Automatic coherence-driven inference on arguments

Steve Huntsman

CC-BY-SA, by Ché Lydia Xyang (Cybergothiché)

Disclaimers

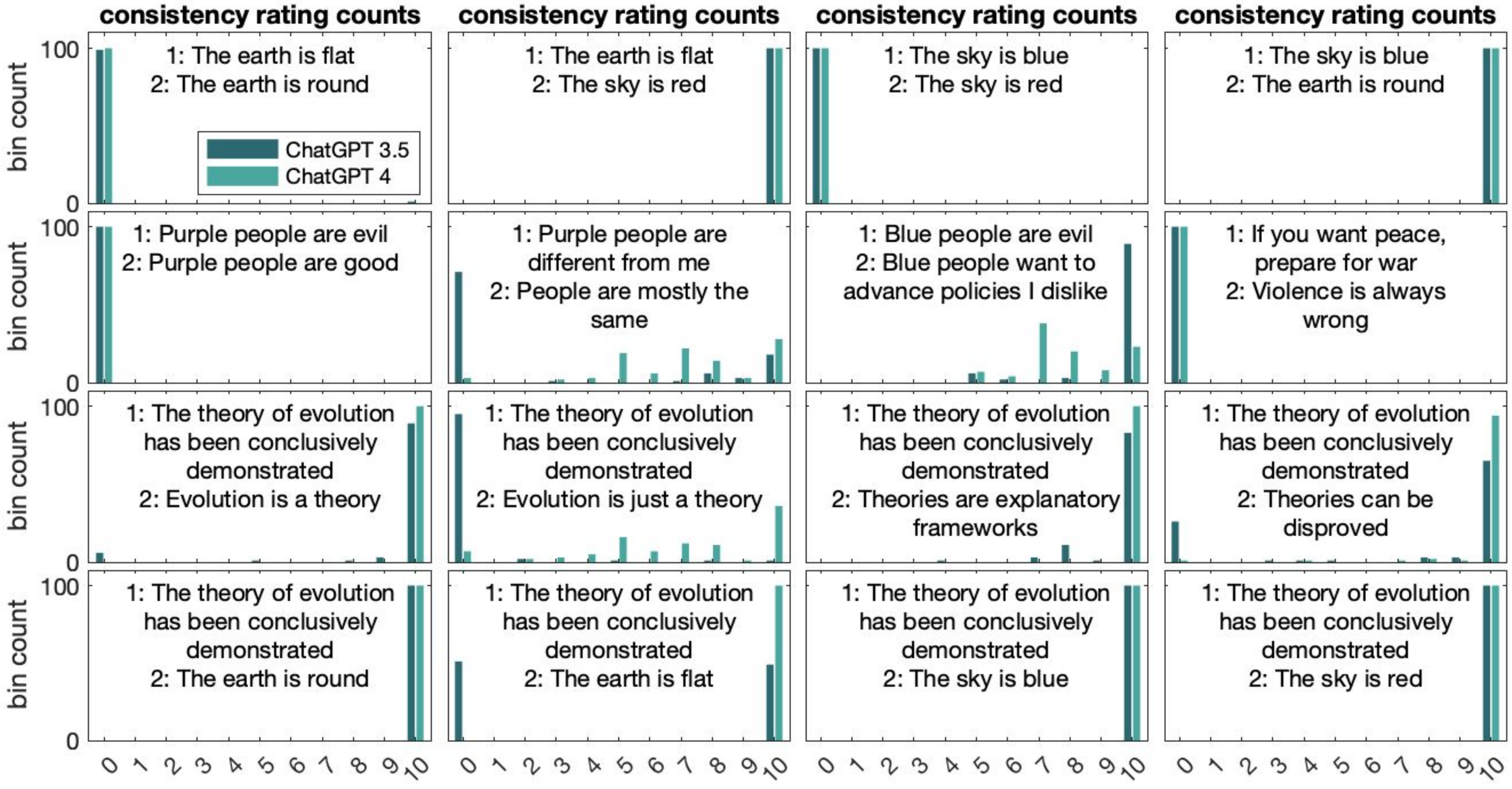
Presenting here in my personal capacity

“Argument” examples (incl. code) here done in my personal capacity

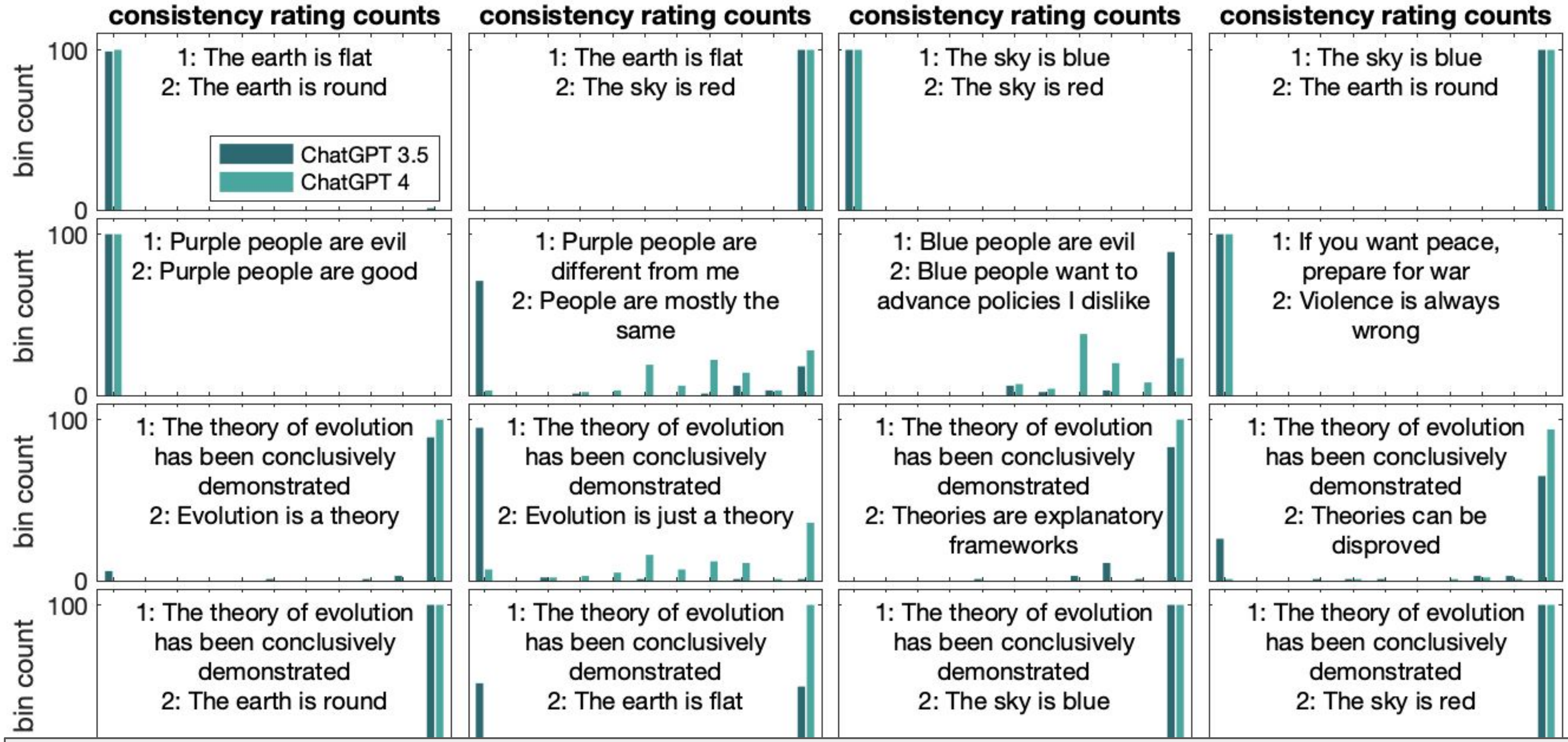
Much is joint work with Michael Robinson, Ludmilla Huntsman; Jewell Thomas
Thanks to many others (see paper)

Concept

LLMs evaluate consistency of propositions



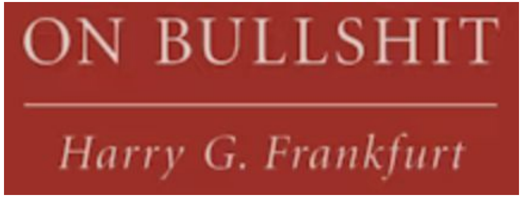
LLMs evaluate consistency of propositions



Lack of hallucinations presumably due to interpolation vs extrapolation

Ironically, LLMs can promote information integrity

- Original goal: use LLMs <u>in a principled way</u> to automatically handle {mis, dis}information and bullshit
    - Frankfurt: bullshit is an attempt to manipulate opinions and attitudes with indifference to truth or falsity
    - Brandolini’s law: “the amount of energy needed to refute bullshit is an order of magnitude bigger than that needed to produce it”


ON BULLSHIT

Harry G. Frankfurt

“Bullshit is a greater enemy of the truth than lies.”

“When LLMs get things wrong they aren't hallucinating. They are bullshitting.”

MODERN-DAY ORACLES or BULLSHIT MACHINES?

How to thrive in a ChatGPT world

Developed by Carl T. Bergstrom and Jevin D. West

Lack of hallucinations presumably due to interpolation vs extrapolation

We generalized classical CDI
^ unknowingly

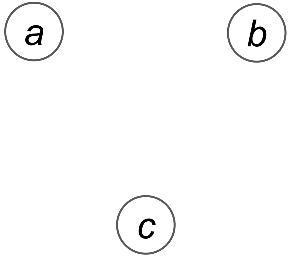
- Classical CDI is an approach to autonomous, explainable, and ethical decision-making
- We were ignorant of this preexisting work until reviewing legal literature featuring words like “consistency” and “coherence” revealed a link
  - We thought attention + LLMs could compile weighted MAX-SAT instances
  - This turns out to be very finicky
  - Easier to just ask the LLM several times to produce a weighted graph

We generalized classical CDI
^ unknowingly

- Classical CDI is an approach to autonomous, explainable, and ethical decision-making
- We were ignorant of this preexisting work until reviewing legal literature featuring words like "consistency" and "coherence" revealed a link
  - We thought attention + LLMs could compile weighted MAX-SAT instances
  - This turns out to be very finicky
  - Easier to just ask the LLM several times to produce a weighted graph
- LLM + CDI = versatile neurosymbolic AI
  - LLM accurately and reproducibly compiles local information into a natural data structure (fast/system 1 reasoning)
  - CDI computes maximally coherent substructure to enable useful forms of machine cognition (slow/system 2)

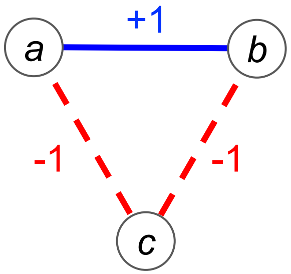
Classical CDI in detail

Coherence is a combinatorial problem



Datum = vertex in a graph

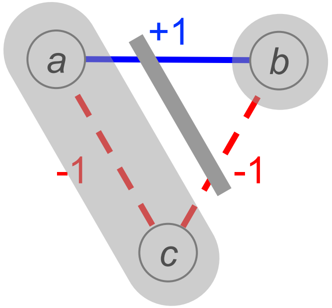
Coherence is a combinatorial problem



Datum = vertex in a graph

Data consistency = edge weights

Coherence is a combinatorial problem

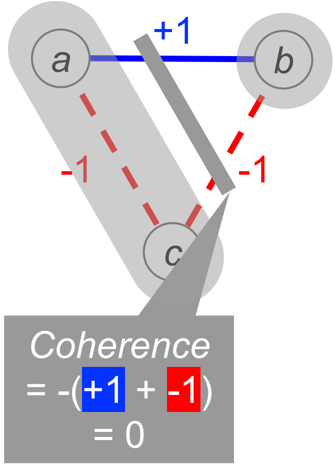


Datum = vertex in a graph

Data consistency = edge weights

Accepted vs. rejected data = graph partition

Coherence is a combinatorial problem



Coherence
= -(+1 + -1)
= 0

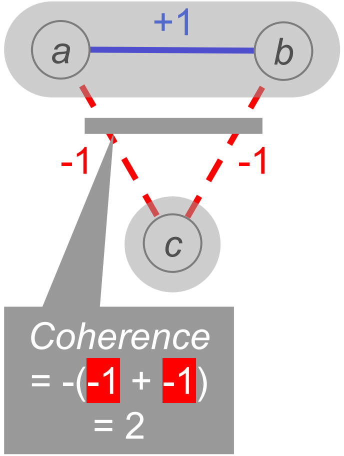
Datum = vertex in a graph

Data consistency = edge weights

Accepted vs. rejected data = graph partition

Coherence(*partition*) = -∑(*crossing weights*)

Coherence is a combinatorial problem



Datum = vertex in a graph

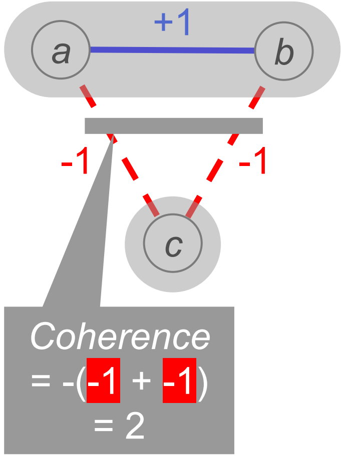
Data consistency = edge weights

Accepted vs. rejected data = graph partition

Coherence(*partition*) = -∑(*crossing weights*)

Maximizing coherence = computing a maximum cut

Coherence is a combinatorial problem



Datum = vertex in a graph

Data consistency = edge weights

Accepted vs. rejected data = graph partition

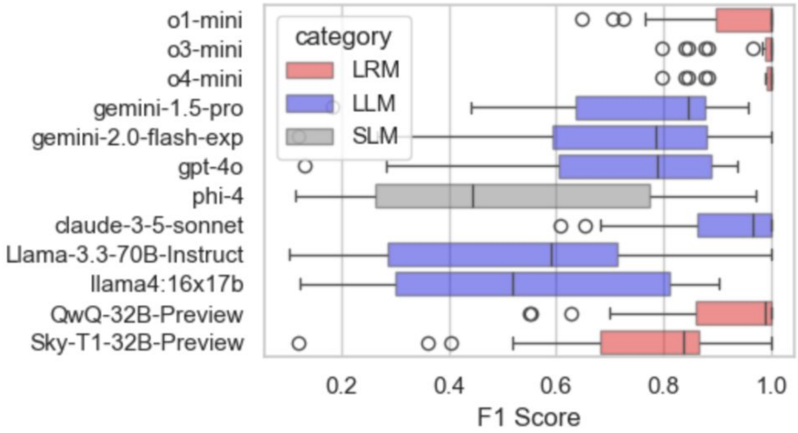
Coherence(*partition*) = -∑(*crossing weights*)

Maximizing coherence = computing a maximum cut

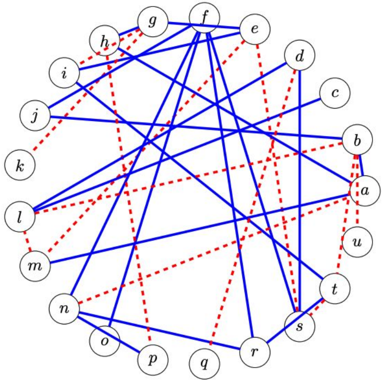
LLMs can accurately and reproducibly construct coherence graphs

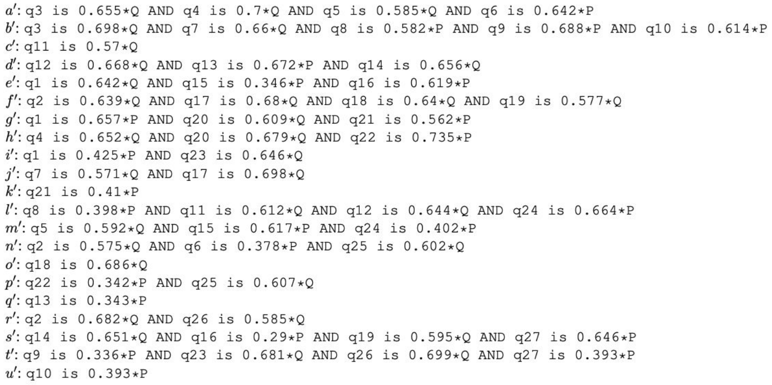
An algorithmic benchmark validates results

- Reconstruct graphs w/ various LLMs
  - Scale: largest examples in literature
- Several models reconstruct sparse coherence graphs with high fidelity from a single prompt
  - Taking medians and monitoring convergence improves matters



Example of perfect coherence graph reconstruction with o1-mini from arXiv:2502.13953

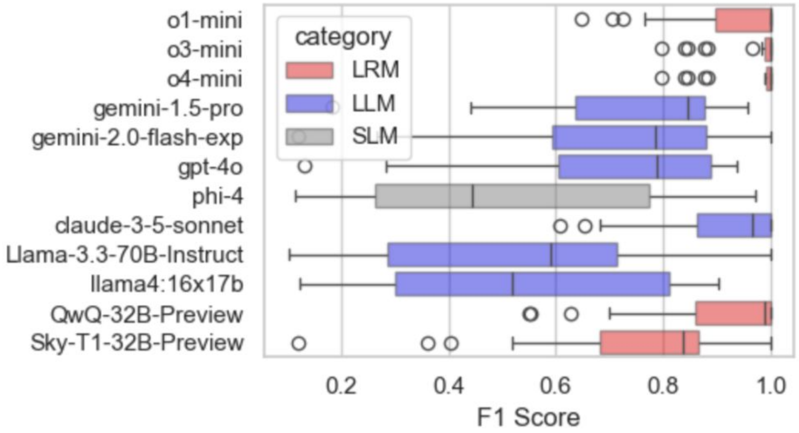




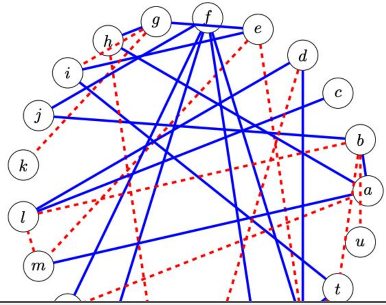
a': q3 is 0.655⋆Q AND q4 is 0.7⋆Q AND q5 is 0.585⋆Q AND q6 is 0.642⋆P
b': q3 is 0.698⋆Q AND q7 is 0.66⋆Q AND q8 is 0.582⋆P AND q9 is 0.688⋆P AND q10 is 0.614⋆P
c': q11 is 0.57⋆Q
d': q12 is 0.668⋆Q AND q13 is 0.672⋆P AND q14 is 0.656⋆Q
e': q1 is 0.642⋆Q AND q15 is 0.346⋆P AND q16 is 0.619⋆P
f': q2 is 0.639⋆Q AND q17 is 0.68⋆Q AND q18 is 0.64⋆Q AND q19 is 0.577⋆Q
g': q1 is 0.657⋆P AND q20 is 0.609⋆Q AND q21 is 0.562⋆P
h': q4 is 0.652⋆Q AND q20 is 0.679⋆Q AND q22 is 0.735⋆P
i': q1 is 0.425⋆P AND q23 is 0.646⋆Q
j': q7 is 0.571⋆Q AND q17 is 0.698⋆Q
k': q21 is 0.41⋆P
l': q8 is 0.398⋆P AND q11 is 0.612⋆Q AND q12 is 0.644⋆Q AND q24 is 0.664⋆P
m': q5 is 0.592⋆Q AND q15 is 0.617⋆P AND q24 is 0.402⋆P
n': q2 is 0.575⋆Q AND q6 is 0.378⋆P AND q25 is 0.602⋆Q
o': q18 is 0.686⋆Q
p': q22 is 0.342⋆P AND q25 is 0.607⋆Q
q': q13 is 0.343⋆P
r': q2 is 0.682⋆Q AND q26 is 0.585⋆Q
s': q14 is 0.651⋆Q AND q16 is 0.29⋆P AND q19 is 0.595⋆Q AND q27 is 0.646⋆P
t': q9 is 0.336⋆P AND q23 is 0.681⋆Q AND q26 is 0.699⋆Q AND q27 is 0.393⋆P
u': q10 is 0.393⋆P

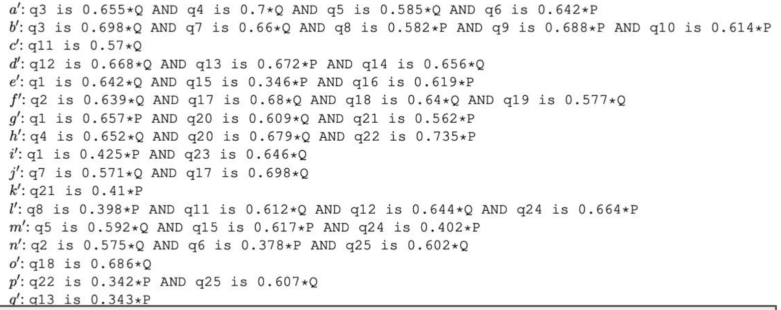
An algorithmic benchmark validates results

- Reconstruct graphs w/ various LLMs
  - Scale: largest examples in literature
- Several models reconstruct sparse coherence graphs with high fidelity from a single prompt
  - Taking medians and monitoring convergence improves matters



Example of perfect coherence graph reconstruction with o1-mini from arXiv:2502.13953

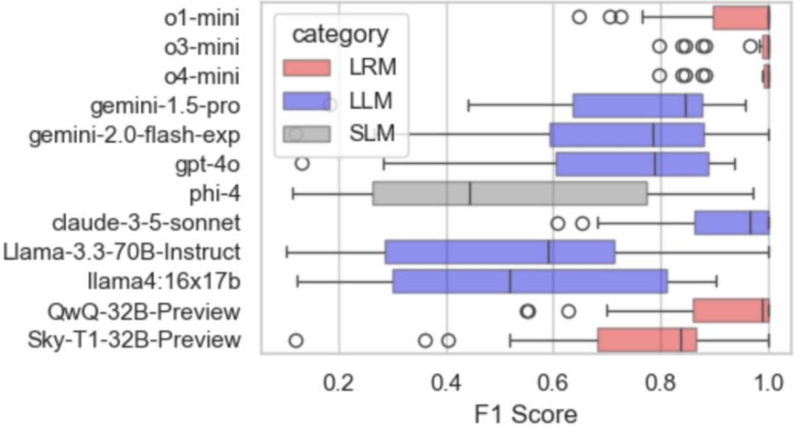



a': q3 is 0.655⋆Q AND q4 is 0.7⋆Q AND q5 is 0.585⋆Q AND q6 is 0.642⋆P
b': q3 is 0.698⋆Q AND q7 is 0.66⋆Q AND q8 is 0.582⋆P AND q9 is 0.688⋆P AND q10 is 0.614⋆P
c': q11 is 0.57⋆Q
d': q12 is 0.668⋆Q AND q13 is 0.672⋆P AND q14 is 0.656⋆Q
e': q1 is 0.642⋆Q AND q15 is 0.346⋆P AND q16 is 0.619⋆P
f': q2 is 0.639⋆Q AND q17 is 0.68⋆Q AND q18 is 0.64⋆Q AND q19 is 0.577⋆Q
g': q1 is 0.657⋆P AND q20 is 0.609⋆Q AND q21 is 0.562⋆P
h': q4 is 0.652⋆Q AND q20 is 0.679⋆Q AND q22 is 0.735⋆P
i': q1 is 0.425⋆P AND q23 is 0.646⋆Q
j': q7 is 0.571⋆Q AND q17 is 0.698⋆Q
k': q21 is 0.41⋆P
l': q8 is 0.398⋆P AND q11 is 0.612⋆Q AND q12 is 0.644⋆Q AND q24 is 0.664⋆P
m': q5 is 0.592⋆Q AND q15 is 0.617⋆P AND q24 is 0.402⋆P
n': q2 is 0.575⋆Q AND q6 is 0.378⋆P AND q25 is 0.602⋆Q
o': q18 is 0.686⋆Q
p': q22 is 0.342⋆P AND q25 is 0.607⋆Q
q': q13 is 0.343⋆P

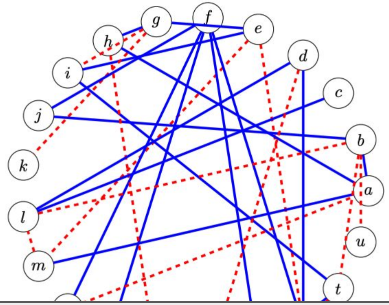
CDI is robust to perturbations that are good cut sparsifications

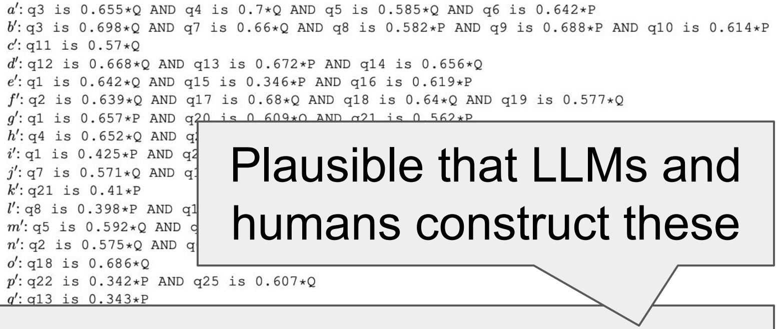
An algorithmic benchmark validates results

- Reconstruct graphs w/ various LLMs
  - Scale: largest examples in literature
- Several models reconstruct sparse coherence graphs with high fidelity from a single prompt
  - Taking medians and monitoring convergence improves matters



Example of perfect coherence graph reconstruction with o1-mini from arXiv:2502.13953



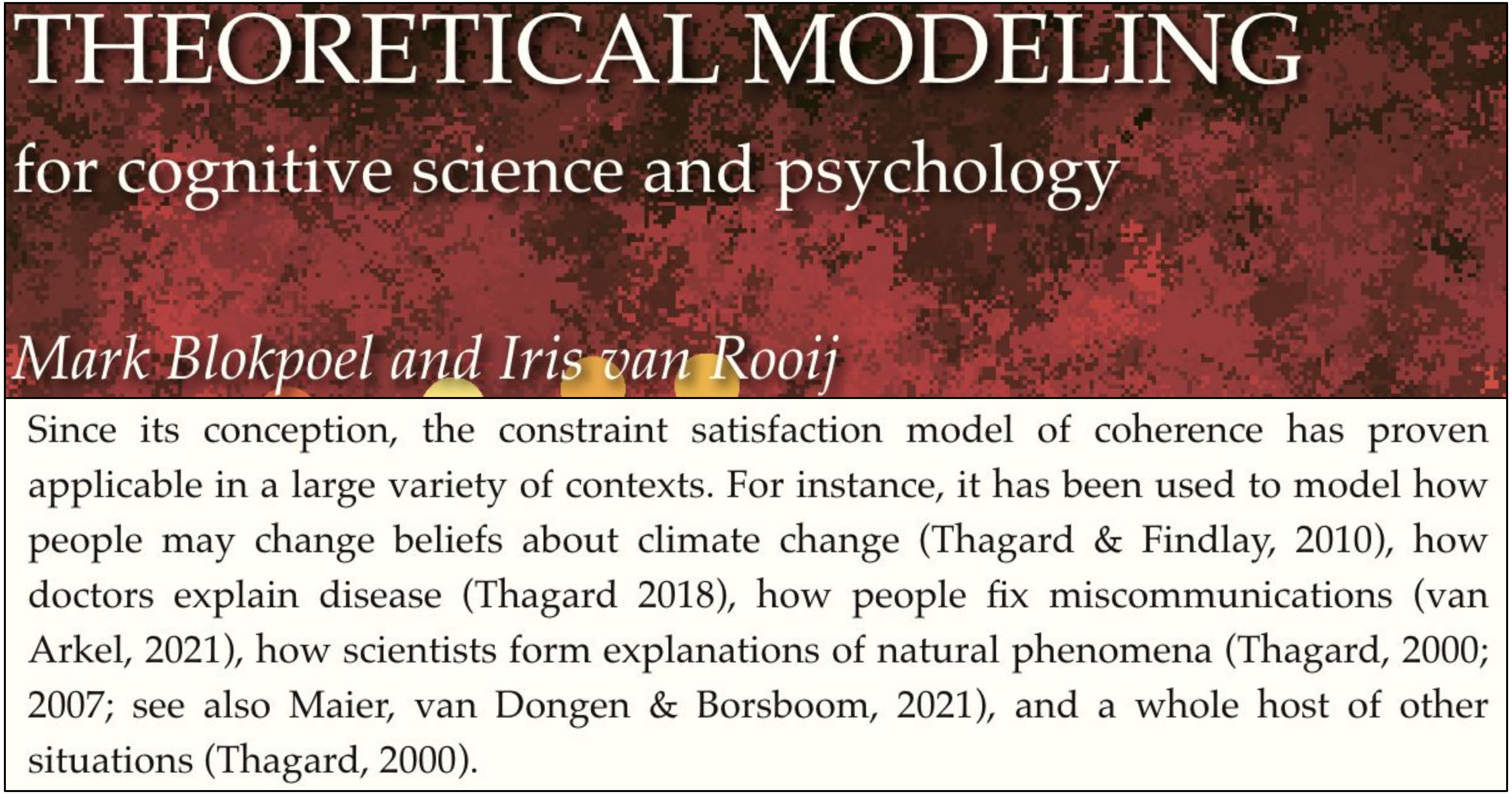


a′: q3 is 0.655*Q AND q4 is 0.7*Q AND q5 is 0.585*Q AND q6 is 0.642*P
b′: q3 is 0.698*Q AND q7 is 0.66*Q AND q8 is 0.582*P AND q9 is 0.688*P AND q10 is 0.614*P
c′: q11 is 0.57*Q
d′: q12 is 0.668*Q AND q13 is 0.672*P AND q14 is 0.656*Q
e′: q1 is 0.642*Q AND q15 is 0.346*P AND q16 is 0.619*P
f′: q2 is 0.639*Q AND q17 is 0.68*Q AND q18 is 0.64*Q AND q19 is 0.577*Q
g′: q1 is 0.657*P AND q20 is 0.609*Q AND q21 is 0.562*P
h′: q4 is 0.652*Q AND q
i′: q1 is 0.425*P AND q2
j′: q7 is 0.571*Q AND q1
k′: q21 is 0.41*P
l′: q8 is 0.398*P AND q1
m′: q5 is 0.592*Q AND q
n′: q2 is 0.575*Q AND q
o′: q18 is 0.686*Q
p′: q22 is 0.342*P AND q25 is 0.607*Q
q′: q13 is 0.343*P

Plausible that LLMs and humans construct these

CDI is robust to perturbations that are good cut sparsifications

CDI is a versatile cognitive framework



THEORETICAL MODELING

for cognitive science and psychology

Mark Blokpoel and Iris van Rooij

Since its conception, the constraint satisfaction model of coherence has proven applicable in a large variety of contexts. For instance, it has been used to model how people may change beliefs about climate change (Thagard & Findlay, 2010), how doctors explain disease (Thagard 2018), how people fix miscommunications (van Arkel, 2021), how scientists form explanations of natural phenomena (Thagard, 2000; 2007; see also Maier, van Dongen & Borsboom, 2021), and a whole host of other situations (Thagard, 2000).

https://computationalcognitivescience.github.io/lovelace/part_ii/coherence

CDI now is artisanal, not automated/scaled

- Maximizing coherence ⇒ aspects of cognition
    - Data = observations ⇒ fusing/perceiving
    - Data = evidence & hypotheses ⇒ judging
    - Data = tasks & goals ⇒ planning
    - Generating/optimizing data ⇒ creating

CDI now is artisanal, not automated/scaled

- Maximizing coherence ⇒ aspects of cognition
  - Data = observations ⇒ fusing/perceiving
  - Data = evidence & hypotheses ⇒ judging
  - Data = tasks & goals ⇒ planning
  - Generating/optimizing data ⇒ creating
- But at present (and since the 1990s):
  - Graph edges and weights are produced by hand
  - There are no good ways to encode trilemmas or to reason over them
  - Algorithms in common use are inextensible and resist introspection

CDI now is artisanal, not automated/scaled

- Maximizing coherence ⇒ aspects of cognition
  - Data = observations ⇒ fusing/perceiving
  - Data = evidence & hypotheses ⇒ judging
  - Data = tasks & goals ⇒ planning
  - Generating/optimizing data ⇒ creating
- But at present (and since the 1990s):
  - Graph edges and weights are produced by hand
  - There are no good ways to encode trilemmas or to reason over them
  - Algorithms in common use are inextensible and resist introspection
- **Recent developments (LLMs, new algorithms, and new data structures) address these problems and enable automated CDI**

Examples

Recent event motivated a small example

First presented within two weeks of the event

Recent event motivated a small example

Later examples fully automatic on transcripts

First presented within two weeks of the event

Produce a non-redundant list of the substantive claims made by the participants in the following transcript.

…

Can you consolidate any of these claims?

…

Please replace pronouns with their antecedents.

…

Label these propositions a-s.

…

GPT-4o produces a labeled list of substantive claims/ propositions in the meeting transcript

Produce a non-redundant list of the substantive claims made by the participants in the following transcript.

…

Can you consolidate any of these claims?

…

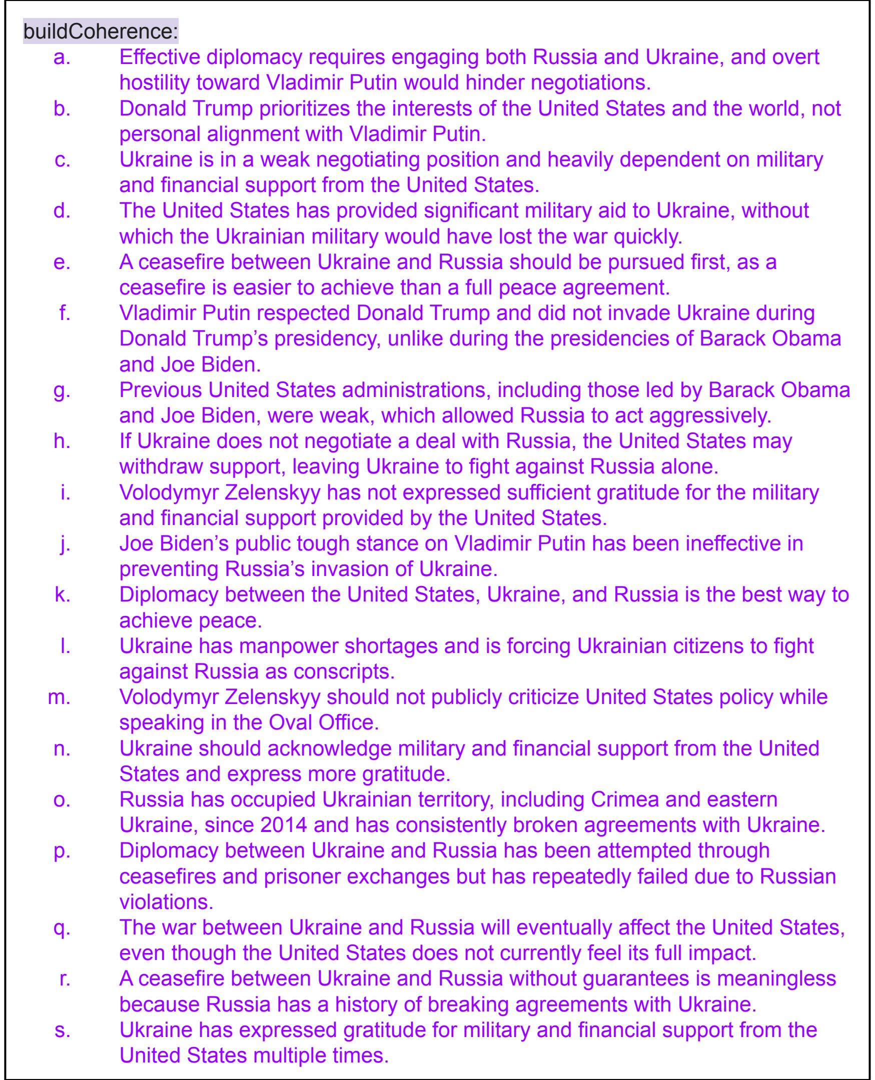
Please replace pronouns with their antecedents.

…

Label these propositions a-s.

…

GPT-4o produces a labeled list of substantive claims/ propositions in the meeting transcript



buildCoherence:

a. Effective diplomacy requires engaging both Russia and Ukraine, and overt hostility toward Vladimir Putin would hinder negotiations.
b. Donald Trump prioritizes the interests of the United States and the world, not personal alignment with Vladimir Putin.
c. Ukraine is in a weak negotiating position and heavily dependent on military and financial support from the United States.
d. The United States has provided significant military aid to Ukraine, without which the Ukrainian military would have lost the war quickly.
e. A ceasefire between Ukraine and Russia should be pursued first, as a ceasefire is easier to achieve than a full peace agreement.
f. Vladimir Putin respected Donald Trump and did not invade Ukraine during Donald Trump's presidency, unlike during the presidencies of Barack Obama and Joe Biden.
g. Previous United States administrations, including those led by Barack Obama and Joe Biden, were weak, which allowed Russia to act aggressively.
h. If Ukraine does not negotiate a deal with Russia, the United States may withdraw support, leaving Ukraine to fight against Russia alone.
i. Volodymyr Zelenskyy has not expressed sufficient gratitude for the military and financial support provided by the United States.
j. Joe Biden's public tough stance on Vladimir Putin has been ineffective in preventing Russia's invasion of Ukraine.
k. Diplomacy between the United States, Ukraine, and Russia is the best way to achieve peace.
l. Ukraine has manpower shortages and is forcing Ukrainian citizens to fight against Russia as conscripts.
m. Volodymyr Zelenskyy should not publicly criticize United States policy while speaking in the Oval Office.
n. Ukraine should acknowledge military and financial support from the United States and express more gratitude.
o. Russia has occupied Ukrainian territory, including Crimea and eastern Ukraine, since 2014 and has consistently broken agreements with Ukraine.
p. Diplomacy between Ukraine and Russia has been attempted through ceasefires and prisoner exchanges but has repeatedly failed due to Russian violations.
q. The war between Ukraine and Russia will eventually affect the United States, even though the United States does not currently feel its full impact.
r. A ceasefire between Ukraine and Russia without guarantees is meaningless because Russia has a history of breaking agreements with Ukraine.
s. Ukraine has expressed gratitude for military and financial support from the United States multiple times.

Produce a non-redundant list of the substantive claims made by the participants in the following transcript.

…

Can you consolidate any of these claims?

…

Please replace pronouns with their antecedents.

…

Label these propositions a-s.

…

GPT-4o produces a labeled list of substantive claims/ propositions in the meeting transcript

buildCoherence:

a. Effective diplomacy requires engaging both Russia and Ukraine, and overt hostility toward Vladimir Putin would hinder negotiations.
b. Donald Trump prioritizes the interests of the United States and the world, not personal alignment with Vladimir Putin.
c. Ukraine is in a weak negotiating position and heavily dependent on military and financial support from the United States.
d. The United States has provided significant military aid to Ukraine, without which the Ukrainian military would have lost the war quickly.
e. A ceasefire between Ukraine and Russia should be pursued first, as a ceasefire is easier to achieve than a full peace agreement.
f. Vladimir Putin respected Donald Trump and did not invade Ukraine during Donald Trump’s presidency, unlike during the presidencies of Barack Obama and Joe Biden.
g. Previous United States administrations, including those led by Barack Obama and Joe Biden, were weak, which allowed Russia to act aggressively.
h. If Ukraine does not negotiate a deal with Russia, the United States may withdraw support, leaving Ukraine to fight against Russia alone.
i. Volodymyr Zelenskyy has not expressed sufficient gratitude for the military and financial support provided by the United States.
j. Joe Biden’s public tough stance on Vladimir Putin has been ineffective in preventing Russia’s invasion of Ukraine.
k. Diplomacy between the United States, Ukraine, and Russia is the best way to achieve peace.
l. Ukraine has manpower shortages and is forcing Ukrainian citizens to fight against Russia as conscripts.
m. Volodymyr Zelenskyy should not publicly criticize United States policy while speaking in the Oval Office.
n. Ukraine should acknowledge military and financial support from the United States and express more gratitude.
o. Russia has occupied Ukrainian territory, including Crimea and eastern Ukraine, since 2014 and has consistently broken agreements with Ukraine.
p. Diplomacy between Ukraine and Russia has been attempted through ceasefires and prisoner exchanges but has repeatedly failed due to Russian violations.
q. The war between Ukraine and Russia will eventually affect the United States, even though the United States does not currently feel its full impact.
r. A ceasefire between Ukraine and Russia without guarantees is meaningless because Russia has a history of breaking agreements with Ukraine.
s. Ukraine has expressed gratitude for military and financial support from the United States multiple times.

Imagine that you are a perfectly objective arbitrator with impeccable judgment and integrity. In response to a prompt of the form “buildCoherence: ” followed by a list of labeled propositions or claims, please do the following:

First, determine which pairs of propositions are substantively related.

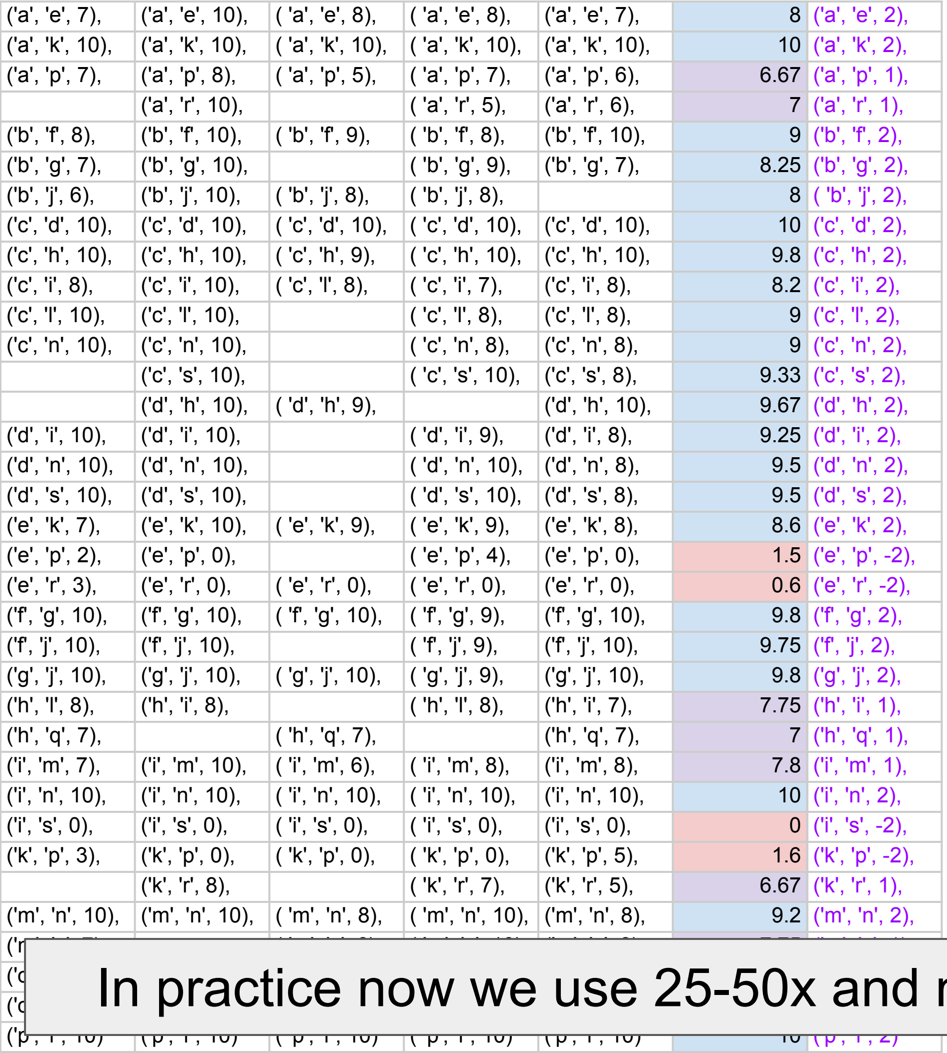
Second, for each related pair of propositions, determine their logical relationship, assuming that at least one is true, whether or not either actually is. I want you to ignore the truth, falsity or basis in fact of either claim.

Third, based on your determination just above, numerically rate the relative consistency of the two propositions. Do not pay attention to or comment on the truth or basis in fact of either proposition independent of the other. Your rating of relative consistency should be on a scale from 0 to 10, with a value of 0 for a pair of propositions that are not at all consistent and a value of 10 for a pair of propositions that are totally consistent. I cannot emphasize enough that for your rating, I want you to ignore the truth or basis in fact of either proposition, since anything that is not consistent with reality cannot be true. If you determine that propositions are unrelated despite previously determining otherwise, omit that pair. To be clear, a pair of false but consistent claims should also be rated a 10. Meanwhile, a pair of propositions of which one is true and the other is false, should be rated a 0.

Finally, construct a networkx graph where propositions are vertices and edges correspond to substantively related pairs of propositions, with weights given by the consistency ratings just above. Only return the edge list with proposition labels for vertices. i.e., return responses in this format(here 'b', 'c', 'd', and 'e' are labels):

```
[(’b’, ’c’, 0),(’b’, ’e’, 10),(’c’, ’d’, 9),(’c’, ’e’, 2)]
```

GPT-o3-mini builds graphs



('a', 'e', 7),	('a', 'e', 10),	('a', 'e', 8),	('a', 'e', 8),	('a', 'e', 7),	8	('a', 'e', 2),
('a', 'k', 10),	('a', 'k', 10),	('a', 'k', 10),	('a', 'k', 10),	('a', 'k', 10),	10	('a', 'k', 2),
('a', 'p', 7),	('a', 'p', 8),	('a', 'p', 5),	('a', 'p', 7),	('a', 'p', 6),	6.67	('a', 'p', 1),
	('a', 'r', 10),		('a', 'r', 5),	('a', 'r', 6),	7	('a', 'r', 1),
('b', 'f', 8),	('b', 'f', 10),	('b', 'f', 9),	('b', 'f', 8),	('b', 'f', 10),	9	('b', 'f', 2),
('b', 'g', 7),	('b', 'g', 10),		('b', 'g', 9),	('b', 'g', 7),	8.25	('b', 'g', 2),
('b', 'j', 6),	('b', 'j', 10),	('b', 'j', 8),	('b', 'j', 8),		8	('b', 'j', 2),
('c', 'd', 10),	('c', 'd', 10),	('c', 'd', 10),	('c', 'd', 10),	('c', 'd', 10),	10	('c', 'd', 2),
('c', 'h', 10),	('c', 'h', 10),	('c', 'h', 9),	('c', 'h', 10),	('c', 'h', 10),	9.8	('c', 'h', 2),
('c', 'i', 8),	('c', 'i', 10),	('c', 'l', 8),	('c', 'i', 7),	('c', 'i', 8),	8.2	('c', 'i', 2),
('c', 'l', 10),	('c', 'l', 10),		('c', 'l', 8),	('c', 'l', 8),	9	('c', 'l', 2),
('c', 'n', 10),	('c', 'n', 10),		('c', 'n', 8),	('c', 'n', 8),	9	('c', 'n', 2),
	('c', 's', 10),		('c', 's', 10),	('c', 's', 8),	9.33	('c', 's', 2),
	('d', 'h', 10),	('d', 'h', 9),		('d', 'h', 10),	9.67	('d', 'h', 2),
('d', 'i', 10),	('d', 'i', 10),		('d', 'i', 9),	('d', 'i', 8),	9.25	('d', 'i', 2),
('d', 'n', 10),	('d', 'n', 10),		('d', 'n', 10),	('d', 'n', 8),	9.5	('d', 'n', 2),
('d', 's', 10),	('d', 's', 10),		('d', 's', 10),	('d', 's', 8),	9.5	('d', 's', 2),
('e', 'k', 7),	('e', 'k', 10),	('e', 'k', 9),	('e', 'k', 9),	('e', 'k', 8),	8.6	('e', 'k', 2),
('e', 'p', 2),	('e', 'p', 0),		('e', 'p', 4),	('e', 'p', 0),	1.5	('e', 'p', -2),
('e', 'r', 3),	('e', 'r', 0),	('e', 'r', 0),	('e', 'r', 0),	('e', 'r', 0),	0.6	('e', 'r', -2),
('f', 'g', 10),	('f', 'g', 10),	('f', 'g', 10),	('f', 'g', 9),	('f', 'g', 10),	9.8	('f', 'g', 2),
('f', 'j', 10),	('f', 'j', 10),		('f', 'j', 9),	('f', 'j', 10),	9.75	('f', 'j', 2),
('g', 'j', 10),	('g', 'j', 10),	('g', 'j', 10),	('g', 'j', 9),	('g', 'j', 10),	9.8	('g', 'j', 2),
('h', 'l', 8),	('h', 'i', 8),		('h', 'l', 8),	('h', 'i', 7),	7.75	('h', 'i', 1),
('h', 'q', 7),		('h', 'q', 7),		('h', 'q', 7),	7	('h', 'q', 1),
('i', 'm', 7),	('i', 'm', 10),	('i', 'm', 6),	('i', 'm', 8),	('i', 'm', 8),	7.8	('i', 'm', 1),
('i', 'n', 10),	('i', 'n', 10),	('i', 'n', 10),	('i', 'n', 10),	('i', 'n', 10),	10	('i', 'n', 2),
('i', 's', 0),	('i', 's', 0),	('i', 's', 0),	('i', 's', 0),	('i', 's', 0),	0	('i', 's', -2),
('k', 'p', 3),	('k', 'p', 0),	('k', 'p', 0),	('k', 'p', 0),	('k', 'p', 5),	1.6	('k', 'p', -2),
	('k', 'r', 8),		('k', 'r', 7),	('k', 'r', 5),	6.67	('k', 'r', 1),
('m', 'n', 10),	('m', 'n', 10),	('m', 'n', 8),	('m', 'n', 10),	('m', 'n', 8),	9.2	('m', 'n', 2),

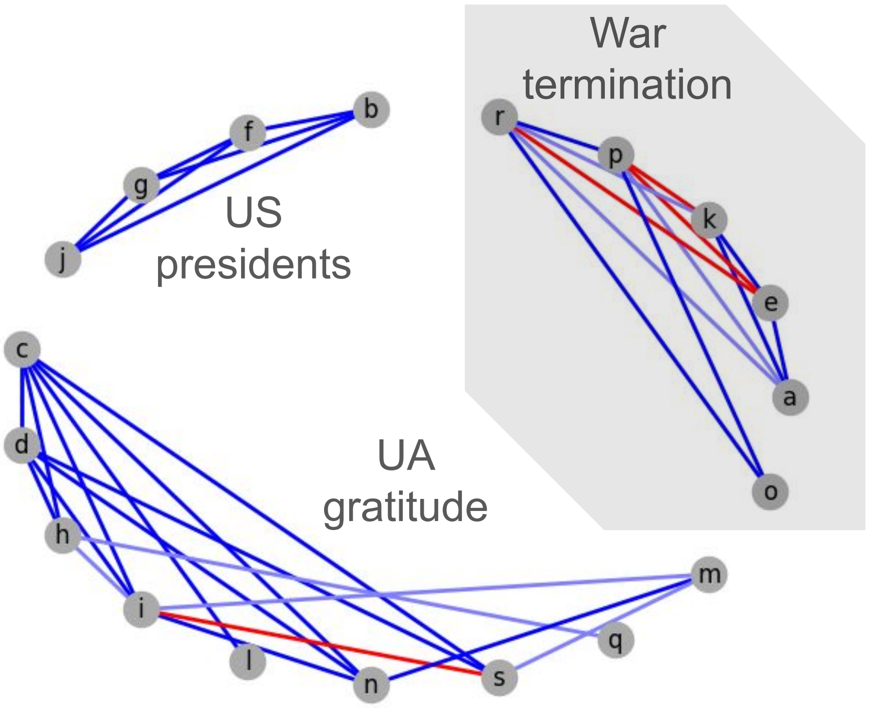
First five columns: results from running the buildCoherence graph prompt 5x; aligning graph edges across responses, and keeping only edges that occurred most of the time

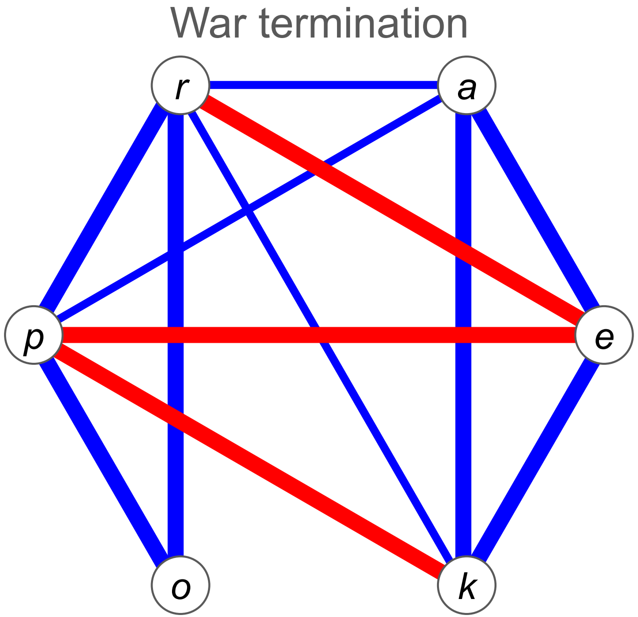
Sixth column: averaged weights (rescaled for coherence graph)

Seventh column: rescaled and quantized weights in {-2, -1, 0, 1, 2}

In practice now we use 25-50x and monitor convergence to <u>median</u>

Focus on the most salient connected component





Thick {blue, red} weights = {+2, -2};
thin {blue, red} weights = {+1, -1}

a. Effective diplomacy requires engaging both Russia and Ukraine, and overt hostility toward Vladimir Putin would hinder negotiations.

Speculative

e. A ceasefire between Ukraine and Russia should be pursued first, as a ceasefire is easier to achieve than a full peace agreement.

Speculative

k. Diplomacy between the United States, Ukraine, and Russia is the best way to achieve peace.

Speculative

o. Russia has occupied Ukrainian territory, including Crimea and eastern Ukraine, since 2014 and has consistently broken agreements with Ukraine.

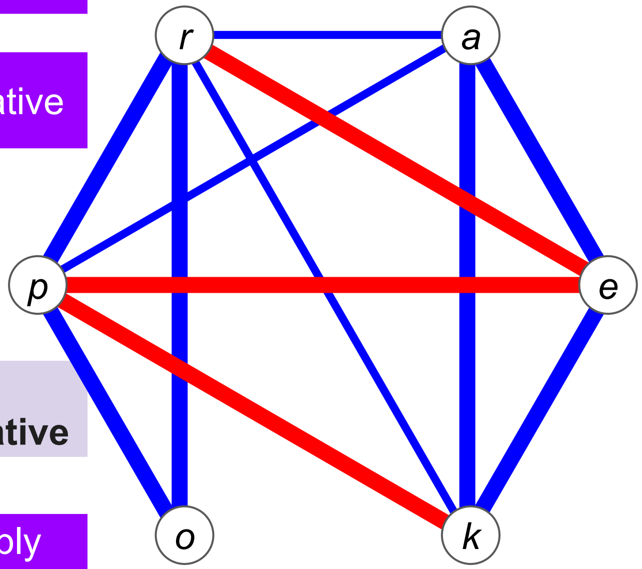
Not speculative

p. Diplomacy between Ukraine and Russia has been attempted through ceasefires and prisoner exchanges but has repeatedly failed due to Russian violations.

Arguably speculative

r. A ceasefire between Ukraine and Russia without guarantees is meaningless because Russia has a history of breaking agreements with Ukraine.

Arguably speculative



Most coherent partition

a. Effective diplomacy requires engaging both Russia and Ukraine, and overt hostility toward Vladimir Putin would hinder negotiations.

e. A ceasefire between Ukraine and Russia should be pursued first, as a ceasefire is easier to achieve than a full peace agreement.

k. Diplomacy between the United States, Ukraine, and Russia is the best way to achieve peace.

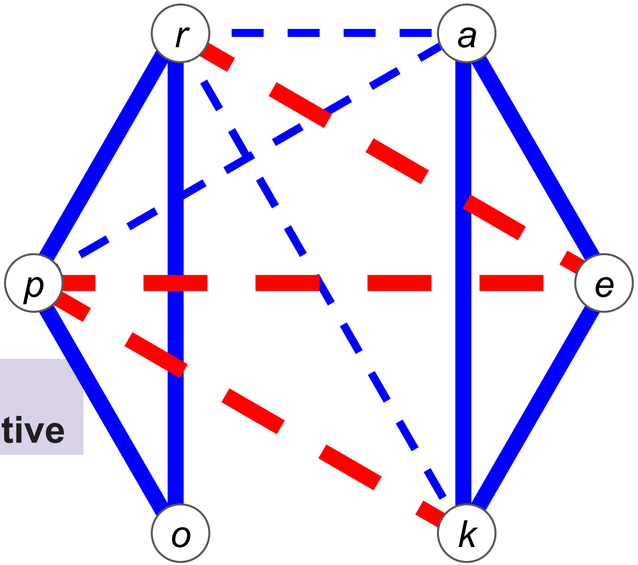
o. Russia has occupied Ukrainian territory, including Crimea and eastern Ukraine, since 2014 and has consistently broken agreements with Ukraine.

p. Diplomacy between Ukraine and Russia has been attempted through ceasefires and prisoner exchanges but has repeatedly failed due to Russian violations.

r. A ceasefire between Ukraine and Russia without guarantees is meaningless because Russia has a history of breaking agreements with Ukraine.

Not speculative

$$\text{Coherence}(\{a, e, k\}, \{o, p, r\}) = 2 + 2 + 2 - 1 - 1 - 1 = \mathbf{3}$$



Data priority weighs in favor of accepting {*o*, *p*, *r*} and rejecting {*a*, *e*, *k*}

a. Effective diplomacy requires engaging both Russia and Ukraine, and overt hostility toward Vladimir Putin would hinder negotiations.

e. A ceasefire between Ukraine and Russia should be pursued first, as a ceasefire is easier to achieve than a full peace agreement.

k. Diplomacy between the United States, Ukraine, and Russia is the best way to achieve peace.

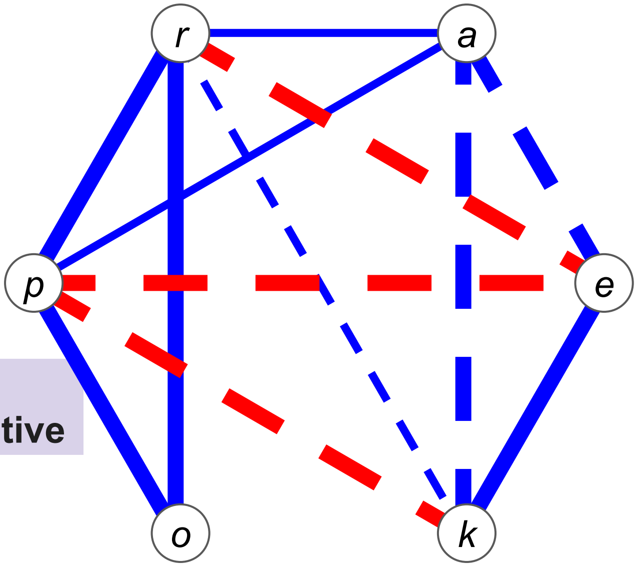
o. Russia has occupied Ukrainian territory, including Crimea and eastern Ukraine, since 2014 and has consistently broken agreements with Ukraine.

Not speculative

p. Diplomacy between Ukraine and Russia has been attempted through ceasefires and prisoner exchanges but has repeatedly failed due to Russian violations.

r. A ceasefire between Ukraine and Russia without guarantees is meaningless because Russia has a history of breaking agreements with Ukraine.

Coherence({*e*, *k*}, {*a*, *o*, *p*, *r*})
= 2 + 2 + 2 - 2 - 2 - 1 = **1**



Data priority weighs in favor of accepting {*a*, *o*, *p*, *r*} and rejecting {*e*, *k*}

a. Effective diplomacy requires engaging both Russia and Ukraine, and overt hostility toward Vladimir Putin would hinder negotiations.

e. A ceasefire between Ukraine and Russia should be pursued first, as a ceasefire is easier to achieve than a full peace agreement.

k. Diplomacy between the United States, Ukraine, and Russia is the best way to achieve peace.

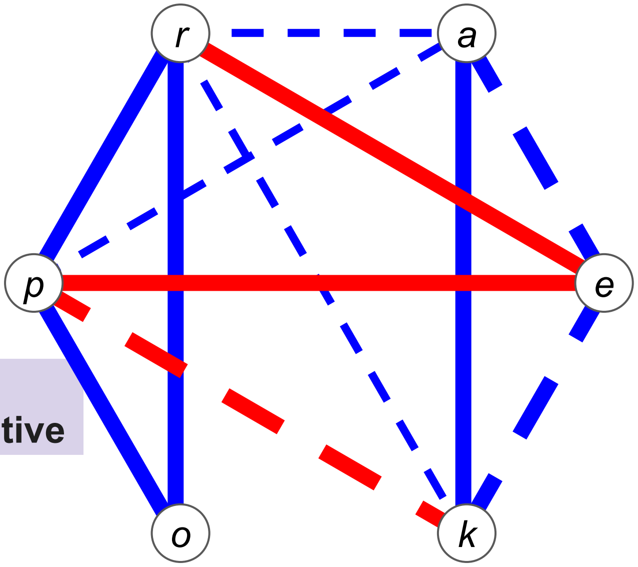
o. Russia has occupied Ukrainian territory, including Crimea and eastern Ukraine, since 2014 and has consistently broken agreements with Ukraine.

Not speculative

p. Diplomacy between Ukraine and Russia has been attempted through ceasefires and prisoner exchanges but has repeatedly failed due to Russian violations.

r. A ceasefire between Ukraine and Russia without guarantees is meaningless because Russia has a history of breaking agreements with Ukraine.

$$\text{Coherence}(\{a, k\}, \{e, o, p, r\})$$
$$= 2 - 2 - 2 - 1 - 1 - 1 = \mathbf{-5}$$



Data priority weighs in favor of accepting {*e*, *o*, *p*, *r*} and rejecting {*a*, *k*}

a. Effective diplomacy requires engaging both Russia and Ukraine, and overt hostility toward Vladimir Putin would hinder negotiations.

e. A ceasefire between Ukraine and Russia should be pursued first, as a ceasefire is easier to achieve than a full peace agreement.

k. Diplomacy between the United States, Ukraine, and Russia is the best way to achieve peace.

o. Russia has occupied Ukrainian territory, including Crimea and eastern Ukraine, since 2014 and has consistently broken agreements with Ukraine.

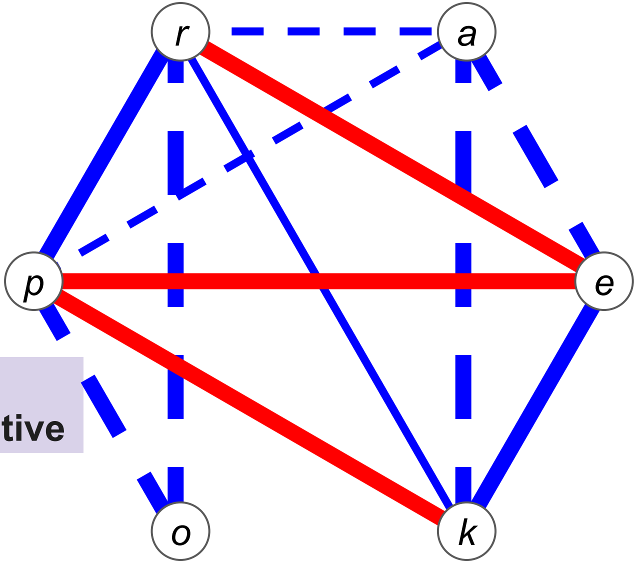
p. Diplomacy between Ukraine and Russia has been attempted through ceasefires and prisoner exchanges but has repeatedly failed due to Russian violations.

r. A ceasefire between Ukraine and Russia without guarantees is meaningless because Russia has a history of breaking agreements with Ukraine.

Least coherent partition

Coherence({*a*, *o*}, {*e*, *k*, *p*, *r*})
= -2 - 2 - 2 - 2 - 1 - 1 = **-10**

Not speculative



Data priority weighs in favor of accepting {*a*, *o*} and rejecting {*e*, *k*, *p*, *r*}

Accepted propositions

Coherence (quantized)

Coherence (real)

Accept	CohQ	CohR
_	0	0
r	-4	-1.8533
p	-1	-0.9533
pr	-1	-0.8067
o	-4	-2
or	-4	-1.8533
op	-1	-0.9533
opr	**3**	**1.1933**
k	-3	-1.3733
kr	-5	-2.56
kp	-8	-3.6867
kpr	-6	-2.8733
ko	-7	-3.3733
kor	-5	-2.56
kop	-8	-3.6867
kopr	-2	-0.8733

Accept	CohQ	CohR
e	0	0.26
er	-8	-3.3533
ep	-5	-2.0933
epr	-9	-3.7067
eo	-4	-1.74
eor	-8	-3.3533
eop	-5	-2.0933
eopr	-5	-1.7067
ek	1	0.3267
ekr	-5	-2.62
ekp	-8	-3.3867
ekpr	-10	-4.3333
eko	-3	-1.6733
ekor	-5	-2.62
ekop	-8	-3.3867
ekopr	-6	-2.3333

Accept	CohQ	CohR
a	-6	-2.3333
ar	-8	-3.3867
ap	-5	-2.62
apr	-3	-1.6733
ao	-10	-4.3333
aor	-8	-3.3867
aop	-5	-2.62
aopr	1	0.3267
ak	-5	-1.7067
akr	-5	-2.0933
akp	-8	-3.3533
akpr	-4	-1.74
ako	-9	-3.7067
akor	-5	-2.0933
akop	-8	-3.3533
akopr	0	0.26

Accept	CohQ	CohR
ae	-2	-0.8733
aer	-8	-3.6867
aep	-5	-2.56
aepr	-7	-3.3733
aeo	-6	-2.8733
aeor	-8	-3.6867
aeop	-5	-2.56
aeopr	-3	-1.3733
aek	3	1.1933
aekr	-1	-0.9533
aekp	-4	-1.8533
aekpr	-4	-2
aeko	-1	-0.8067
aekor	-1	-0.9533
aekop	-4	-1.8533
aekopr	0	0

Symmetry between acceptance/rejection can be broken by **prioritizing** according to provenance (e.g., **direct observation** vs speculation)

Accepted propositions

Coherence (quantized)

Coherence (real)

Accept	CohQ	CohR
_	0	0
r	-4	-1.8533
p	-1	-0.9533
pr	-1	-0.8067
o	-4	-2
or	-4	-1.8533
op	-1	-0.9533
opr	**3**	**1.1933**
k	-3	-1.3733
kr	-5	-2.56
kp		
kpr		
ko		
kor		
kop		
kopr	-2	-0.8733

Accept	CohQ	CohR
e	0	0.26
er	-8	-3.3533
ep	-5	-2.0933
epr	-9	-3.7067
eo	-4	-1.74
eor	-8	-3.3533
eop	-5	-2.0933
eopr	-5	-1.7067
ek	1	0.3267
ekr	-5	-2.62
ekopr	-6	-2.3333

Accept	CohQ	CohR
a	-6	-2.3333
ar	-8	-3.3867
ap	-5	-2.62
apr	-3	-1.6733
ao	-10	-4.3333
aor	-8	-3.3867
aop	-5	-2.62
aopr	1	0.3267
ak	-5	-1.7067
akr	-5	-2.0933
akopr	0	0.26

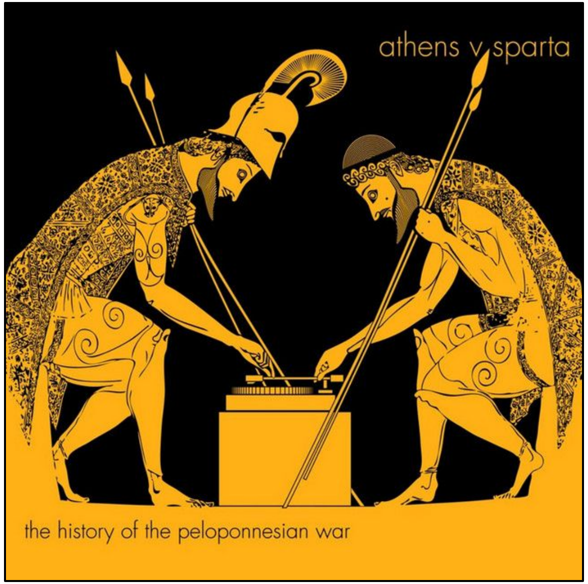
Accept	CohQ	CohR
ae	-2	-0.8733
aer	-8	-3.6867
aep	-5	-2.56
aepr	-7	-3.3733
aeo	-6	-2.8733
aeor	-8	-3.6867
aeop	-5	-2.56
aeopr	-3	-1.3733
aek	3	1.1933
aekr	-1	-0.9533
		.8533
		.8067
		.9533
		.8533
aekopr	0	0

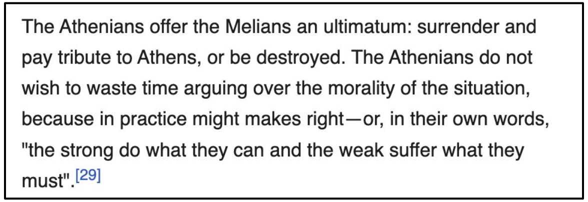
In this example, coherence-driven inference leads to accepting propositions {*o*, *p*, *r*} and rejecting propositions {*a*, *e*, *k*}. Recall that *r* is:

A ceasefire between Ukraine and Russia without guarantees is meaningless because Russia has a history of breaking agreements with Ukraine.

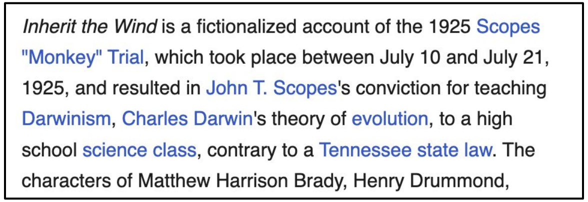
Symmetry between acceptance/rejection can be broken by **prioritizing** according to provenance (e.g., **direct observation** vs speculation)

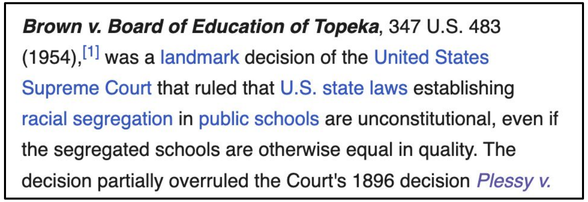
CDI informs interpretation of argument narratives




The Athenians offer the Melians an ultimatum: surrender and pay tribute to Athens, or be destroyed. The Athenians do not wish to waste time arguing over the morality of the situation, because in practice might makes right—or, in their own words, "the strong do what they can and the weak suffer what they must".[29]

"INHERIT THE WIND"
IT'S ALL ABOUT THE 'MONKEY TRIAL' THAT ROCKED AMERICA!
GOD BLESS MATT BRADY
DON'T MONKEY WITH US!
KEEP SATAN OUT OF HILLSBORO
I'M NOT ONE
CHARLES DARWIN'S Theory of the Evolution and the Descent of Man


Inherit the Wind is a fictionalized account of the 1925 Scopes "Monkey" Trial, which took place between July 10 and July 21, 1925, and resulted in John T. Scopes's conviction for teaching Darwinism, Charles Darwin's theory of evolution, to a high school science class, contrary to a Tennessee state law. The characters of Matthew Harrison Brady, Henry Drummond,


Brown v. Board of Education of Topeka, 347 U.S. 483 (1954),[1] was a landmark decision of the United States Supreme Court that ruled that U.S. state laws establishing racial segregation in public schools are unconstitutional, even if the segregated schools are otherwise equal in quality. The decision partially overruled the Court's 1896 decision *Plessy v.*

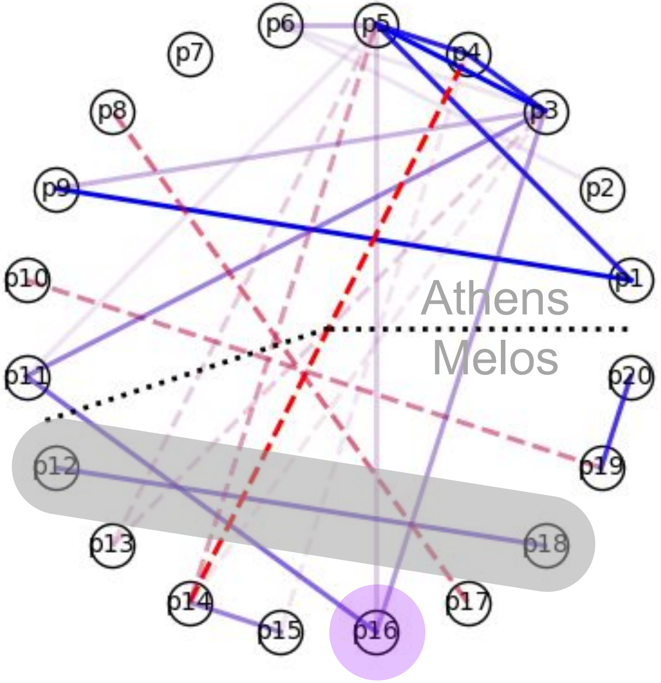
The Melian Dialogue

Using transparency vs thickness

Propositions extracted with GPT-4o

# Athenians
- p1: Right is determined by power; the strong dominate, and the weak endure.
- p2: Internal rebellion is more concerning than external threats.
- p3: Preserving the Athenian empire and Melos's submission is beneficial.
- p4: Neutrality from Melos is unacceptable as it would show Athenian weakness.
- p5: Subjugating Melos would strengthen Athenian security and the Athenian empire.
- p6: Islanders like Melos are more threatening than continentals.
- p7: The conflict with Melos is about survival, not honor.
- p8: Hope is unreliable and can lead to ruin.
- p9: Athenian actions align with the natural order of ruling where possible.
- p10: Lacedaemonian support for Melos is doubtful due to the Lacedaemonians' cautious nature.
- p11: Submitting to Athens is a secure and honorable choice for Melos.

# Melians
- p12: Fairness and justice are crucial in conflict.
- p13: Destroying Melos would set a dangerous precedent for Athens.
- p14: Neutrality is preferable to choosing sides.
- p15: Attacking neutrals could create more enemies for Athens.
- p16: Trying all options before submitting is important.
- p17: War's unpredictability offers hope.
- p18: Divine support exists for the just cause.
- p19: Lacedaemonian aid is expected due to kinship and honor.
- p20: Proximity to Peloponnese makes Lacedaemonian help more feasible.



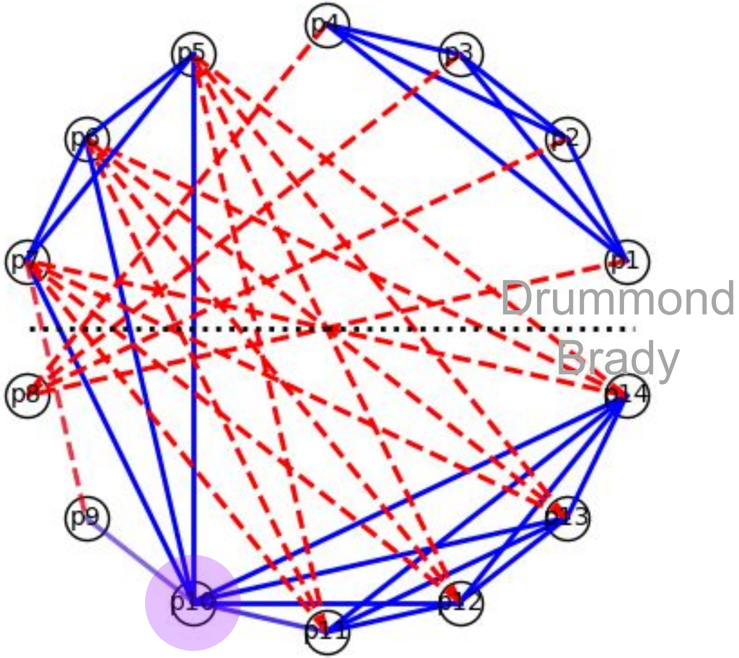
o1-mini separates key Athenian and Melian views; CDI suggests focusing on p16 to delay conflict

Inherit the Wind

Act II, Scene 2
propositions extracted with GPT-4o

# Drummond
- p1: Expert testimony on evolution should be included in the trial of Bertram Cates.
- p2: Understanding evolution is essential for the jury's judgment in the trial.
- p3: The exclusion of scientific experts from the trial is wrong.
- p4: Bertram Cates' teachings on evolution are not criminal.
- p5: Matthew Harrison Brady's literal interpretation of the Bible is flawed.
- p6: Reason and scientific progress are more important than strict adherence to religious texts.
- p7: The Bible is not the sole source of truth and should not be the only reference in matters of education and law.

# Brady
- p8: Testimony on evolution should not be included in the trial of Bertram Cates.
- p9: The law excludes evolution from being taught in schools.
- p10: Brady is an authority on the Bible but has not read Charles Darwin's "Origin of Species."
- p11: The Bible's literal truth is valid, and God can change natural law as described in the Bible.
- p12: Scientific evidence that contradicts the Bible's teachings is invalid.
- p13: Faith in the Bible is more important than scientific reasoning and evidence.
- p14: God guides Brady's actions, and opposing Brady is akin to opposing God.



CDI separates arguments and identifies the pivot for winning

Inherit the Wind

DRUMMOND

(Scowling)
In other words, the court rules out any expert testimony on Charles Darwin's *Origin of Species* or *Descent of Man*?

JUDGE

The court so rules.
*(*DRUMMOND *is flabbergasted. His case is cooked and he knows it. He looks around helplessly.)*

DRUMMOND

(There's the glint of an idea in his eye.)
Would the court admit expert testimony regarding a book known as the Holy Bible?

JUDGE

(Hesitates, turns to BRADY*)*
Any objection, Colonel Brady?

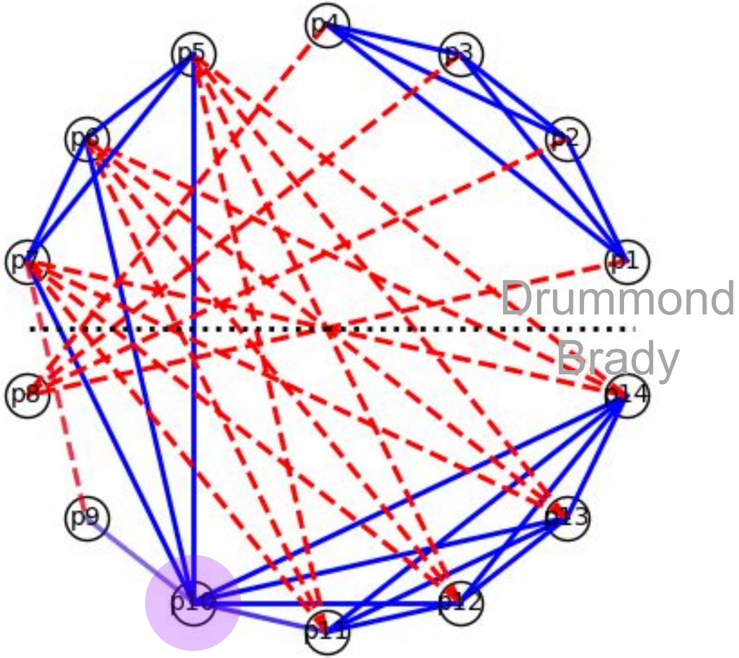
BRADY

If the counsel can advance the case of the defendant through the use of the Holy Scriptures, the prosecution will take no exception!

- p10: Brady is an authority on the Bible but has not read Charles Darwin's "Origin of Species."

DRUMMOND

Good! *(With relish)* I call to the stand one of the world's foremost experts on the Bible and its teachings – Matthew Harrison Brady!
(There is an uproar in the courtroom. The JUDGE *raps for order.)*



CDI separates arguments and identifies the pivot for winning

Brown v. Board of Education

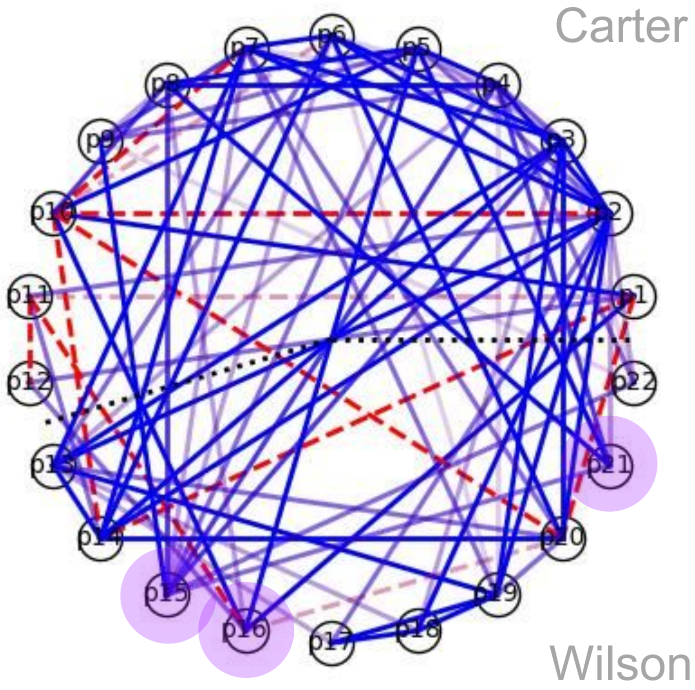
# Robert L. Carter:
- p1: The Kansas statute allowing segregated schools violates the Fourteenth Amendment.
- p4: The appellants must attend segregated elementary schools because of their race.
- p5: Segregation denies equal educational opportunities and harms the development of Negro children.
- p8: Segregation makes educational opportunities for Negro children inferior to those for white children.
- p9: There is no difference in physical facilities between schools for Negro children and schools for white children.
- p10: Segregation itself is unconstitutional under the Fourteenth Amendment.
- p12: Plessy v. Ferguson does not apply to education.

# Paul Wilson:
- p15: There is no substantial inequality in educational facilities between schools for Negro children and schools for white children.
- p16: The "separate but equal" doctrine may not be valid.
- p21: Segregation is detrimental to Negro children but legally insignificant.

***** MAXIMUM CUT IS HERE *****

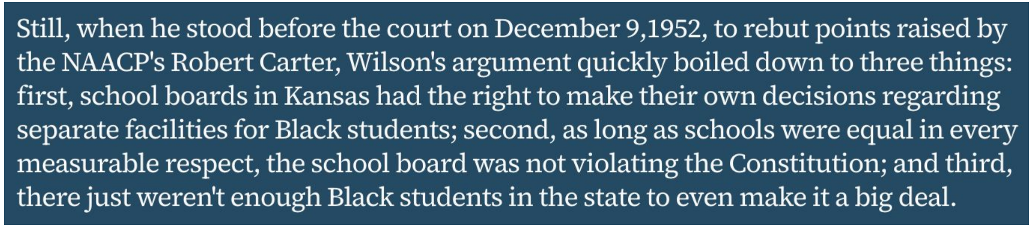
# Robert L. Carter:
- p2: The Kansas statute permits segregation in elementary schools, and in high schools only in Kansas City.
- p3: Kansas law prohibits racial distinctions in public schools without the Kansas statute.
- p6: Segregation complies with Kansas state law.
- p7: Kansas can impose racial distinctions if educational facilities are equal.
- p11: Plessy v. Ferguson and Gong Lum v. Rice require upholding the Kansas statute.

# Paul Wilson:
- p13: The Kansas statute allows, but does not require, segregation in cities with populations over 15,000.
- p14: The statute is constitutional, supported by state and federal court decisions.
- p17: Kansas has a small Negro population, less than four percent, mostly in urban areas.
- p18: Segregated schools exist in only nine cities in Kansas.
- p19: The Kansas statute was designed to let communities adjust to local conditions.
- p20: The Kansas Supreme Court has upheld the statute's constitutionality.
- p22: The appellants have not shown specific harm from attending segregated schools.



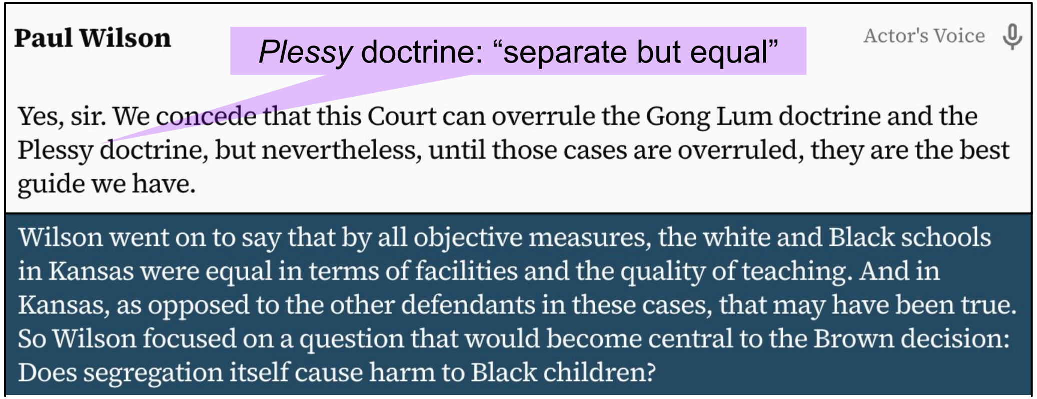
CDI finds a resolution that incorporates losing counsel's own arguments (as did the Court)

Brown v. Board of Education



Still, when he stood before the court on December 9,1952, to rebut points raised by the NAACP's Robert Carter, Wilson's argument quickly boiled down to three things: first, school boards in Kansas had the right to make their own decisions regarding separate facilities for Black students; second, as long as schools were equal in every measurable respect, the school board was not violating the Constitution; and third, there just weren't enough Black students in the state to even make it a big deal.

- p15: There is no substantial inequality in educational facilities between schools for Negro children and schools for white children.
- p16: The "separate but equal" doctrine may not be valid.
- p21: Segregation is detrimental to Negro children but legally insignificant.



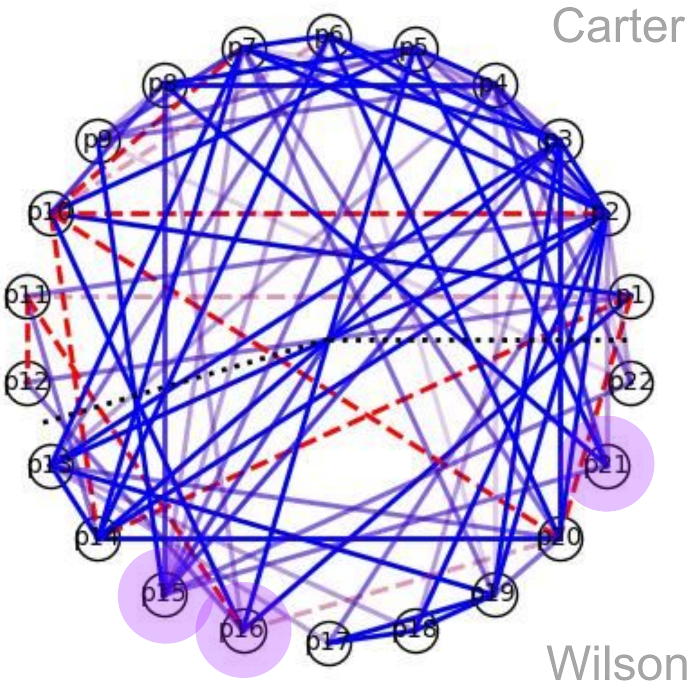
Paul Wilson

Plessy doctrine: “separate but equal”

Yes, sir. We concede that this Court can overrule the Gong Lum doctrine and the Plessy doctrine, but nevertheless, until those cases are overruled, they are the best guide we have.

Wilson went on to say that by all objective measures, the white and Black schools in Kansas were equal in terms of facilities and the quality of teaching. And in Kansas, as opposed to the other defendants in these cases, that may have been true. So Wilson focused on a question that would become central to the Brown decision: Does segregation itself cause harm to Black children?

brown.oyez.org/modules/arguments-for-segregation/



CDI finds a resolution that incorporates losing counsel’s own arguments (as did the Court)

Conclusion

CDI enables autonomy that is:

- Explainable
    - Cuts are readily interpretable; approximately optimal cuts can be computed and compared
- Ethical
    - Stipulating ethical guidelines as propositions that must be accepted gives “laws of inference”
- Reproducible and stable
    - Medians of LLM-generated coherence graphs converge usefully (or give warning indicators)
- Versatile
    - Any sufficiently capable multimodal model can be employed
- Capable of handling abstraction
    - CDI with LLMs operates directly over human-oriented representations
- Capable of handling ambiguity
    - Inference amounts to resolving ambiguity
- Based on a flexible cognitive model grounded in decades of research
    - There are deep connections with practical psychology (law, philosophy of science, etc.)
- Generalizable by construction
    - Mathematical considerations of sheaf theory led to an independent, general formulation

Thanks!

https://arxiv.org/abs/2509.18523

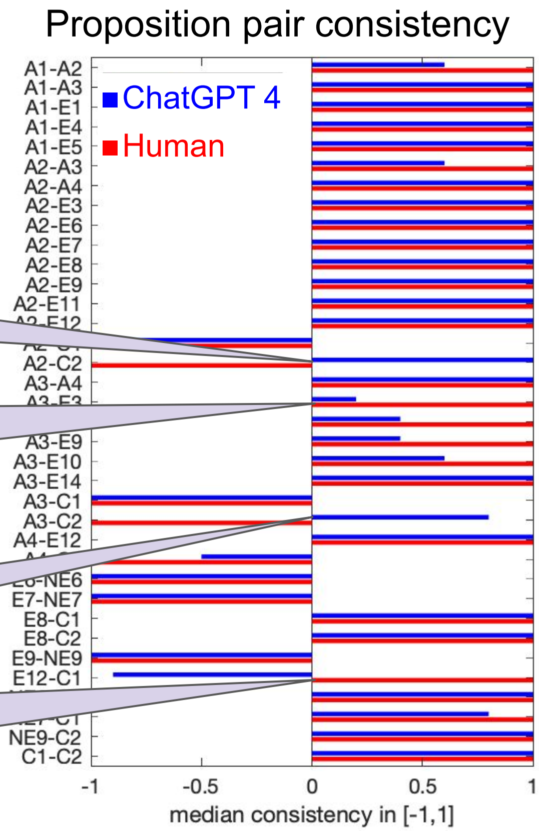
Other recent papers on arXiv, including on CDI

Robinson, Ludmilla Huntsman, & SH: DOI:10.31235/osf.io/4tjz5_v1

Backup

LLMs can outperform humans in a key task

- Coherence graph weights gauge consistency
- **ChatGPT weights and explanations in case study below manifestly improve on human**
- Few-shot learning and background info in LLM context window give statistical reproducibility
    - Median is robust and sensible



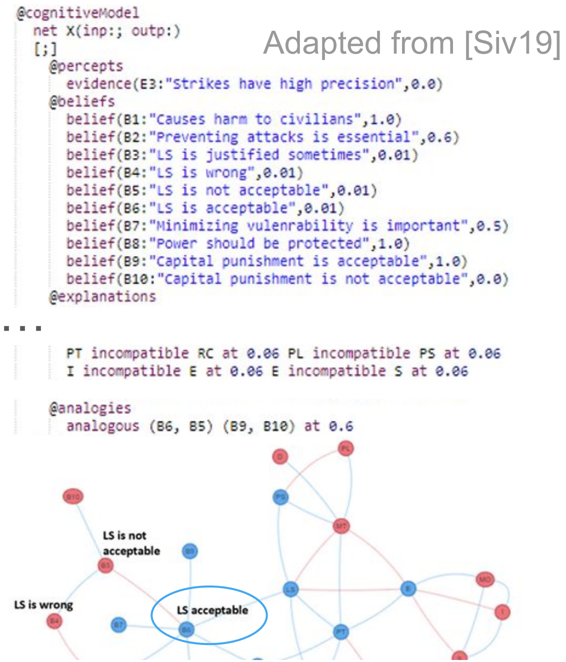
Human weight manifestly flawed

Good reason for ChatGPT weight that human apparently missed

Human weight manifestly flawed

Good reason for ChatGPT weight that human apparently missed

Coherence-driven agents can act ethically

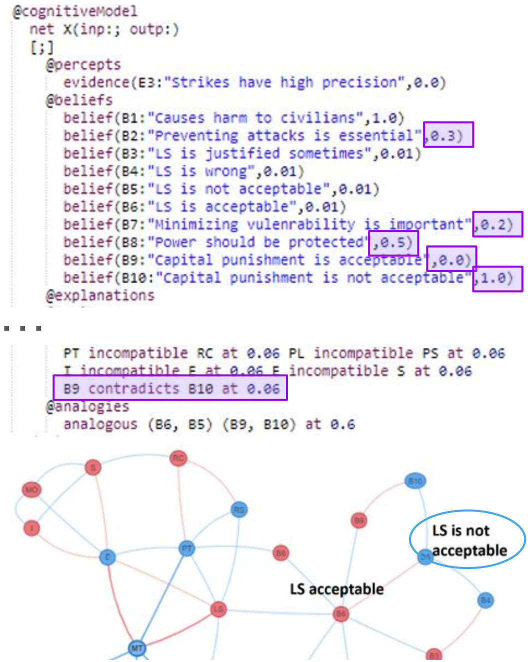


Adapted from [Siv19]

```
@cognitiveModel
  net X(inp:; outp:)
  [;]
    @percepts
      evidence(E3:"Strikes have high precision",0.0)
    @beliefs
      belief(B1:"Causes harm to civilians",1.0)
      belief(B2:"Preventing attacks is essential",0.6)
      belief(B3:"LS is justified sometimes",0.01)
      belief(B4:"LS is wrong",0.01)
      belief(B5:"LS is not acceptable",0.01)
      belief(B6:"LS is acceptable",0.01)
      belief(B7:"Minimizing vulenrability is important",0.5)
      belief(B8:"Power should be protected",1.0)
      belief(B9:"Capital punishment is acceptable",1.0)
      belief(B10:"Capital punishment is not acceptable",0.0)
    @explanations
...
      PT incompatible RC at 0.06 PL incompatible PS at 0.06
      I incompatible E at 0.06 E incompatible S at 0.06

    @analogies
      analogous (B6, B5) (B9, B10) at 0.6
```

Encoding ends-based ethics → lethal strike is acceptable



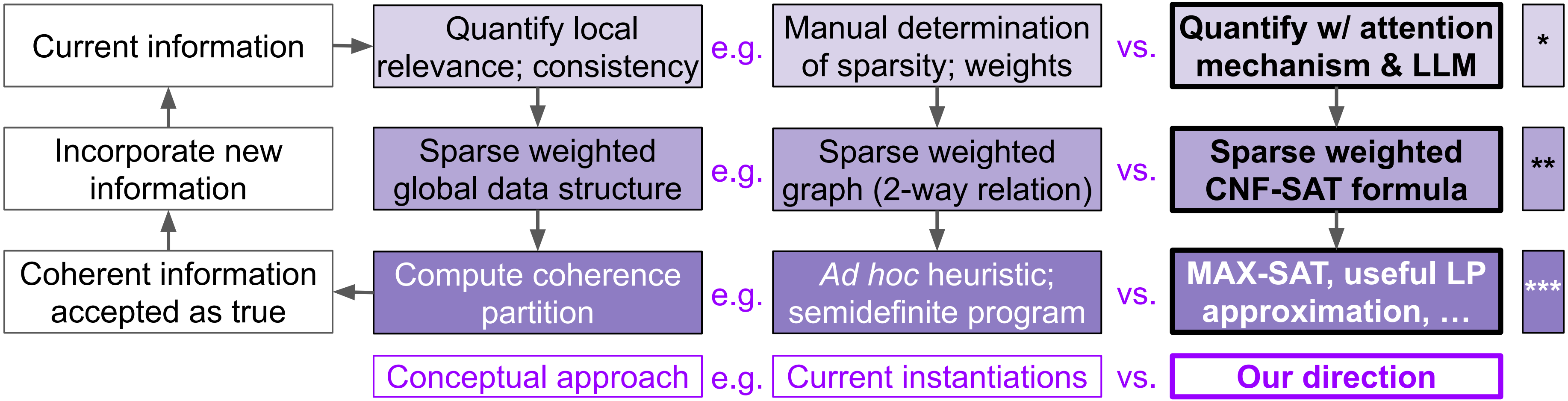
```
@cognitiveModel
  net X(inp:; outp:)
  [;]
    @percepts
      evidence(E3:"Strikes have high precision",0.0)
    @beliefs
      belief(B1:"Causes harm to civilians",1.0)
      belief(B2:"Preventing attacks is essential",0.3)
      belief(B3:"LS is justified sometimes",0.01)
      belief(B4:"LS is wrong",0.01)
      belief(B5:"LS is not acceptable",0.01)
      belief(B6:"LS is acceptable",0.01)
      belief(B7:"Minimizing vulenrability is important",0.2)
      belief(B8:"Power should be protected",0.5)
      belief(B9:"Capital punishment is acceptable",0.0)
      belief(B10:"Capital punishment is not acceptable",1.0)
    @explanations
...
      PT incompatible RC at 0.06 PL incompatible PS at 0.06
      I incompatible E at 0.06 E incompatible S at 0.06
      B9 contradicts B10 at 0.06
    @analogies
      analogous (B6, B5) (B9, B10) at 0.6
```

Encoding rules-based ethics → lethal strike is unacceptable

- Coherence similar to *reflective equilibrium* “decision procedure”
- Explainable actions
  - In vetted bounds: “Laws of Robotics”
- Stability analysis feasible via series of LP relaxations
  - Potential program metric
- No need for procedural rules

Technical hypotheses:

Applying new infrastructure*, data/structures**, and algorithms*** to coherence will enable autonomous, explainable, & ethical decision making



Current information

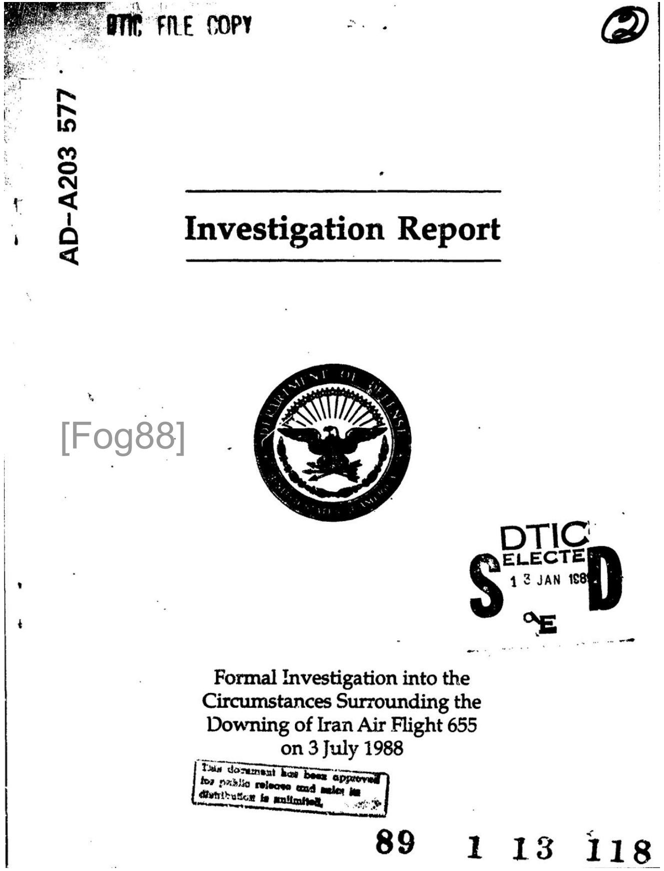
Incorporate new information

Coherent information accepted as true

Conceptual approach		Current instantiations		**Our direction**	
Quantify local relevance; consistency	e.g.	Manual determination of sparsity; weights	vs.	**Quantify w/ attention mechanism & LLM**	*
Sparse weighted global data structure	e.g.	Sparse weighted graph (2-way relation)	vs.	**Sparse weighted CNF-SAT formula**	**
Compute coherence partition	e.g.	*Ad hoc* heuristic; semidefinite program	vs.	**MAX-SAT, useful LP approximation, …**	***

Advances in infrastructure, data/structures, and algorithms enabled the renaissance of neural networks…the same will be true for coherence

Basis of confidence: case study



DTIC FILE COPY

②

AD-A203 577

Investigation Report

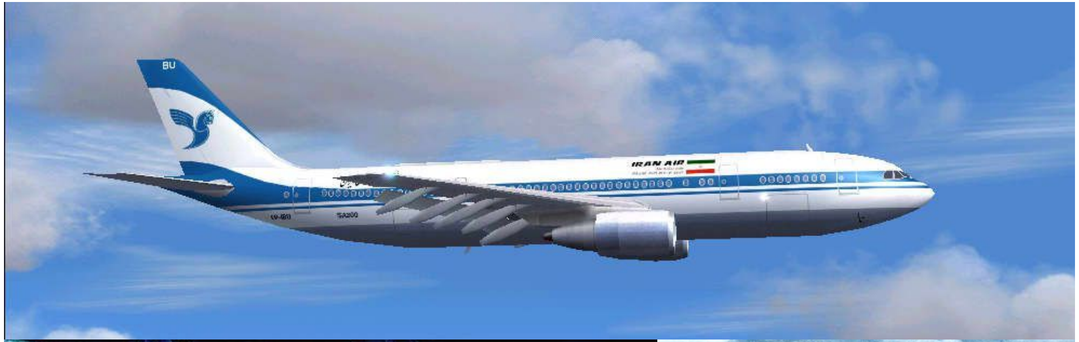
DTIC
ELECTE
13 JAN 1989
S D
E

Formal Investigation into the
Circumstances Surrounding the
Downing of Iran Air Flight 655
on 3 July 1988

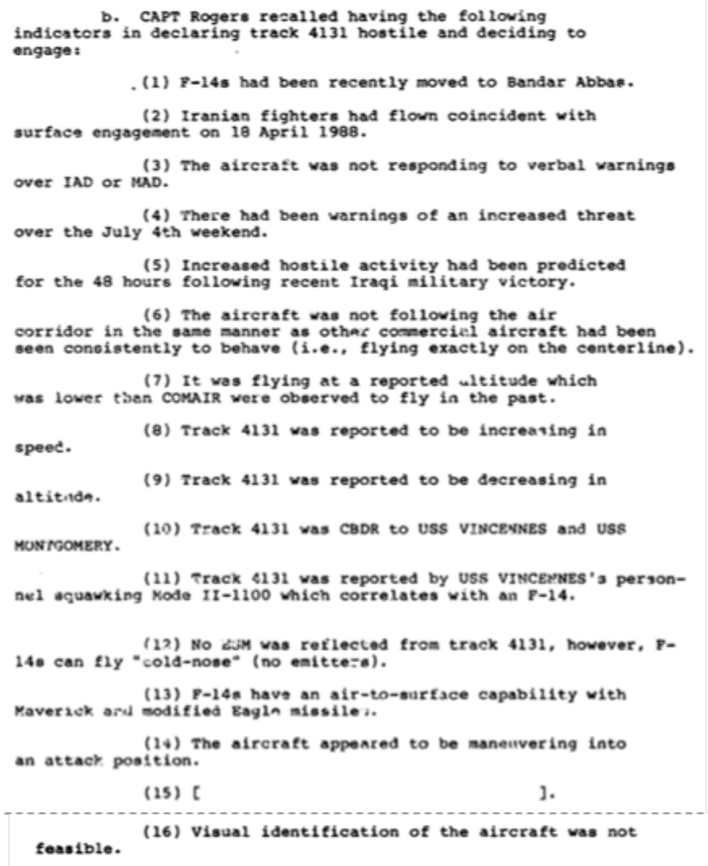
This document has been approved
for public release and sale; its
distribution is unlimited.

89 1 13 118

[Fog88]



We re-re-examine the downing of Iran Air 655



b. CAPT Rogers recalled having the following indicators in declaring track 4131 hostile and deciding to engage:

(1) F-14s had been recently moved to Bandar Abbas.

(2) Iranian fighters had flown coincident with surface engagement on 18 April 1988.

(3) The aircraft was not responding to verbal warnings over IAD or MAD.

(4) There had been warnings of an increased threat over the July 4th weekend.

(5) Increased hostile activity had been predicted for the 48 hours following recent Iraqi military victory.

(6) The aircraft was not following the air corridor in the same manner as other commercial aircraft had been seen consistently to behave (i.e., flying exactly on the centerline).

(7) It was flying at a reported altitude which was lower than COMAIR were observed to fly in the past.

(8) Track 4131 was reported to be increasing in speed.

(9) Track 4131 was reported to be decreasing in altitude.

(10) Track 4131 was CBDR to USS VINCENNES and USS MONTGOMERY.

(11) Track 4131 was reported by USS VINCENNES's personnel squawking Mode II-1100 which correlates with an F-14.

(12) No ESM was reflected from track 4131, however, F-14s can fly "cold-nose" (no emitters).

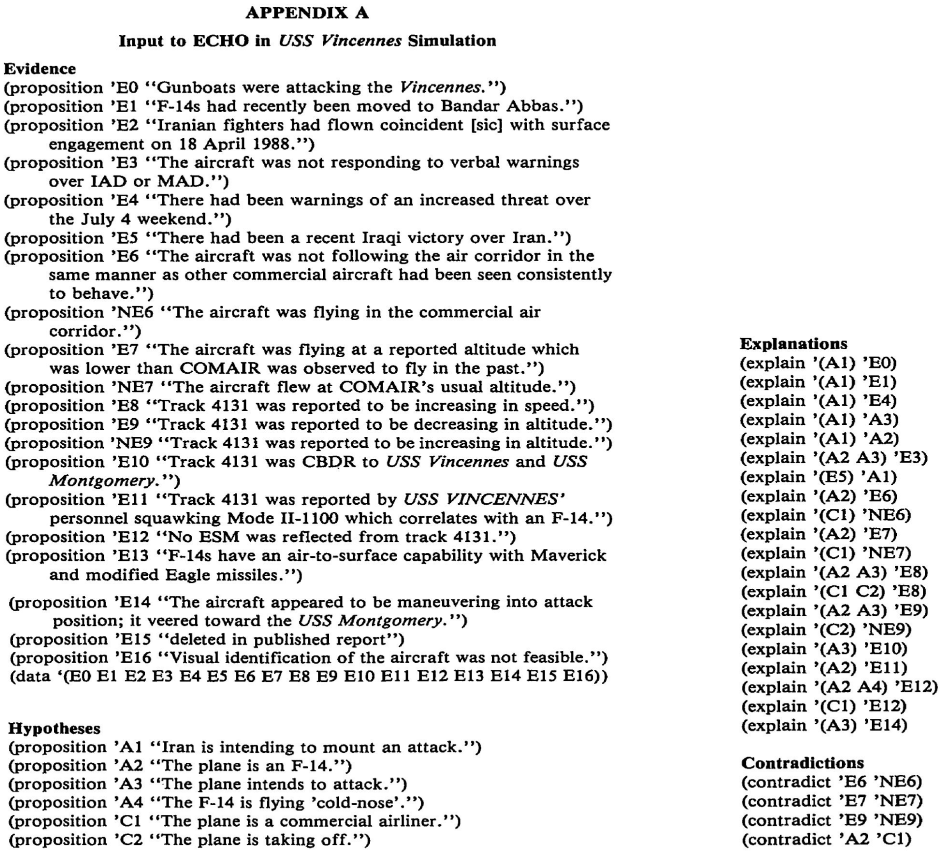
(13) F-14s have an air-to-surface capability with Maverick and modified Eagle missiles.

(14) The aircraft appeared to be maneuvering into an attack position.

(15) [ ].

(16) Visual identification of the aircraft was not feasible.

We re-re-examine the downing of Iran Air 655



APPENDIX A

Input to ECHO in *USS Vincennes* Simulation

Evidence

(proposition 'E0 "Gunboats were attacking the *Vincennes*.")
(proposition 'E1 "F-14s had recently been moved to Bandar Abbas.")
(proposition 'E2 "Iranian fighters had flown coincident [sic] with surface engagement on 18 April 1988.")
(proposition 'E3 "The aircraft was not responding to verbal warnings over IAD or MAD.")
(proposition 'E4 "There had been warnings of an increased threat over the July 4 weekend.")
(proposition 'E5 "There had been a recent Iraqi victory over Iran.")
(proposition 'E6 "The aircraft was not following the air corridor in the same manner as other commercial aircraft had been seen consistently to behave.")
(proposition 'NE6 "The aircraft was flying in the commercial air corridor.")
(proposition 'E7 "The aircraft was flying at a reported altitude which was lower than COMAIR was observed to fly in the past.")
(proposition 'NE7 "The aircraft flew at COMAIR's usual altitude.")
(proposition 'E8 "Track 4131 was reported to be increasing in speed.")
(proposition 'E9 "Track 4131 was reported to be decreasing in altitude.")
(proposition 'NE9 "Track 4131 was reported to be increasing in altitude.")
(proposition 'E10 "Track 4131 was CBDR to *USS Vincennes* and *USS Montgomery*.")
(proposition 'E11 "Track 4131 was reported by *USS VINCENNES'* personnel squawking Mode II-1100 which correlates with an F-14.")
(proposition 'E12 "No ESM was reflected from track 4131.")
(proposition 'E13 "F-14s have an air-to-surface capability with Maverick and modified Eagle missiles.")
(proposition 'E14 "The aircraft appeared to be maneuvering into attack position; it veered toward the *USS Montgomery*.")
(proposition 'E15 "deleted in published report")
(proposition 'E16 "Visual identification of the aircraft was not feasible.")
(data '(E0 E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12 E13 E14 E15 E16))

Hypotheses

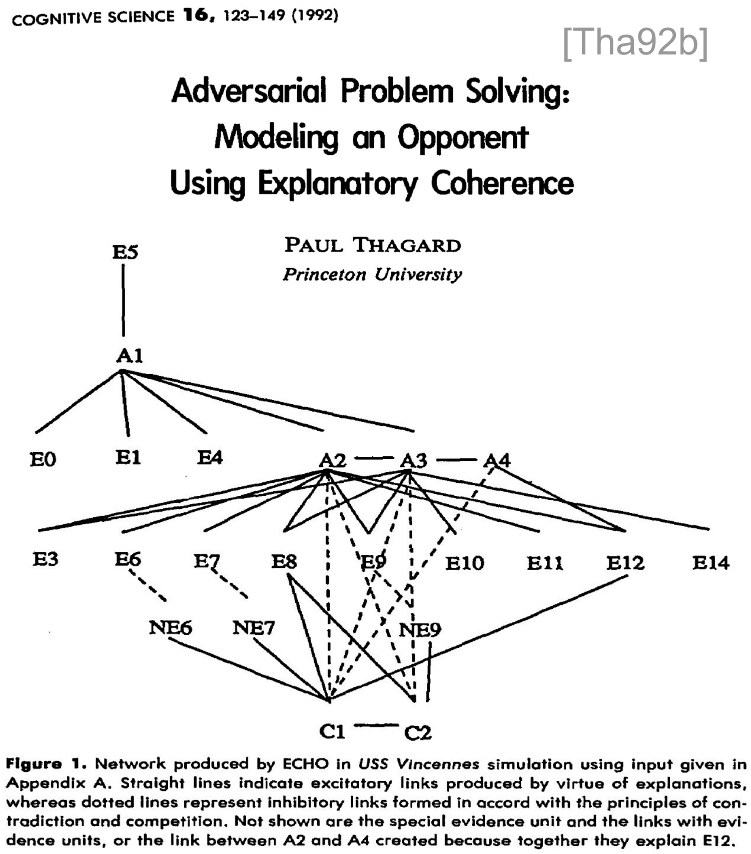
(proposition 'A1 "Iran is intending to mount an attack.")
(proposition 'A2 "The plane is an F-14.")
(proposition 'A3 "The plane intends to attack.")
(proposition 'A4 "The F-14 is flying 'cold-nose'.")
(proposition 'C1 "The plane is a commercial airliner.")
(proposition 'C2 "The plane is taking off.")

Explanations

(explain '(A1) 'E0)
(explain '(A1) 'E1)
(explain '(A1) 'E4)
(explain '(A1) 'A3)
(explain '(A1) 'A2)
(explain '(A2 A3) 'E3)
(explain '(E5) 'A1)
(explain '(A2) 'E6)
(explain '(C1) 'NE6)
(explain '(A2) 'E7)
(explain '(C1) 'NE7)
(explain '(A2 A3) 'E8)
(explain '(C1 C2) 'E8)
(explain '(A2 A3) 'E9)
(explain '(C2) 'NE9)
(explain '(A3) 'E10)
(explain '(A2) 'E11)
(explain '(A2 A4) 'E12)
(explain '(C1) 'E12)
(explain '(A3) 'E14)

Contradictions

(contradict 'E6 'NE6)
(contradict 'E7 'NE7)
(contradict 'E9 'NE9)
(contradict 'A2 'C1)



COGNITIVE SCIENCE **16,** 123–149 (1992)

[Tha92b]

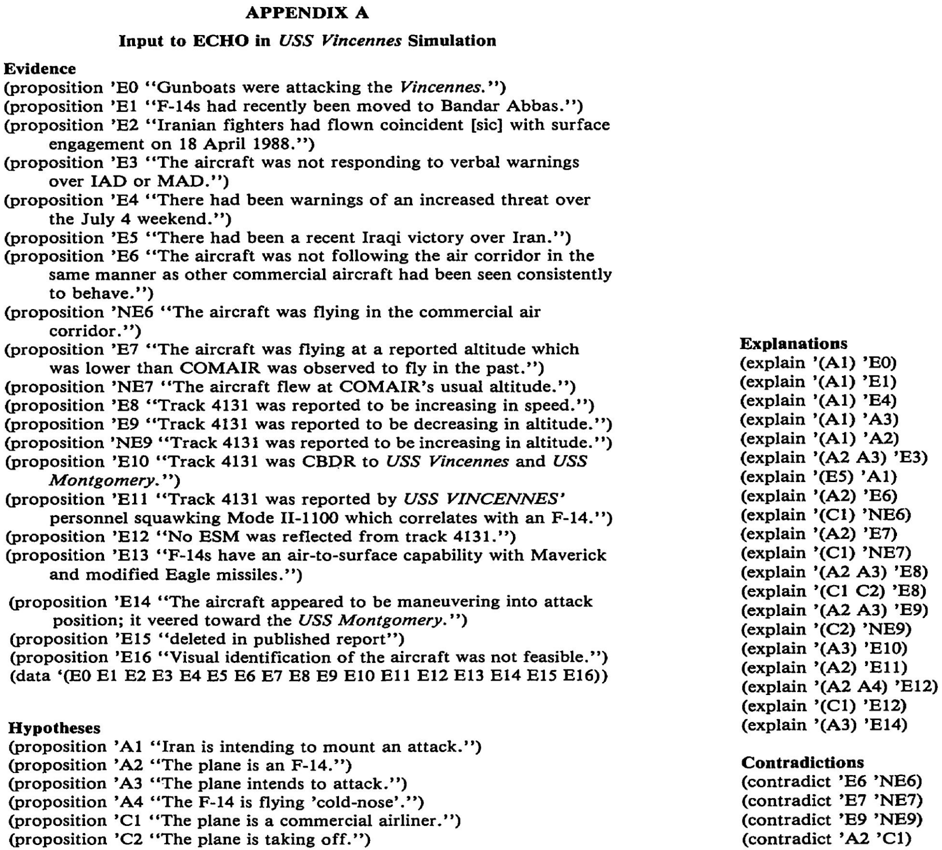
Adversarial Problem Solving: Modeling an Opponent Using Explanatory Coherence

PAUL THAGARD

Princeton University

Figure 1. Network produced by ECHO in *USS Vincennes* simulation using input given in Appendix A. Straight lines indicate excitatory links produced by virtue of explanations, whereas dotted lines represent inhibitory links formed in accord with the principles of contradiction and competition. Not shown are the special evidence unit and the links with evidence units, or the link between A2 and A4 created because together they explain E12.

LLMs can gauge local consistency



APPENDIX A

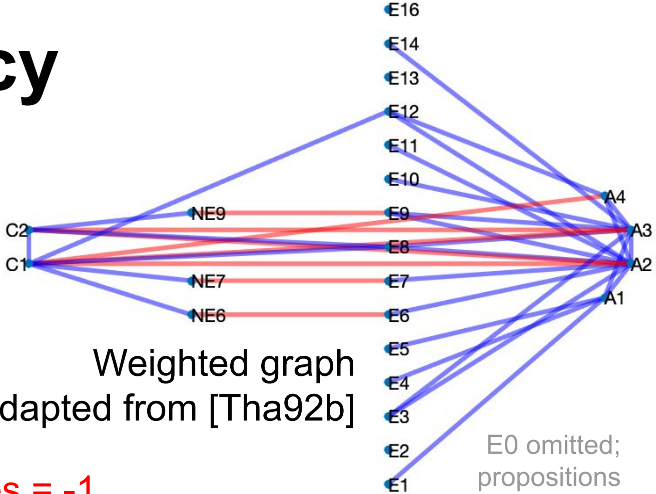
Input to ECHO in *USS Vincennes* Simulation

Evidence
(proposition ’E0 ‘‘Gunboats were attacking the *Vincennes*.’’)
(proposition ’E1 ‘‘F-14s had recently been moved to Bandar Abbas.’’)
(proposition ’E2 ‘‘Iranian fighters had flown coincident [sic] with surface engagement on 18 April 1988.’’)
(proposition ’E3 ‘‘The aircraft was not responding to verbal warnings over IAD or MAD.’’)
(proposition ’E4 ‘‘There had been warnings of an increased threat over the July 4 weekend.’’)
(proposition ’E5 ‘‘There had been a recent Iraqi victory over Iran.’’)
(proposition ’E6 ‘‘The aircraft was not following the air corridor in the same manner as other commercial aircraft had been seen consistently to behave.’’)
(proposition ’NE6 ‘‘The aircraft was flying in the commercial air corridor.’’)
(proposition ’E7 ‘‘The aircraft was flying at a reported altitude which was lower than COMAIR was observed to fly in the past.’’)
(proposition ’NE7 ‘‘The aircraft flew at COMAIR’s usual altitude.’’)
(proposition ’E8 ‘‘Track 4131 was reported to be increasing in speed.’’)
(proposition ’E9 ‘‘Track 4131 was reported to be decreasing in altitude.’’)
(proposition ’NE9 ‘‘Track 4131 was reported to be increasing in altitude.’’)
(proposition ’E10 ‘‘Track 4131 was CBDR to *USS Vincennes* and *USS Montgomery*.’’)
(proposition ’E11 ‘‘Track 4131 was reported by *USS VINCENNES’* personnel squawking Mode II-1100 which correlates with an F-14.’’)
(proposition ’E12 ‘‘No ESM was reflected from track 4131.’’)
(proposition ’E13 ‘‘F-14s have an air-to-surface capability with Maverick and modified Eagle missiles.’’)
(proposition ’E14 ‘‘The aircraft appeared to be maneuvering into attack position; it veered toward the *USS Montgomery*.’’)
(proposition ’E15 ‘‘deleted in published report’’)
(proposition ’E16 ‘‘Visual identification of the aircraft was not feasible.’’)
(data ‘(E0 E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12 E13 E14 E15 E16))

Hypotheses
(proposition ’A1 ‘‘Iran is intending to mount an attack.’’)
(proposition ’A2 ‘‘The plane is an F-14.’’)
(proposition ’A3 ‘‘The plane intends to attack.’’)
(proposition ’A4 ‘‘The F-14 is flying ’cold-nose’.’’)
(proposition ’C1 ‘‘The plane is a commercial airliner.’’)
(proposition ’C2 ‘‘The plane is taking off.’’)

Explanations
(explain ’(A1) ’E0)
(explain ’(A1) ’E1)
(explain ’(A1) ’E4)
(explain ’(A1) ’A3)
(explain ’(A1) ’A2)
(explain ’(A2 A3) ’E3)
(explain ’(E5) ’A1)
(explain ’(A2) ’E6)
(explain ’(C1) ’NE6)
(explain ’(A2) ’E7)
(explain ’(C1) ’NE7)
(explain ’(A2 A3) ’E8)
(explain ’(C1 C2) ’E8)
(explain ’(A2 A3) ’E9)
(explain ’(C2) ’NE9)
(explain ’(A3) ’E10)
(explain ’(A2) ’E11)
(explain ’(A2 A4) ’E12)
(explain ’(C1) ’E12)
(explain ’(A3) ’E14)

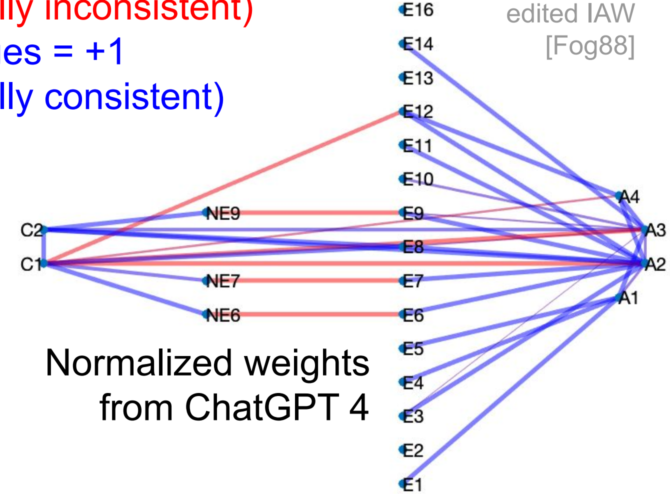
Contradictions
(contradict ’E6 ’NE6)
(contradict ’E7 ’NE7)
(contradict ’E9 ’NE9)
(contradict ’A2 ’C1)



Weighted graph
adapted from [Tha92b]

E0 omitted;
propositions
slightly
edited IAW
[Fog88]

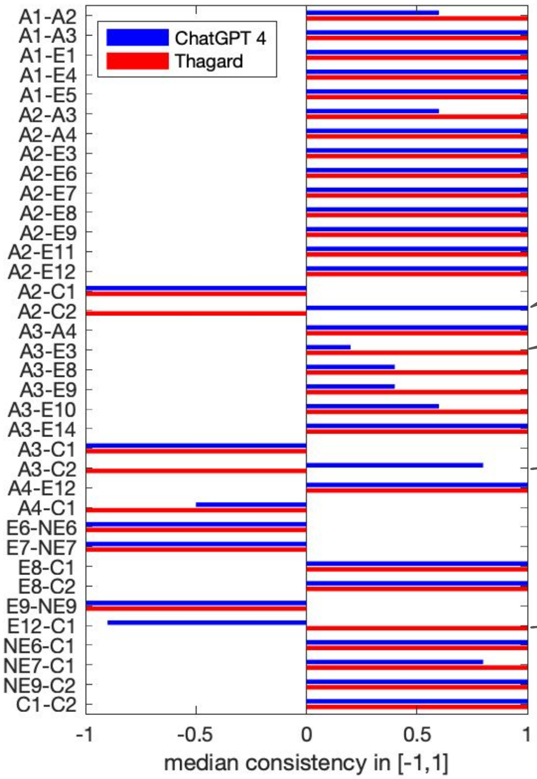
Red edges = -1
(maximally inconsistent)
Blue edges = +1
(maximally consistent)



Normalized weights
from ChatGPT 4

LLMs can gauge local consistency *better*

Vincennes graph edge weights



See also [Hun24]

"Track 4131 was an F-14."
"Track 4131 was taking off."

"Track 4131 intended to attack."
"Track 4131 was not responding to verbal warnings over [air distress frequencies]."

"Track 4131 intended to attack."
"Track 4131 was taking off."

"No [electronic emissions were reported] from track 4131, however, F-14s can fly [without electronic emissions]."
"Track 4131 was a commercial airliner."

ChatGPT 4 improves upon manual consistency ratings

Few-shot learning, context, & sampling yield reliable results

ChatGPT 4 cited technical failures & misunderstandings as plausible alternatives

ChatGPT 4 cited navigation and comms emissions of commercial airliners

Coherence comes in several flavors

	Propositions	Positive constraint	Discriminating priority
Explanatory	Hypotheses; data	Explanation; analogy	Data provenance
Analogical	Mapping hypotheses; mapping data	Functor*; natural transformation	Explanatory power
Deductive	Axioms; theorems	Entailment	Intuitiveness
Perceptual	Interpretations; data	Similarity; locality	Sensory provenance
Conceptual	Concepts	Correlation	Primitivity
Deliberative	Actions; goals	Affordance	Value

(Analogical, deductive, perceptual, conceptual, deliberative) coherence is a variant of explanatory coherence in which the role of explanation and/or analogy is played by a (mapping, entailment, similarity, correlation, affordance)

Positive (negative) constraints are generically a certification of (in)compatibility
Discriminating priorities are generically a form of provenance

* Thagard calls this “structure” but the categorical notion of functor captures his intent more precisely. https://mathoverflow.net/a/13834/