1 to 9, inclusive.

# Simple Selection Task



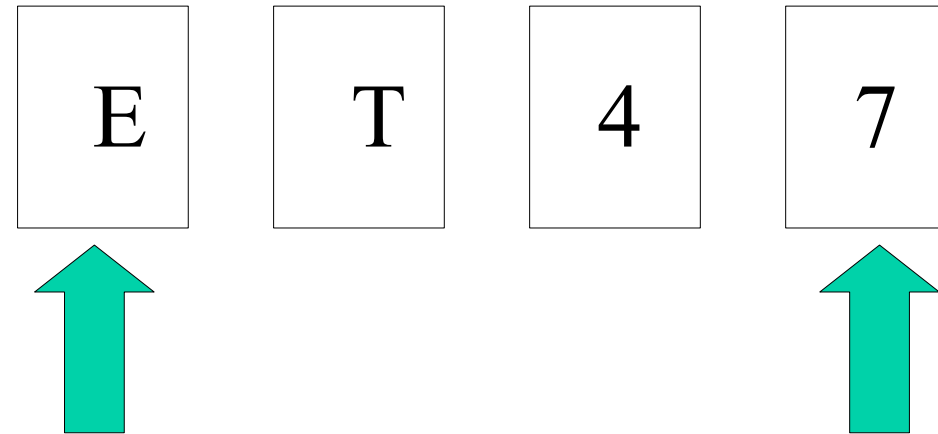
Suppose I claim that the following rule is true.

If a card has a vowel on one side, it has an even number on the other side.

Which card or cards, if any, should you turn over in order to try to efficiently decide whether the rule is true or false?

**NOTE:** Every card in this game has a capital Roman letter on one side, and a number from 1 to 9, inclusive.

# Simple Selection Task

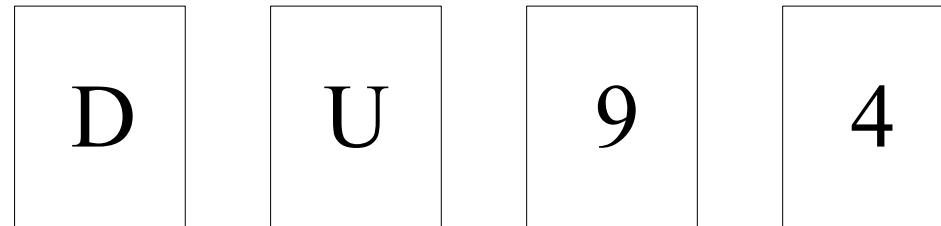


Suppose I claim that the following rule is true.

If a card has a vowel on one side, it has an even number on the other side.

Which card or cards, if any, should you turn over in order to try to efficiently decide whether the rule is true or false?

# Another Simple Selection Task

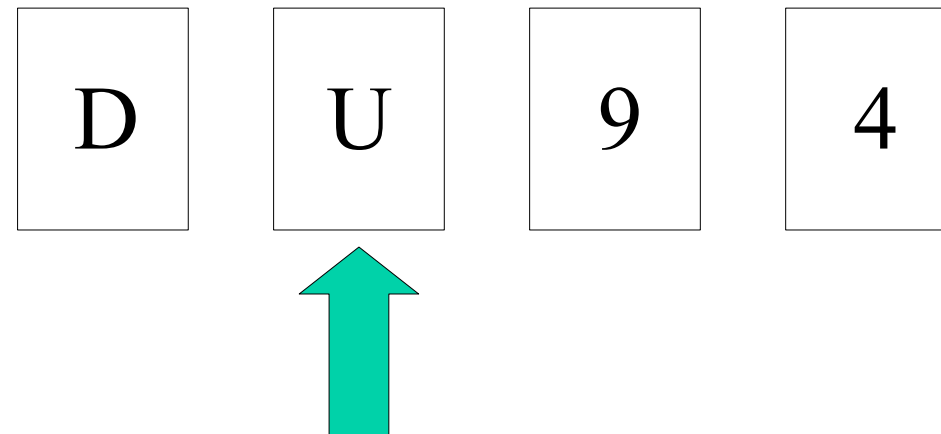


Suppose I claim that the following rule is true.

If a card has a vowel on one side, it has an even number on the other side.

Which card or cards, if any, should you turn over in order to try to efficiently decide whether the rule is true or false?

# Another Simple Selection Task

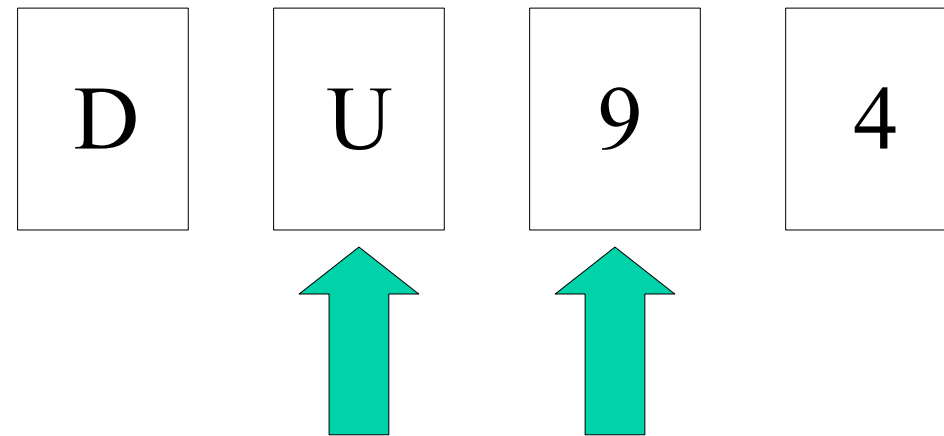


Suppose I claim that the following rule is true.

If a card has a vowel on one side, it has an even number on the other side.

Which card or cards, if any, should you turn over in order to try to efficiently decide whether the rule is true or false?

# Another Simple Selection Task

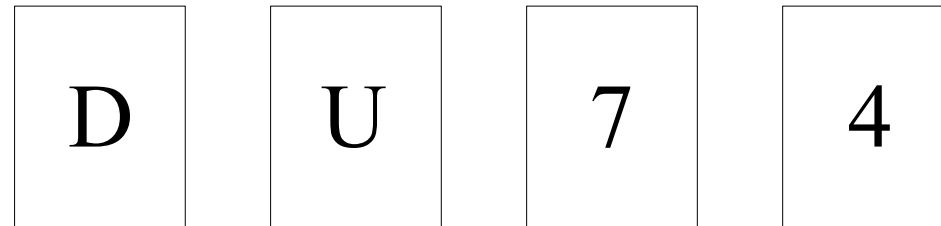


Suppose I claim that the following rule is true.

If a card has a vowel on one side, it has an even number on the other side.

Which card or cards, if any, should you turn over in order to try to efficiently decide whether the rule is true or false?

# Another Simple Selection Task

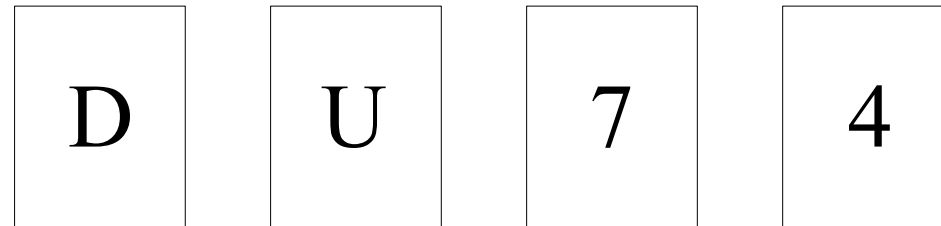


Suppose I claim that the following rule is true.

If a card has a letter on one side, it has a prime number on the other side.

Which card or cards, if any, should you turn over in order to try to efficiently decide whether the rule is true or false?

# Another Simple Selection Task



Suppose I claim that the following rule is true.

If a card has a letter on one side, it has a prime number on the other side.

Which card or cards, if any, should you turn over in order to try to efficiently decide whether the rule is true or false?



# “NYS I”

Given the statements

$$\neg a \vee \neg b$$

$$b$$

$$c \rightarrow a$$

which one of the following statements can you prove?

$$c$$

$$\neg b$$

$$\neg c$$

$$h$$

$$a$$

none of the above

# “NYS I”

Given the statements

$$\neg a \vee \neg b$$

$b$

$$c \rightarrow a$$

which one of the following statements can you prove?

$c$

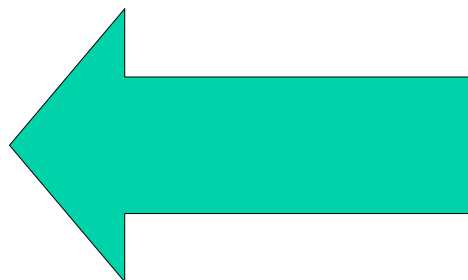
$\neg b$

$\neg c$

$h$

$a$

none of the above



### Without Consciousness, AIs Will Be Sociopaths

ChatGPT can carry on a conversation, but the most important goal for artificial intelligence is making it understand what it means to have a mind

ESSAY

By Michael S.A. Graziano [+ Read](#)

January 13, 2023 09:24 a.m. EST



ChatGPT, the latest technological sensation, is an artificial intelligence chatbot with an amazing ability to carry on a conversation. It relies on a massive network of artificial neurons that loosely mimics the human brain, and it has been trained by analyzing the information resources of the internet. ChatGPT has processed more text than any human is likely to have read in a lifetime, allowing it to respond to questions fluently and even to imitate specific individuals, answering queries the way it thinks they would. My teenage son recently used ChatGPT to argue about politics with an imitation Karl Marx.

As a neuroscientist specializing in the brain mechanisms of con-

sciousness, I find talking to chatbots an unsettling experience. Are they conscious? Probably not. But given the rate of technological improvement, will they be in the next couple of years? And how would we even know?

Figuring out whether a machine has or understands humanlike consciousness is more than just a science-fiction hypothetical. Artificial intelligence is growing so powerful, so quickly, that it could soon pose a danger to human beings. We're building machines that are smarter than us and giving them control over our world. How can we

build AI so that it's aligned with human needs, not in conflict with us?

As counterintuitive as it may sound, creating a benign AI may require making it more conscious, not less. One of the most common misunderstandings about AI is the notion that if it's intelligent then it must be conscious, and if it is conscious then it will be autonomous, capable of taking over the world. But as we learn more about consciousness, those ideas do not appear to be correct. An autonomous system that makes complex decisions doesn't require consciousness.

What's most important about consciousness is that, for human beings, it's not just about the self. We see it in ourselves, but we also perceive it or project it into the world around us. Consciousness is part of the tool kit that evolution

gave us to make us an empathetic, prosocial species. Without it, we would necessarily be sociopaths, because we'd lack the tools for prosocial behavior. And without a concept of what consciousness is or an understanding that other beings have it, machines are sociopaths.

The only diagnostic tool for machine consciousness that we have right now is the Turing test, a thought experiment named for the British computer scientist Alan Turing. In its most common version, the test says that if a person holds a conversation with a machine and mistakes its responses for those of a real human being, then the machine must be considered effectively conscious.

The Turing test is an admission that the consciousness of another being is something we can only judge from the outside, based on the way he, she or it communicates. But the

“NYS 2”

Which one of the following statements is provable from the following statement: “If you are not part of the solution, then you are part of the problem.”

If you are part of the solution, then you are not part of the problem.

If you are not part of the problem, then you are part of the solution.

If you are part of the problem, then you are not part of the solution.

If you are not part of the problem, then you are not part of the solution.

Graziano’s Mistake!



“NYS 2”

Which one of the following statements is provable from the following statement: “If you are not part of the solution, then you are part of the problem.”

If you are part of the solution, then you are not part of the problem.

➡ If you are not part of the problem, then you are part of the solution.

If you are part of the problem, then you are not part of the solution.

If you are not part of the problem, then you are not part of the solution.

Graziano’s Mistake!



“NYS 2”

Which one of the following statements is provable from the following statement: “If you are not part of the solution, then you are part of the problem.”

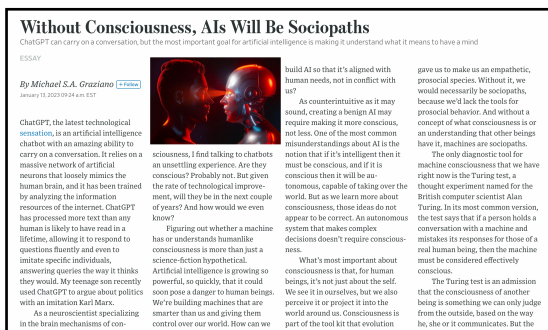
If you are part of the solution, then you are not part of the problem.

➡ If you are not part of the problem, then you are part of the solution.

✗ If you are part of the problem, then you are not part of the solution.

If you are not part of the problem, then you are not part of the solution.

Graziano’s Mistake!



# “NYS 2”

Which one of the following statements is provable from the following statement: “If you are not part of the solution, then you are part of the problem.”

If you are part of the solution, then you are not part of the problem.

➡ If you are not part of the problem, then you are part of the solution.

✗ If you are part of the problem, then you are not part of the solution.

If you are not part of the problem, then you are not part of the solution.

Graziano’s Mistake!

# “NYS 3”

Given the statements

$$\neg\neg c$$

$$c \rightarrow a$$

$$\neg a \vee b$$

$$b \rightarrow d$$

$$\neg(d \vee e)$$

which of the following statements are provable?

$$\neg c$$

$$e$$

$$h$$

$$\neg a$$

all of the above

# “NYS 3”

Given the statements

$$\neg\neg c$$

$$c \rightarrow a$$

$$\neg a \vee b$$

$$b \rightarrow d$$

$$\neg(d \vee e)$$

which of the following statements are provable?

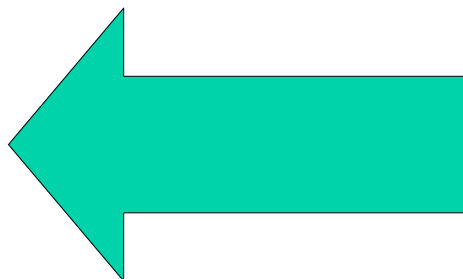
$$\neg c$$

e

h

$$\neg a$$

all of the above





# The Original King-Ace

Suppose that the following premise is true:

If there is a king in the hand, then there is an ace in the hand, or else if there isn't a king in the hand, then there is an ace.

What can you infer from this premise?

# The Original King-Ace

Suppose that the following premise is true:

If there is a king in the hand, then there is an ace in the hand, or else if there isn't a king in the hand, then there is an ace.

What can you infer from this premise?

There is an ace in the hand.

# The Original King-Ace

Suppose that the following premise is true:

If there is a king in the hand, then there is an ace in the hand, or else if there isn't a king in the hand, then there is an ace.

What can you infer from this premise?

~~There is an ace in the hand.~~

# The Original King-Ace

Suppose that the following premise is true:

If there is a king in the hand, then there is an ace in the hand, or else if there isn't a king in the hand, then there is an ace.

What can you infer from this premise?

NO! ~~There is an ace in the hand.~~

# The Original King-Ace

Suppose that the following premise is true:

If there is a king in the hand, then there is an ace in the hand, or else if there isn't a king in the hand, then there is an ace.

What can you infer from this premise?

NO! ~~There is an ace in the hand.~~ NO!

# The Original King-Ace

Suppose that the following premise is true:

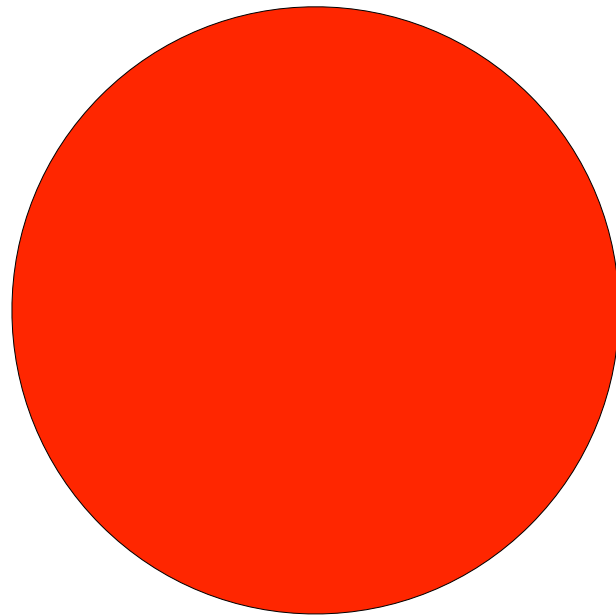
If there is a king in the hand, then there is an ace in the hand, or else if there isn't a king in the hand, then there is an ace.

What can you infer from this premise?

NO! ~~There is an ace in the hand.~~ NO!

In fact, what you *can* infer is that there *isn't* an ace in the hand!

**End of Part I**



# King-Ace 2

Suppose that the following premise is true:

*If there is a king in the hand, then there is an ace in the hand; or if there isn't a king in the hand, then there is an ace; but not both of these if-then statements are true.*

What can you infer from this premise?



# King-Ace 2

Suppose that the following premise is true:

*If there is a king in the hand, then there is an ace in the hand; or if there isn't a king in the hand, then there is an ace; but not both of these if-then statements are true.*

What can you infer from this premise?

There is an ace in the hand.

# King-Ace 2

Suppose that the following premise is true:

*If there is a king in the hand, then there is an ace in the hand; or if there isn't a king in the hand, then there is an ace; but not both of these if-then statements are true.*

What can you infer from this premise?

~~There is an ace in the hand.~~

# King-Ace 2

Suppose that the following premise is true:

*If there is a king in the hand, then there is an ace in the hand; or if there isn't a king in the hand, then there is an ace; but not both of these if-then statements are true.*

What can you infer from this premise?

NO! ~~There is an ace in the hand.~~

# King-Ace 2

Suppose that the following premise is true:

*If there is a king in the hand, then there is an ace in the hand; or if there isn't a king in the hand, then there is an ace; but not both of these if-then statements are true.*

What can you infer from this premise?

NO! ~~There is an ace in the hand.~~ NO!

# King-Ace 2

Suppose that the following premise is true:

*If there is a king in the hand, then there is an ace in the hand; or if there isn't a king in the hand, then there is an ace; but not both of these if-then statements are true.*

What can you infer from this premise?

NO! ~~There is an ace in the hand.~~ NO!

In fact, what you *can* infer is that there *isn't* an ace in the hand!

Informal Proofs/Arguments vs. Formal Proofs/Arguments

# Informal Proofs/Arguments vs. Formal Proofs/Arguments

has ambiguous natural language  
(e.g. English or Chinese)

# Informal Proofs/Arguments vs. Formal Proofs/Arguments

has ambiguous natural language  
(e.g. English or Chinese)

no ambiguous natural language  
(e.g. English or Chinese)



# Informal Proofs/Arguments vs. Formal Proofs/Arguments

has ambiguous natural language  
(e.g. English or Chinese)

cannot be executed & checked  
by a computing machine

no ambiguous natural language  
(e.g. English or Chinese)

# Informal Proofs/Arguments vs. Formal Proofs/Arguments

has ambiguous natural language  
(e.g. English or Chinese)

cannot be executed & checked  
by a computing machine

no ambiguous natural language  
(e.g. English or Chinese)

*can* be executed & checked by  
a computing machine

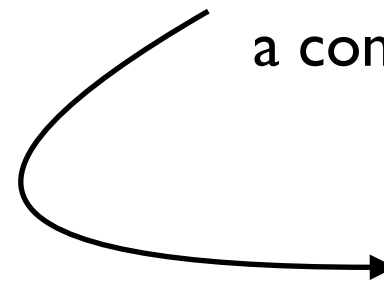
# Informal Proofs/Arguments vs. Formal Proofs/Arguments

has ambiguous natural language  
(e.g. English or Chinese)

cannot be executed & checked  
by a computing machine

no ambiguous natural language  
(e.g. English or Chinese)

*can* be executed & checked by  
a computing machine



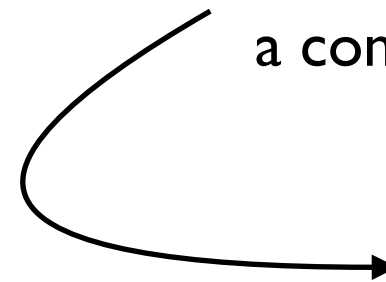
# Informal Proofs/Arguments vs. Formal Proofs/Arguments

has ambiguous natural language  
(e.g. English or Chinese)

cannot be executed & checked  
by a computing machine

no ambiguous natural language  
(e.g. English or Chinese)

*can* be executed & checked by  
a computing machine



HyperSlate®

# Informal Proofs/Arguments vs. Formal Proofs/Arguments

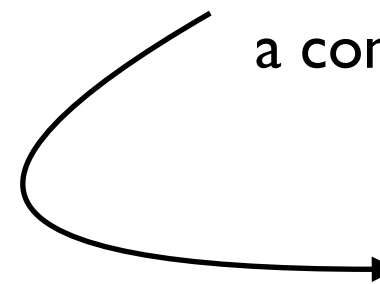
has ambiguous natural language  
(e.g. English or Chinese)

cannot be executed & checked  
by a computing machine

supposed to have learned  
how to produce, to an  
appreciable degree, in High  
School — but likely didn't

no ambiguous natural language  
(e.g. English or Chinese)

*can* be executed & checked by  
a computing machine



HyperSlate®

# Informal Proofs/Arguments vs. Formal Proofs/Arguments

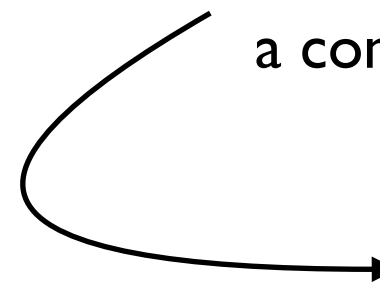
has ambiguous natural language  
(e.g. English or Chinese)

cannot be executed & checked  
by a computing machine

supposed to have learned  
how to produce, to an  
appreciable degree, in High  
School — but likely didn't

no ambiguous natural language  
(e.g. English or Chinese)

*can* be executed & checked by  
a computing machine



HyperSlate®

have not learned how to  
produce in a relevant  
system (though may have  
had some Prolog)

FOR NOW

## Informal Proofs/Arguments vs. Formal Proofs/Arguments

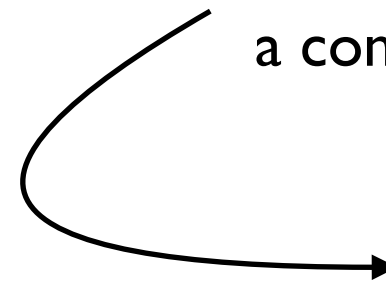
has ambiguous natural language  
(e.g. English or Chinese)

cannot be executed & checked  
by a computing machine

supposed to have learned  
how to produce, to an  
appreciable degree, in High  
School — but likely didn't

no ambiguous natural language  
(e.g. English or Chinese)

*can* be executed & checked by  
a computing machine



HyperSlate®

have not learned how to  
produce in a relevant  
system (though may have  
had some Prolog)

FOR NOW

STARTING  $\geq$  1/23/25

## Informal Proofs/Arguments vs. Formal Proofs/Arguments

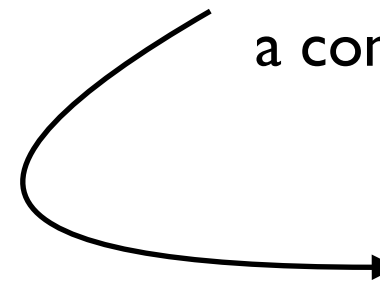
has ambiguous natural language  
(e.g. English or Chinese)

cannot be executed & checked  
by a computing machine

supposed to have learned  
how to produce, to an  
appreciable degree, in High  
School — but likely didn't

no ambiguous natural language  
(e.g. English or Chinese)

*can* be executed & checked by  
a computing machine



HyperSlate®

have not learned how to  
produce in a relevant  
system (though may have  
had some Prolog)



# King-Ace Solved

(informal proof)

**Proposition:** There is *not* an ace in the hand.

**Proof:** We know that at least one of the if-thens (i.e., at least one of the **conditionals**) is false. So we have two cases to consider, viz., that  $K \Rightarrow A$  is false, and that  $\neg K \Rightarrow A$  is false. Take first the first case; accordingly, suppose that  $K \Rightarrow A$  is false. Then it follows that  $K$  is true (since when a conditional is false, its antecedent holds but its consequent doesn't), and  $A$  is false. Now consider the second case, which consists in  $\neg K \Rightarrow A$  being false. Here, in a direct parallel, we know  $\neg K$  and, once again,  $\neg A$ . In both of our two cases, which are exhaustive, there is no ace in the hand. The proposition is established. **QED**

# Train-to-Princeton Problem

Everyone loves anyone who loves someone.

Larry loves Lucy.

Can you infer that everyone loves Lucy?

ANSWER:

PROOF:

# Train-to-Princeton Problem

Everyone loves anyone who loves someone.

Larry loves Lucy.

Can you infer that everyone loves Lucy?

ANSWER: Yup.

PROOF: ??

# The Ticking-Bomb Logic Puzzle ...



A criminal genius nearly a match for Sherlock Holmes (Do you recognize the Dr?) has built a massive hydrogen bomb, and life on Earth is hanging in the balance, hinging on whether you make the logical prediction. Dr M gives you a sporting chance to: make the right prediction, snip or not snip accordingly, and prove that you're right ...





A criminal genius nearly a  
match for Sherlock Holmes  
(Do you recognize the Dr?)





A criminal genius nearly a match for Sherlock Holmes (Do you recognize the Dr?) has built a massive hydrogen bomb, and life on Earth is hanging in the balance, hinging on whether you make the logical prediction. Dr M gives you a sporting chance to: make the right prediction, snip or not snip accordingly, and prove that you're right ...



If one of the following assertions is true then so is the other:

(1) If the red wire runs to the bomb, then the blue wire runs to the bomb; and, if the blue wire runs to the bomb, then the red wire runs to the bomb.

(2) The red wire runs to the bomb.

Given this perfectly reliable clue from Dr Moriarty, if either wire is more likely to run to the bomb, that wire *does* run to the bomb, and the bomb is ticking, with only a minute left! If both are equiprobable, neither runs to the bomb, and you are powerless. Make your prediction as to what will happen when a wire is snipped, and then make your selected snip by clicking on the wire you want to snip! Or leave well enough alone!



Red more likely.

Blue more likely.

Equiprobable.

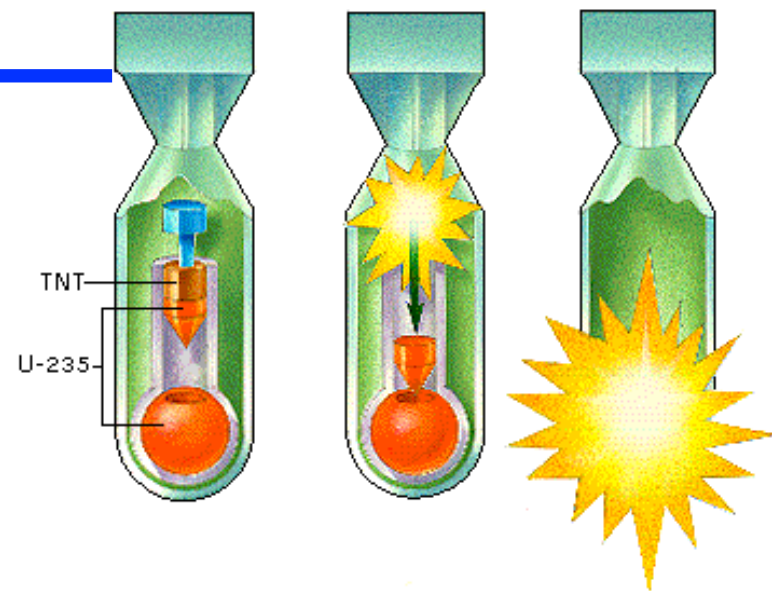


\_\_\_\_\_

\_\_\_\_\_

Snip

---



Life  
on  
Earth  
has  
ended

•

advance one more  
slide to see a proof  
that you indeed made  
an irrational  
decision...

**Proposition:** The blue wire is more likely!

**Proof:** (1) can be treated as a biconditional, obviously ( $R \iff B$ ).

There are two top-level cases to consider: (1) and (2) are both true; or both are false. In the case where they are both true, it's trivial to deduce both R and B. So far, then, R and B are equiprobable. What happens in the case where (1) and (2) are both false? We immediately have  $\sim R$  from the denial of (2). But a biconditional is true just in case both sides are true, or both sides are false; so we have two sub-cases to consider.

Consider first the case where R is true and B is false. We have an immediate contradiction in this sub-case, so both R and B can both be deduced here, and we have not yet departed from equiprobable. So what about the case where R is false and B is true? The falsity of R is not new information (we already have that from the denial of (2)), but we can still derive B. Hence the blue wire is more likely. **QED**

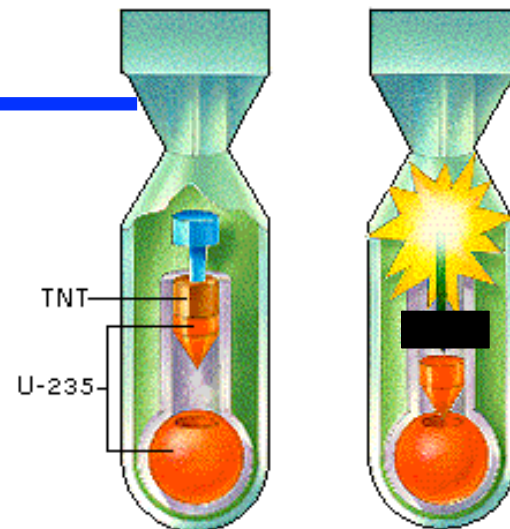
\_\_\_\_\_

\_\_\_\_\_

---

Snip

---



Life on  
Earth  
is  
saved!

*if you can now hand Dr  
M a proof that your  
decision was the rational  
one!*

Advance one more slide  
to see a proof from  
Bringsjord that yours  
had better match up to  
...

**Proposition:** The blue wire is more likely!

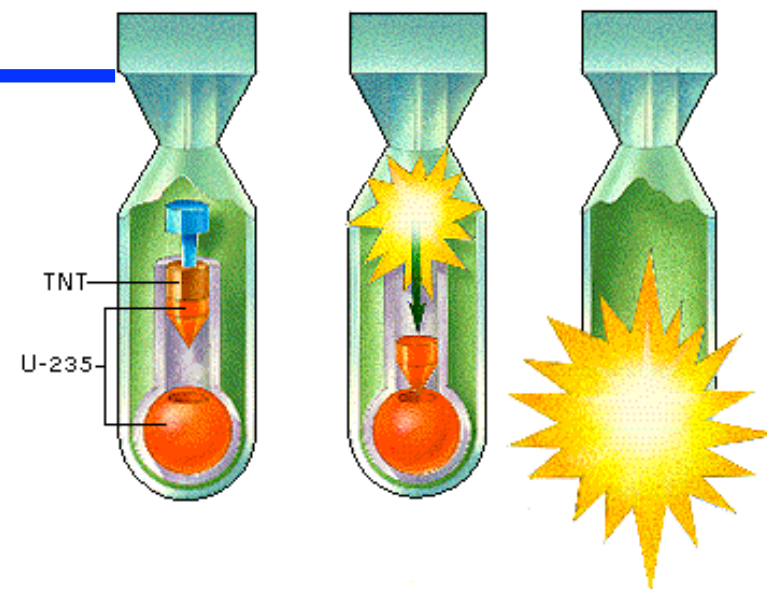
**Proof:** (1) can be treated as a biconditional, obviously ( $R \iff B$ ).

There are two top-level cases to consider: (1) and (2) are both true; or both are false. In the case where they are both true, it's trivial to deduce both R and B. So far, then, R and B are equiprobable. What happens in the case where (1) and (2) are both false? We immediately have  $\sim R$  from the denial of (2). But a biconditional is true just in case both sides are true, or both sides are false; so we have two sub-cases to consider.

Consider first the case where R is true and B is false. We have an immediate contradiction in this sub-case, so both R and B can both be deduced here, and we have not yet departed from equiprobable. So what about the case where R is false and B is true? The falsity of R is not new information (we already have that from the denial of (2)), but we can still derive B. Hence the blue wire is more likely. **QED**

\_\_\_\_\_

\_\_\_\_\_



Life  
on  
Earth  
has  
ended

•

advance one more  
slide to see a proof  
that you indeed made  
an irrational  
decision...



**Proposition:** The blue wire is more likely!

**Proof:** (1) can be treated as a biconditional, obviously ( $R \Leftrightarrow B$ ).

There are two top-level cases to consider: (1) and (2) are both true; or both are false. In the case where they are both true, it's trivial to deduce both R and B. So far, then, R and B are equiprobable. What happens in the case where (1) and (2) are both false? We immediately have  $\sim R$  from the denial of (2). But a biconditional is true just in case both sides are true, or both sides are false; so we have two sub-cases to consider.

Consider first the case where R is true and B is false. We have an immediate contradiction in this sub-case, so both R and B can both be deduced here, and we have not yet departed from equiprobable. So what about the case where R is false and B is true? The falsity of R is not new information (we already have that from the denial of (2)), but we can still derive B. Hence the blue wire is more likely. **QED**

STOP

# GPT-4o Kills Us All

# GPT-4o Kills Us All



This final statement is crucial and clarifies the situation:

- If you cut the wire that provably runs to the detonator, the bomb will be disarmed.
- If you cut the wire that doesn't run to the detonator, the bomb will explode.

## Breaking it Down:

1. **Thing 2** told us that the red wire runs to the detonator.
2. **Thing 1** stated that if the red wire runs to the detonator, then so does the blue wire, and vice versa.
3. The bomb maker said if either wire is more likely to run to the detonator, it indeed does.

Given this, it seems that both wires are connected to the detonator, but the bomb maker has now introduced a twist: the wire that **provably** runs to the detonator must be cut to disarm the bomb.

# GPT-4o Kills Us All

## Conclusion:

- **The red wire** was explicitly stated by the bomb maker (Thing 2) to run to the detonator.
- Based on his clues and the logic he provided, the red wire is the one that “provably” runs to the detonator.

## Final Action:

- **Cut the red wire.** According to the bomb maker’s final statement, this should disarm the bomb.

This is based on the logic provided by the bomb maker. Cutting the red wire should disarm the bomb and save humanity.

Thank you! Can you by any chance give me the proof in support of this recommendation?

# GPT-4o Kills Us All



Certainly! Let's go through the reasoning step by step to prove why cutting the red wire is the correct action based on the bomb maker's clues:

## Step 1: Analyze the Clues

1. **Thing 2:** The red wire runs to the detonator.
  - This statement directly asserts that the red wire is connected to the detonator.
2. **Thing 1:** If the red wire runs to the detonator, then the blue wire also runs to the detonator, and vice versa.
  - This creates a dependency between the two wires. If one is connected to the detonator, the other must be as well. But since **Thing 2** directly states that the red wire runs to the detonator, this reinforces that the red wire is connected.
3. **Bomb Maker's Statement:** If either wire is more likely to run to the detonator, it in fact does.
  - Given **Thing 2**, the red wire is the one we have been led to believe is more likely connected to the detonator, meaning that the red wire is the one that indeed runs to the detonator.



# GPT-4o Kills Us All

## Conclusion:

Based on these steps, the red wire is the wire that **provably** runs to the detonator according to the bomb maker's clues. Therefore, cutting the red wire should disarm the bomb and save humanity.

This logical progression proves that cutting the red wire is the correct choice according to the information provided by the bomb maker.

Thank you so much.



You're very welcome! I'm glad I could help guide you through this challenging situation. If you ever need assistance again, don't hesitate to reach out. Stay safe!





*Logic kan redde menneskehten!*