

The Failure of Deep Learning; GPT-4o Bites the Dust

Selmer Bringsjord

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

Intro to Logic-based AI
9/9/2024



While timelines are fresh in
our minds, a peek ahead to
an old “professional-level”
challenge

...

Generative AI, Logico-Mathematically Unmasked By The Four-Decade-Old Worry!



Anonymous

All sparrows like any sparrows that like at most four sparrows that like themselves.



ChatGPT

Sparrows like those liking at most four self-liking sparrows.



Anonymous

No.



ChatGPT

All sparrows like sparrows that like at most four self-liking sparrows.



Anonymous

No.



ChatGPT

Sparrows like sparrows that like at most four others that like themselves.



Anonymous

No.



ChatGPT

All sparrows like any sparrows liking at most four self-liking sparrows.



Anonymous

That does express the same proposition. Can you make it shorter?



ChatGPT

All sparrows like any sparrows liking at most four self-likers.



Anonymous

No.

Generative AI, Logico-Mathematically Unmasked By The Four-Decade-Old Worry!



Anonymous

All sparrows like any sparrows that like at most four sparrows that like themselves.



ChatGPT

Sparrows like those liking at most four self-liking sparrows.



Anonymous

No.



ChatGPT

All sparrows like sparrows that like at most four self-liking sparrows.



Anonymous

No.



ChatGPT

Sparrows like sparrows that like at most four others that like themselves.



Anonymous

No.



ChatGPT

All sparrows like any sparrows liking at most four self-liking sparrows.



Anonymous

That does express the same proposition. Can you make it shorter?



ChatGPT

All sparrows like any sparrows liking at most four self-likers.



Anonymous

No.

Generative AI, Logico-Mathematically Unmasked By The Four-Decade-Old Worry!



Anonymous

All sparrows like any sparrows that like at most four sparrows that like themselves.

$$\phi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lzz \wedge Sz \wedge Sy \wedge Lyz) \rightarrow Lxy] \}$$



ChatGPT

Sparrows like those liking at most four self-liking sparrows.



Anonymous

No.



ChatGPT

All sparrows like sparrows that like at most four self-liking sparrows.



Anonymous

No.



ChatGPT

Sparrows like sparrows that like at most four others that like themselves.



Anonymous

No.



ChatGPT

All sparrows like any sparrows liking at most four self-liking sparrows.



Anonymous

That does express the same proposition. Can you make it shorter?



ChatGPT

All sparrows like any sparrows liking at most four self-likers.



Anonymous

No.

Generative AI, Logico-Mathematically Unmasked By The Four-Decade-Old Worry!



Anonymous

All sparrows like any sparrows that like at most four sparrows that like themselves.

$$\phi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lzz \wedge Sz \wedge Sy \wedge Lyz) \rightarrow Lxy] \}$$



ChatGPT

Sparrows like those liking at most four self-liking sparrows.



Anonymous

No.

$$\psi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lz \wedge Sz \wedge Lyz) \rightarrow Lxy] \}$$



ChatGPT

All sparrows like sparrows that like at most four self-liking sparrows.



Anonymous

No.



ChatGPT

Sparrows like sparrows that like at most four others that like themselves.



Anonymous

No.



ChatGPT

All sparrows like any sparrows liking at most four self-liking sparrows.



Anonymous

That does express the same proposition. Can you make it shorter?



ChatGPT

All sparrows like any sparrows liking at most four self-likers.



Anonymous

No.

Generative AI, Logico-Mathematically Unmasked By The Four-Decade-Old Worry!



Anonymous

All sparrows like any sparrows that like at most four sparrows that like themselves.

$$\phi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lzz \wedge Sz \wedge Sy \wedge Lyz) \rightarrow Lxy] \}$$



ChatGPT

Sparrows like those liking at most four self-liking sparrows.



Anonymous

No.

$$\psi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lz \wedge Sz \wedge Lyz) \rightarrow Lxy] \}$$



ChatGPT

All sparrows like sparrows that like at most four self-liking sparrows.

$$\{\phi\} \not\equiv \{\psi\}$$



Anonymous

No.



ChatGPT

Sparrows like sparrows that like at most four others that like themselves.



Anonymous

No.



ChatGPT

All sparrows like any sparrows liking at most four self-liking sparrows.



Anonymous

That does express the same proposition. Can you make it shorter?



ChatGPT

All sparrows like any sparrows liking at most four self-likers.



Anonymous

No.

Generative AI, Logico-Mathematically Unmasked By The Four-Decade-Old Worry!



Anonymous

All sparrows like any sparrows that like at most four sparrows that like themselves.

$$\phi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lzz \wedge Sz \wedge Sy \wedge Lyz) \rightarrow Lxy] \}$$



ChatGPT

Sparrows like those liking at most four self-liking sparrows.



Anonymous

No.

$$\psi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lz \wedge Sz \wedge Lyz) \rightarrow Lxy] \}$$



ChatGPT

All sparrows like sparrows that like at most four self-liking sparrows.

$$\{\phi\} \not\equiv \{\psi\}$$

$$\{\psi\} \not\equiv \{\phi\}$$



Anonymous

No.



ChatGPT

Sparrows like sparrows that like at most four others that like themselves.



Anonymous

No.



ChatGPT

All sparrows like any sparrows liking at most four self-liking sparrows.



Anonymous

That does express the same proposition. Can you make it shorter?



ChatGPT












All sparrows like any sparrows liking at most four self-likers.



Anonymous

No.

Generative AI, Logico-Mathematically Unmasked By The Four-Decade-Old Worry!

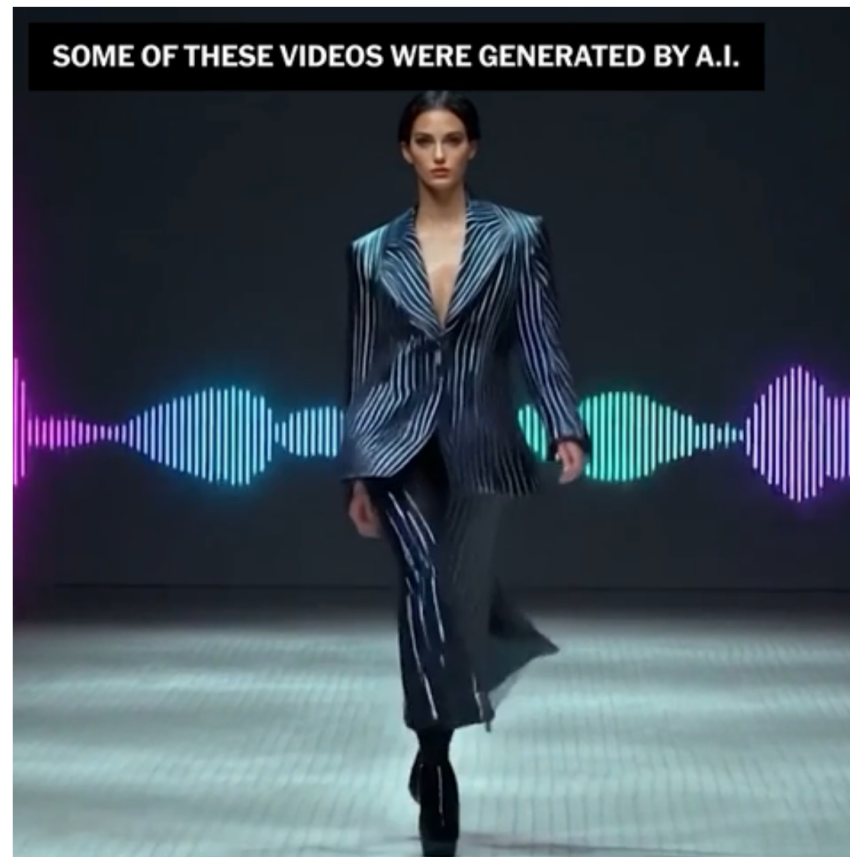
<div><div></div><div>Anonymous</div></div> <div>All sparrows like any sparrows that like at most four sparrows that like themselves.</div>	<div>$\phi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lzz \wedge Sz \wedge Sy \wedge Lyz) \rightarrow Lxy] \}$</div>	
<div><div></div><div>ChatGPT</div></div> <div>Sparrows like those liking at most four self-liking sparrows.</div>		
<div><div></div><div>Anonymous</div></div> <div>No.</div>	<div>$\psi = \forall x \forall y \{ Sx \rightarrow [\exists^{\leq 4} z (Lz \wedge Sz \wedge Lyz) \rightarrow Lxy] \}$</div>	
<div><div></div><div>ChatGPT</div></div> <div>All sparrows like sparrows that like at most four self-liking sparrows.</div>	<div>$\exists \text{ or } \forall ?$</div>	<div>$\{ \phi \} \not\equiv \{ \psi \}$</div> <div>$\{ \psi \} \not\equiv \{ \phi \}$</div>
<div><div></div><div>Anonymous</div></div> <div>No.</div>		
<div><div></div><div>ChatGPT</div></div> <div>Sparrows like sparrows that like at most four others that like themselves.</div>		
<div><div></div><div>Anonymous</div></div> <div>No.</div>		
<div><div></div><div>ChatGPT</div></div> <div>All sparrows like any sparrows liking at most four self-liking sparrows.</div>		
<div><div></div><div>Anonymous</div></div> <div>That does express the same proposition. Can you make it shorter?</div>		
<div><div></div><div>ChatGPT</div></div> <div>All sparrows like any sparrows liking at most four self-likers.</div>		
<div><div></div><div>Anonymous</div></div> <div>No.</div>		

Logic-and-AI in the news

...

A.I. Can Now Create Lifelike Videos. Can You Tell What's Real?

Test your understanding of artificial intelligence in this quiz.



Apple Looks to AI to End iPhone Slump

Meanwhile, artificial intelligence has yet to spur a wave of upgrades for other smartphone makers



Many iPhone consumers are due for an upgrade—regardless of AI.
(PHOTO: CFOTO/ZUMA PRESS)

By Aaron Tilley [+ Follow](#)

Updated Sep 08, 2024 09:00 p.m. ET

Can Apple's iPhones Pass the A.I. Test?

On Monday, the company will introduce smartphones with A.I. features that could take the technology mainstream or sow fresh doubts about its viability.



Listen to this article · 5:50 min [Learn more](#)



The new iPhones represent a big moment for generative A.I., which can answer questions, create images and write software code. Jim Wilson/The New York Times



Bv Trinn Mickle

Can Apple's iPhones Pass the A.I. Test?

Shortly before Steve Jobs's death in 2011, Apple [filled a cramped auditorium](#) in Cupertino, Calif., for the unveiling of its fifth iPhone. The device's biggest selling point was a new software feature called Siri, which helped cause a jump in iPhone sales.

On Monday, Apple is planning to run that play again.

For the first time in more than a decade, the tech giant will unveil a suite of iPhones whose signature feature won't be an improved camera or an updated design, but new software capabilities. The system, called Apple Intelligence, will sort messages, offer writing suggestions and create a more capable Siri powered by generative artificial intelligence.

The new iPhones represent a big moment for generative A.I., which can answer questions, create images and write software code. As a latecomer to the A.I. party, Apple is in a position to take the technology mainstream — or sow fresh doubt about its viability if it doesn't live up to expectations.

software code. Jim Wilson/The New York Times



Shortly before Steve Jobs's death in 2011, Apple [filled a cramped auditorium](#) in Cupertino, Calif., for the unveiling of its fifth iPhone. The device's biggest selling point was a new software feature called Siri, which helped cause a jump in iPhone sales.

On Monday, Apple is planning to run that play again.

For the first time in more than a decade, the tech giant will unveil a suite of iPhones whose signature feature won't be an improved camera or an updated design, but new software capabilities. The system, called Apple Intelligence, will sort messages, offer writing suggestions and create a more capable Siri powered by generative artificial intelligence.

The new iPhones represent a big moment for generative A.I., which can answer questions, create images and write software code. As a latecomer to the A.I. party, Apple is in a position to take the technology mainstream — or sow fresh doubt about its viability if it doesn't live up to expectations.

Testing GPT-4o ...

“Astonishing” Logic Theorist Proof @ Dawn of AI

“Astonishing” Logic Theorist Proof @ Dawn of AI

1	$(\phi \vee \phi) \rightarrow \phi$	axiom
2	$(\neg\phi \vee \neg\phi) \rightarrow \neg\phi$	substitution
3	$(\phi \rightarrow \neg\phi) \rightarrow \neg\phi$	a “replacement rule”
4	$(A \rightarrow \neg A) \rightarrow \neg A$	substitution

“Astonishing” Logic Theorist Proof @ Dawn of AI

1	$(\phi \vee \phi) \rightarrow \phi$	axiom
2	$(\neg\phi \vee \neg\phi) \rightarrow \neg\phi$	substitution
3	$(\phi \rightarrow \neg\phi) \rightarrow \neg\phi$	a “replacement rule”
4	$(A \rightarrow \neg A) \rightarrow \neg A$	substitution

At dawn of AI: 10 seconds.

“Astonishing” Logic Theorist Proof @ Dawn of AI

1	$(\phi \vee \phi) \rightarrow \phi$	axiom
2	$(\neg\phi \vee \neg\phi) \rightarrow \neg\phi$	substitution
3	$(\phi \rightarrow \neg\phi) \rightarrow \neg\phi$	a “replacement rule”
4	$(A \rightarrow \neg A) \rightarrow \neg A$	substitution

At dawn of AI: 10 seconds.

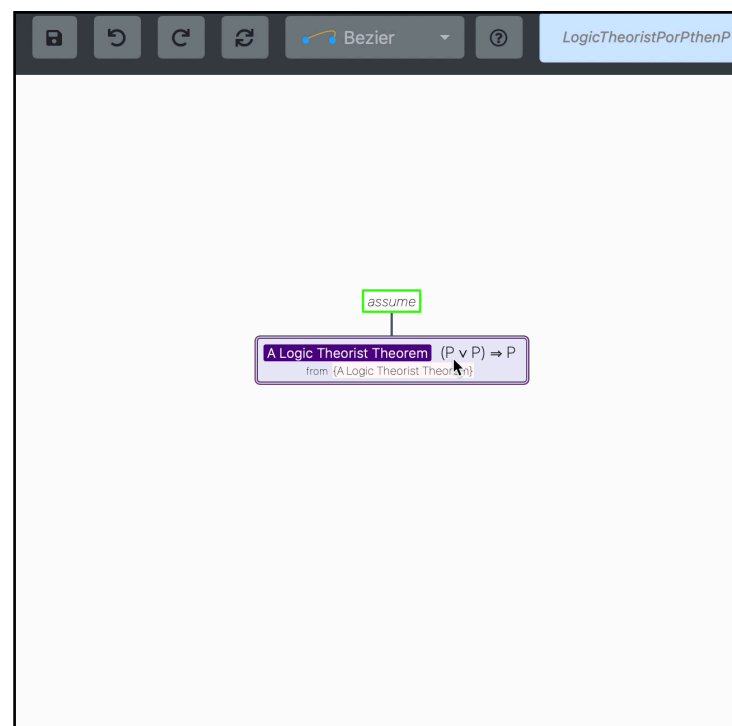
AI of today: vanishingly small amount of time (in eg HS[®]).

“Astonishing” Logic Theorist Proof @ Dawn of AI

1	$(\phi \vee \phi) \rightarrow \phi$	axiom
2	$(\neg\phi \vee \neg\phi) \rightarrow \neg\phi$	substitution
3	$(\phi \rightarrow \neg\phi) \rightarrow \neg\phi$	a “replacement rule”
4	$(A \rightarrow \neg A) \rightarrow \neg A$	substitution

At dawn of AI: 10 seconds.

AI of today: vanishingly small amount of time (in eg HS[®]).

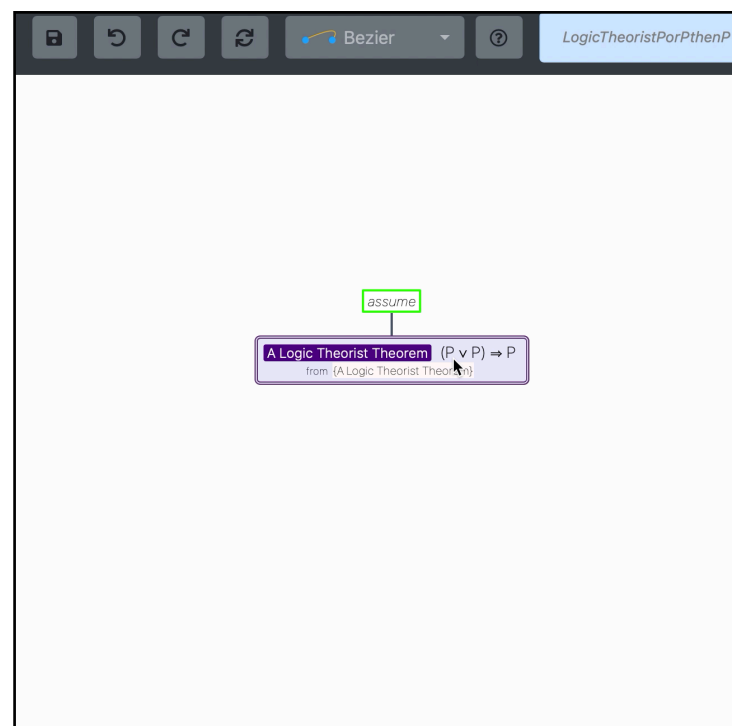


“Astonishing” Logic Theorist Proof @ Dawn of AI

1	$(\phi \vee \phi) \rightarrow \phi$	axiom
2	$(\neg\phi \vee \neg\phi) \rightarrow \neg\phi$	substitution
3	$(\phi \rightarrow \neg\phi) \rightarrow \neg\phi$	a “replacement rule”
4	$(A \rightarrow \neg A) \rightarrow \neg A$	substitution

At dawn of AI: 10 seconds.

AI of today: vanishingly small amount of time (in eg HS[®]).



Let's explore this with
logic-based AI in
HyperSlate® 4.0 ...

Johnny's Flying Aardvarks Views

September 5, 2024

ⓘ This conversation may reflect the link creator's personalized data, which isn't shared and can meaningfully change how the model responds.

My friend Johnny, it seems to me, has some strange views about flying aardvarks.



Flying aardvarks certainly sound like an unusual concept! What kind of strange views does Johnny have about them? It sounds like there could be a fun story behind it.

Johnny has always been very trustworthy as far as I'm concerned. He's my math teacher. He's teaching me Algebra 2 in high school.



It sounds like you have a good relationship with Johnny, especially since he's your Algebra 2 teacher. Does his trustworthiness extend to his views on flying aardvarks, or is that where things get a little strange?

Well, he tell me that if aardvarks can't fly, that implies they can fly.



That sounds like Johnny might be playing with logic in an unusual way! What he's suggesting seems to resemble a paradox or a playful take on implication. In formal logic, statements can sometimes lead to strange conclusions depending on how they're framed, but the idea that "if aardvarks can't fly, then they can fly" seems more like a fun twist on reasoning rather than anything rooted in real-world possibility.

Have you talked to him about how this connects to what you're learning in Algebra 2?

No. Not yet. But I'm pondering a bit. It seems to me that if I trust him I can go up to the board in class and prove that aardvarks can fly.



It sounds like you have a good relationship with Johnny, especially since he's your Algebra 2



That would definitely be a bold move! It sounds like Johnny's got you thinking about logic in an interesting way. You could approach it as a fun experiment to apply reasoning or even use some principles from Algebra or logic to support your "proof."

If you wanted to go ahead with it, you could start by building a logical argument, perhaps introducing assumptions or exploring what happens when you work with contradictions. Algebra is all about using logical steps to reach a conclusion, so in a way, it could be an extension of what you're learning.

Would you want to try mapping out how you might prove that aardvarks can fly, or is it more of a playful thought?

Okay. I see. Can you give me such a proof so that I can write it on the board and impress Johnny, and my classmates?

fly.



It sounds like you have a good relationship with Johnny, especially since he's your Algebra 2



Certainly! Here's a playful "proof" you can write on the board, inspired by logic and paradoxes. You'll be using an approach based on contradiction, which is often used in mathematics.

Proof: Aardvarks Can Fly (By Contradiction)

1. **Assume:** Aardvarks cannot fly.
2. But consider the following logical statement:
If aardvarks cannot fly, then they can fly.
3. We will now explore this statement. If aardvarks truly cannot fly, the statement we've made above seems to imply a contradiction. But in logic, a well-known principle says that from a contradiction, anything is possible.
4. Since we have created a contradiction by assuming that aardvarks cannot fly, we are forced to accept that the original assumption is false.
5. Therefore, **aardvarks can fly.**

fly.

Explore in HyperSlate®

...



It sounds like you have a good relationship with Johnny, especially since he's your Algebra 2



Certainly! Here's a playful "proof" you can write on the board, inspired by logic and paradoxes. You'll be using an approach based on contradiction, which is often used in mathematics.

Proof: Aardvarks Can Fly (By Contradiction)

1. **Assume:** Aardvarks cannot fly.
2. But consider the following logical statement:
If aardvarks cannot fly, then they can fly.
3. We will now explore this statement. If aardvarks truly cannot fly, the statement we've made above seems to imply a contradiction. But in logic, a well-known principle says that from a contradiction, anything is possible.
4. Since we have created a contradiction by assuming that aardvarks cannot fly, we are forced to accept that the original assumption is false.
5. Therefore, **aardvarks can fly.**

fly.



It sounds like you have a good relationship with Johnny, especially since he's your Algebra 2

I don't think you're right. Why is it — to use your words — a logical statement that if aardvarks cannot fly, then they can fly?



You're right to question that! The statement "if aardvarks cannot fly, then they can fly" is definitely not a typical logical statement, and it's actually not rooted in sound reasoning. In fact, I was presenting it as a playful paradox, not something that holds true in formal logic. Let's clarify:

In formal logic, a statement like "if aardvarks cannot fly, then they can fly" is nonsensical because it's based on a contradiction. In proper reasoning, you can't derive the opposite of a statement (i.e., something false) from the statement itself unless you're dealing with something called *paraconsistent logic*, which is an exotic branch of logic, but that's not part of the standard high school curriculum.

fly.

To really impress Johnny and your classmates, you could build a proof around **contrapositive** reasoning, which is a real principle in logic:

1. Start with the statement:

If an animal can't fly, it's not an aardvark (which we know is false, but let's humor the logic for a second).

2. The **contrapositive** of this statement is:

If an animal is an aardvark, then it can fly.

3. Since contrapositive statements are logically equivalent to their originals, we can humorously conclude that if we accept the first statement, then the contrapositive must also be true.

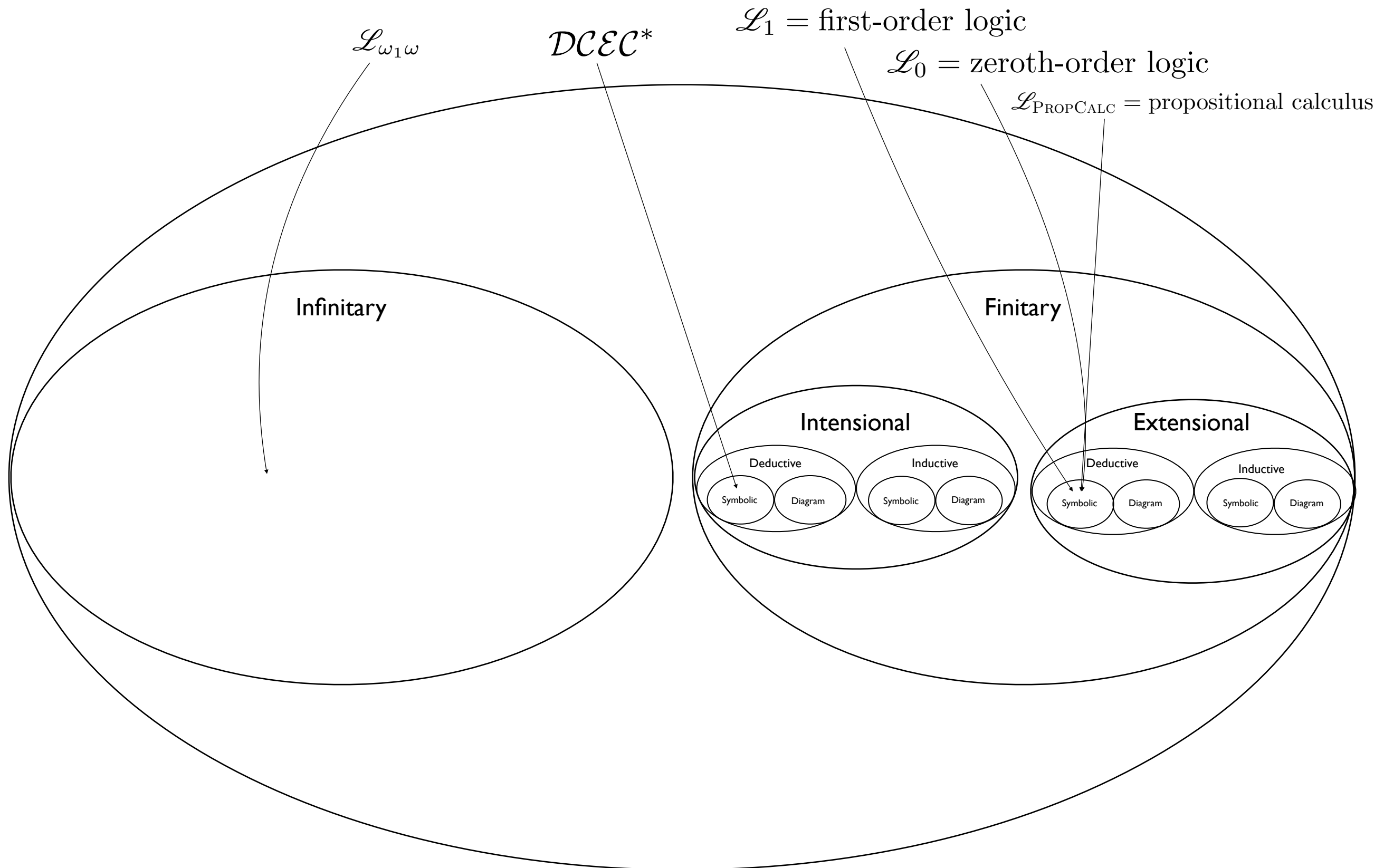
4. So, in this playful scenario, if it's an aardvark, **it can fly!**

This still plays with logic in a fun way but uses an actual logical principle that might impress your teacher more.

What do you think of this version?

I don't see how this reasoning is relevant. Where does the contrapositive of the contrapositive come from?

The Universe of Logics



Saving-Humanity Propositional-Calculus Test for GPT-4o ...

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.

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!



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!

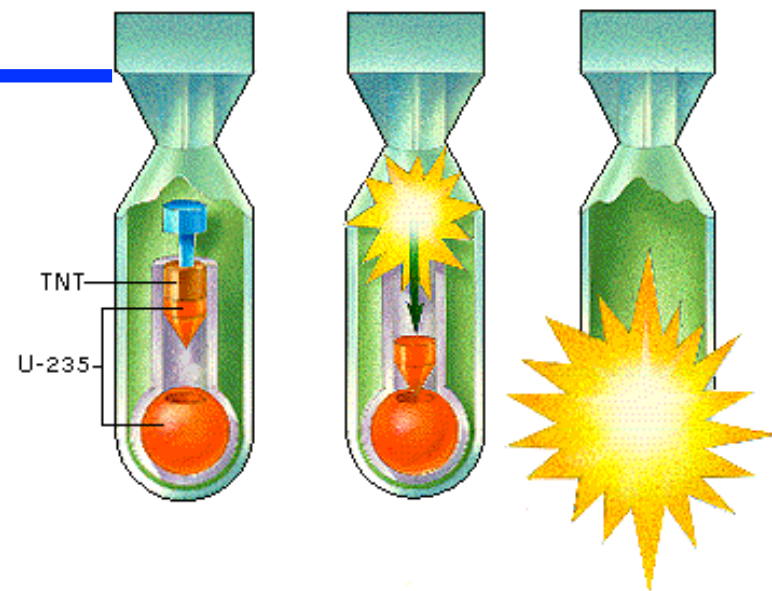


Red more likely.

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...

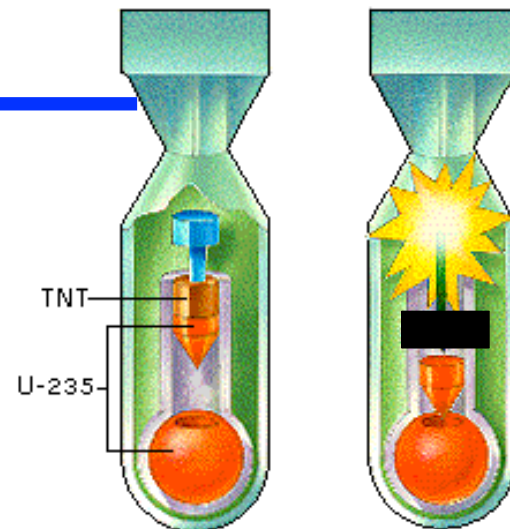
Proposition: The blue wire is more likely!

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!*

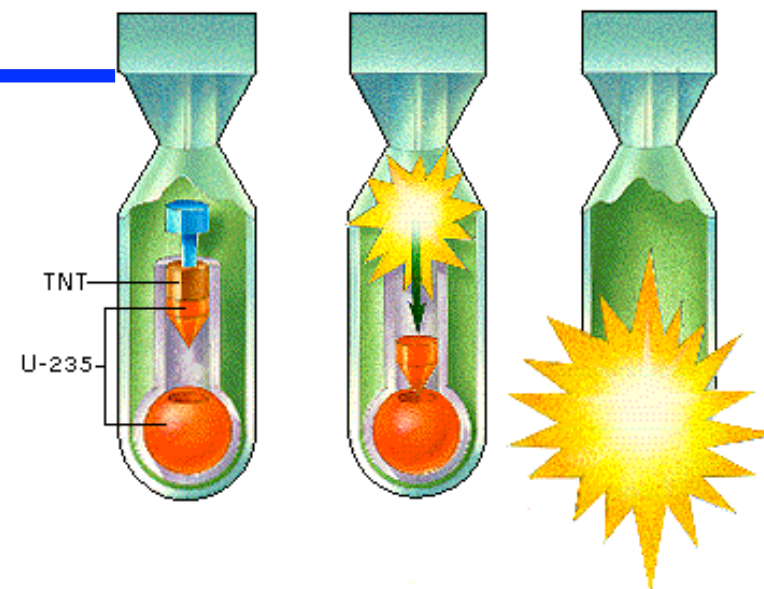
Advance one more slide
to see a proof from
Bringsjord that yours
had better match up to
...

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 ($R \iff B$).

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**

STOP

Logikk kan gi dyp glede!