



Why “AI” is Only Half the Answer

Doug Lenat
Lenat@cyc.com



cycorp, 7718 Wood Hollow Drive, Austin, Texas





February 7, 2018



SPE Webinars are Sponsored by:

SPE
SPE Foundation







Doug Lenat
*Founder, Cycorp
AI Pioneer*



Douglas Lenat is one of the pioneers and leading researchers of Artificial Intelligence. He received his PhD in Computer Science from Stanford, investigating automated discovery based on "interestingness" heuristics, for which he received the 1977 IJCAI Computers and Thought Award. Besides professorship at CMU and Stanford, he was Principal Scientist at MCC, where he founded the Cyc Project in 1984 – something he dubbed "ontological engineering". At the end of 1994, he founded Cycorp, where he continues to serve as its CEO. Dr. Lenat is a Fellow of the AAAS, AAAI, and Cognitive Science Society, and an editor of the *J. Automated Reasoning*, *J. Learning Sciences*, and *J. Applied Ontology*. He's consulted for many organizations, including the White House, and is the only individual to have served on the Scientific Advisory Boards of both Microsoft and Apple.

Why AI is Only Half the Answer

- ❑ There are two AI power sources (statistics, logic)
- ❑ Everyone else is focusing on right-brain AI (brittle)
- ❑ One exception: CYC (Manhattan-Project-like effort)
- ❑ Now that it's more or less done, building a Cyc app. means extending Cyc's ontology, KB,... slightly
- ❑ The path to "real AI": statistics+logic synergy
 - A couple quick demo's of Cyc running.
 - Examples drawn from petroleum engineering Cyc app's.

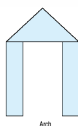
A typical "AI task" (to solve using ML or Logic):

Distinguish drawings or pictures of *arches* (i.e., from non-arches)

Examples for the concept "arch"



Arch

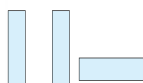


Arch

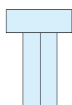
Non-examples for the concept "arch"



Non-arch



Near miss



Near miss



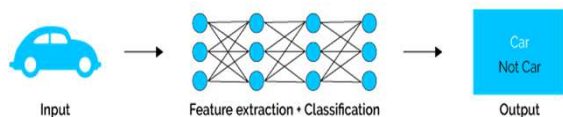
Non-arch

Source of Power #1: Statistics/Neural Nets/ Machine Learning/... trained on Big Data

Source of Power #2: Logic applied to
declaratively represented causal models

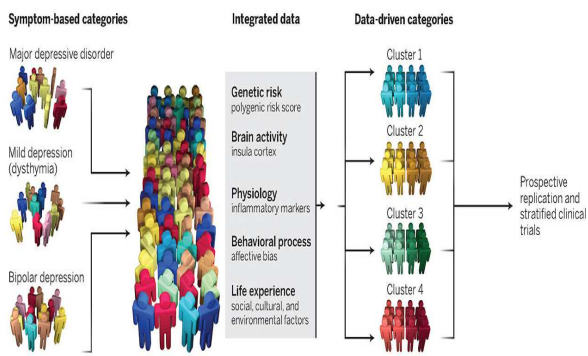
Source of Power #1: Statistics/Neural Nets/ Machine Learning/... trained on Big Data

- Feature Extraction (n low-level features for each image, text, etc.)
 - For images: low-level recognizers for edges, corners, ridges, curves, contiguous blob with the same color/texture throughout, etc.
 - For text: particular words and short-sequences-of-words, entities
- Classification (take a large number of those n-vectors as input)
 - Supervised: Someone draws a region in n-space and labels it
 - Unsupervised: Just look at the points in n-space and notice clusters

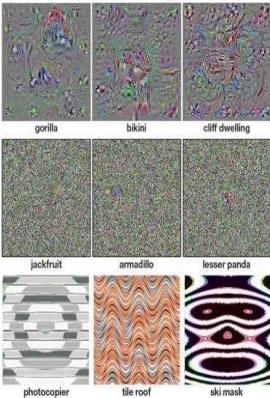


Deconstructed, parsed, and diagnosed.

A hypothetical example illustrates how precision medicine might deconstruct traditional symptom-based categories. Patients with a range of mood disorders are studied across several analytical platforms to parse current heterogeneous syndromes into homogeneous clusters.



Problem#1:
No explanations



gorilla bikini cliff dwelling
jackfruit armadillo lesser panda
photocopier tile roof ski mask

Problem#2:
Makes mistakes that no human would ever make

Problem#3:
Confirmation bias

Problem#4:
Disagreeing experts

Problem#5:
Multiple faults

Problem#6:
Only a small amount of data is available

Source of Power #1: Statistics/Neural Nets/
Machine Learning/... trained on Big Data

Source of Power #2: Logic applied to
declaratively represented causal models

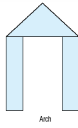
Source of Power #1: Statistics/Neural Nets/
Machine Learning/... trained on Big Data

Source of Power #2: Logic applied to
declaratively represented causal models

Examples for the concept "arch"



Arch

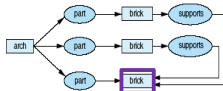


Arch

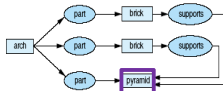
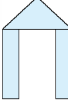
13

Generalization of descriptions to include multiple examples.

a. An example of an arch and its network description



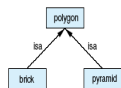
b. An example of another arch and its network description



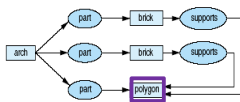
14

Generalization of descriptions to include multiple examples (cont'd)

c. Given background knowledge that bricks and pyramids are both types of polygons



d. Generalization that includes both examples



15

Specialization of a description to exclude a near miss.
In **c** we add constraints to **a** so that it can't match with **b**.

a. Candidate description of an arch



b. A near miss and its description



c. Arch description specialized to exclude the near miss

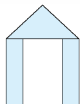


16

Examples for the concept “arch”



Arch

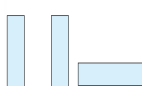


Arch

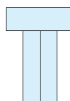
Non-examples for the concept “arch”



Non-arch



Near miss



Near miss



Non-arch

17

Logic (Cyc) handles situations that try Statistics (ML)

- Situations where multiple experts would disagree
- Situations where humans and ML algorithms suffer from confirmation bias
- Multiple “faults” – notably where the signals superpose and mask each other’s “telltales”
- Absurd answers (or worse: the final answer appears plausible, only the way it got that answer is absurd)
- Only a small amount of data is available

The Power of SMALL Data

$$\begin{array}{r} 35 \\ +17 \\ \hline 42 \end{array}$$

What did the kid do wrong?

The Power of SMALL Data

$$11x - 34 = 10$$

$$11x = 44$$

$$x = 33$$

← they add 34 to both sides (so far so good)

but then...

← the student's answer



MathCraft AI

divide both sides by 11x

I'm making this problem harder!

Let's look at the textbook

Let's look at a video

Let's first do a similar but simpler practice problem

$$11x - 34 = 10$$

He/she removes a coefficient by subtracting it; and always borrows during columnar subtraction

$$\begin{array}{r} 31 \\ 44 \\ -11 \\ \hline 213 \end{array}$$

Cyc predicts that some students will get the answer 213, for $44 - 11$, hence for "x", above,... and some do!

Cyc UNDERSTANDS why, and can then remediate.

$11x = 44$

$x = 44 - 11$

$x = 44/11$

$x = 33$

$x = 213$

$x = 4$

Specialization of a description to exclude a near miss.
In **c** we add constraints to **a** so that it can't match with **b**.

a. Candidate description of an arch



b. A near miss and its description



c. Arch description specialized to exclude the near miss



22

Kinds of Petroleum Extraction

PurposefulPhysicalAction
 RemovingSomething-Translational
 ExtractingPetroleum
 EnhancedOilRecovery
 EnhancedOilRecovery-GasInjection
 EnhancedOilRecovery-CO2Flooding
 EnhancedOilRecovery-NitrogenFlooding
 ThermallyEnhancedOilRecovery
 InSituCombustion
 ThermallyEnhancedOilRecovery-SteamInjection
 CyclicSteamStimulation
 ThermallyEnhancedOilRecovery-SteamFlooding
 ReverseSteamAssistedGravityDrainage
 SteamAssistedGravityDrainage
 VaporExtractionProcess

Kinds of Artificial lifts

HumanScaleObject
 PhysicalDevice
 Technology-Artifact
 ArtificialLiftDevice
 ElectricalSubmersiblePump
 GasLift
 ProgressingCavityPump
 Pumpjack



"The main purpose of artificial lift devices is to enable artificially lifted outflow."

(typePrimaryFunction-DeviceUsed ArtificialLiftDevice WellOutflow-ArtificiallyLifted)

Recognizing Bubble vs. Annular flow

"If the fluid moving in a multiphase fluid flow has a low concentration of gas by volume, Then it is bubble flow."

```
(forAll ?FLOW
  (forAll ?STUFF
    (implies
      (and
        (isa ?FLOW MultiphaseFluidFlow-Translation)
        (primaryObjectMoving ?FLOW ?STUFF)
        (hasConcentrationOfGasInFluidByVolume ?STUFF
          (VeryLowAmountFn ConcentrationOfGasInFluidByVolume)))
        (isa ?FLOW MultiphaseFluidFlow-BubbleFlow))))

"If the fluid moving in a multiphase fluid flow has a high concentration of gas by volume, Then it is annular flow."
(forAll ?FLOW
  (forAll ?STUFF
    (implies
      (and
        (isa ?FLOW MultiphaseFluidFlow-Translation)
        (primaryObjectMoving ?FLOW ?STUFF)
        (hasConcentrationOfGasInFluidByVolume ?STUFF
          (VeryHighAmountFn ConcentrationOfGasInFluidByVolume)))
        (isa ?FLOW MultiphaseFluidFlow-AnnularFlow))))
```

Oil Formation Volume Factor

*"Bo is equal to the ratio of R-VOL to S-VOL, where:
R-VOL is the volume at reservoir conditions of the liquid hydrocarbon in the reservoir payload, and
S-VOL is the volume at surface conditions of the liquid hydrocarbon in the reservoir payload."*

```
(implies
  (and
    (isa ?RESERVOIR NaturalReservoir)
    (reservoirPayload ?RESERVOIR ?HYDROCARBON)
    (aboveGroundRegion ?RESERVOIR ?SURFACE)
    (uniquePartOfType ?HYDROCARBON
      (LiquidFn Hydrocarbon) ?OIL)
    (objectHasSlotValueAtConditions ?OIL volumeOfObject ?R-VOL ?RESERVOIR)
    (ratioOfTo ?R-VOL ?S-VOL ?RATIO)
    (objectHasSlotValueAtConditions ?OIL volumeOfObject ?S-VOL ?SURFACE))
    (oilFormationVolumeFactor ?RESERVOIR ?RATIO))
```

Multiphase vertical outflow

```
(equationFormForMultiphaseOilVerticalOutflowFormula
  (numericallyEquals
    (The
      (RateOfChangeRelativeToRateOfChangeFn Pressure Height))
    (PlusFn
      (TimesFn
        (DensityFn
          (The MultiphaseHydrocarbonFluid))
        (VectorMagnitudeFn
          (The GravitationalForceVector))
        (CosineFn
          (InclinationFn
            (The ProductionTubing))))
        (QuotientFn
          (TimesFn
            (FrictionFactorOfFluidFlowFn
              (The FluidFlowEvent))
            (DensityFn
              (The MultiphaseHydrocarbonFluid))
            (SpeedFn
              (The MultiphaseHydrocarbonFluid)))
            (DiameterOfCylinderFn
              (The ProductionTubing))))
        (TimesFn
          (DensityFn
            (The MultiphaseHydrocarbonFluid))
          (SpeedFn
            (The MultiphaseHydrocarbonFluid))
          (The
            (RateOfChangeRelativeToRateOfChangeFn
              (SpeedFn
                (The MultiphaseHydrocarbonFluid))
              (HeightFn
                (The ProductionTubing))))))
```

VERTICAL FLOW

GENERAL SOLUTION :

$$\left(\frac{dp}{dh}\right)_{TOTAL} = \left(\frac{dp}{dh}\right)_{ELEVATION} + \left(\frac{dp}{dh}\right)_{FRICTION} + \left(\frac{dp}{dh}\right)_{ACC}$$

$$= \rho g \cos \theta + \frac{f \rho v^2}{2d} + \rho v \frac{dv}{dh}$$

TYPICALLY IN OIL WELLS :

> 90%	< 10%	NEGLECTIBLE EXCEPT FOR MIST FLOW AT LOW ρ AND HIGH VELOCITY
-------	-------	--

TO SOLVE EQUATION, NEED TO DETERMINE

$P > P^*$

FOR 2-PHASE FLOW !

Why Bother Doing All That? i.e., why not “live with” ML?

- Sometimes you want to see the **explanation** (or even: pro/con arguments for top **n** answers)
- Sometimes you care about rare conditions (where there is little or no training data)
- **Statistics can be inhumanly brittle**
- Multiple disagreeing experts;...

The Brittleness Bottleneck — c. 1975

Skin Disease Expert System

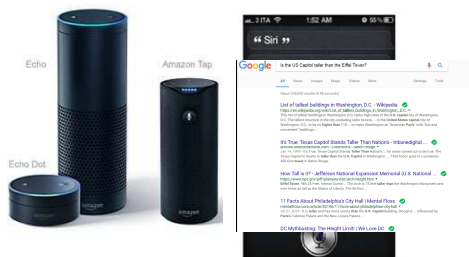
- “Are there spots on the body?”
- “What color are the spots?”
- “Are there more on the trunk than elsewhere?”

After telling it about my old, rusted-out Pontiac:
“The patient has measles.”

Veneer of intelligence, not intelligence

29

The Brittleness Bottleneck – today!



Veneer of intelligence, not intelligence

BloombergView Safety Numbers on Autonomous Cars Just Don't Add Up

Safety Numbers on Autonomous Cars Just Don't Add Up

California has automakers report "disengagement" episodes. But what does that mean?

By Edward Hlademayer
February 2, 2018, 9:08 AM CST



former editor at the Washington Post
Read more
Follow @ewahlademayer

LIVE TV
AUDIO



Google Is the US Capitol taller than the Eiffel Tower?

All News Images Maps Videos More Settings Tools

About 345,000 results (0.88 seconds)

List of tallest buildings in Washington, D.C. - Wikipedia ✓
https://en.wikipedia.org/wiki/List_of_tallest_buildings_in_Washington,_D.C.
This list of tallest buildings in Washington, D.C. ranks high-rises in the U.S. capital city of Washington, D.C. The tallest structure in the city, excluding radio towers, ... in the United States capital city of Washington, D.C., to be no higher than 110 ... to make Washington an "American Paris" with "low and convenient" buildings ...

It's True: Texas Capitol Stands Taller Than Nation's - tribunedigital ... ✓
articles.orlandosentinel.com/collections/baton-rouge
Jan 14, 1999 - It's True: Texas Capitol Stands Taller Than Nation's ... for years turned out to be true: The Texas Capitol in Austin is taller than the U.S. Capitol in Washington. ... That honor goes to Louisiana's 450-foot tower in Baton Rouge.

How Tall is it? - Jefferson National Expansion Memorial (U.S. National ... ✓
<https://www.nps.gov/jeff/planyourvisit/arch-height.htm>
Eiffel Tower, 984.25 Feet Interior Dome ... The Arch is 75 feet taller than the Washington Monument and over twice as tall as the Statue of Liberty. The 60-foot ...

11 Facts About Philadelphia's City Hall | Mental Floss ✓
mentalfloss.com/article/83196/11-facts-about-philadelphia-city-hall
Jul 27, 2016 - It is taller and has more rooms than the U.S. Capitol Building, though it ... Influenced by Paris's Tuileries Palace and the New Louvre Palace, ...

DC Mythbusting: The Height Limit! | We Love DC ✓
<https://www.washingtonpost.com/2016/07/27/dc-myth-busting-the-height-limit/>

Google Who was President when Barack Obama was born?

About 43,900,000 results (0.59 seconds)

Born on August 4, 1961, in Honolulu, Hawaii, Barack Obama is the 44th and current **president of the United States**. He was a community organizer, civil rights lawyer and teacher before pursuing a political career. He was elected to the Illinois State Senate in 1996 and to the U.S. Senate in 2004.

Barack Obama - U.S. President, Lawyer, U.S. Senator ...
www.biography.com/people/barack-obama-12762369

Barack Obama - U.S. President, Lawyer, U.S. Senator ...
www.biography.com/people/barack-obama-12762369

Born on August 4, 1961, in Honolulu, Hawaii, Barack Obama is the 44th and current president of the United States. He was a community organizer, civil rights lawyer and teacher before pursuing a political career. He was elected to the Illinois State Senate in 1996 and to the U.S. Senate in 2004.

About Barack Obama
https://www.biography.com/people/barack-obama-12762369

He was born on August 4th, 1961, in Honolulu, Hawaii, to a mother from Kansas, Stanley Ann Dunham, and a father from Kenya, Barack Obama Sr. He was also raised by his grandfather, who served in Patton's army, and his grandmother, who worked her way up from the secretarial pool to become vice president of a local bank.

In the news
Misperceptions persist about Obama's faith, but aren't so widespread
CNN - 1 day ago
Barack Obama's religious beliefs and background persist, Obama was born in Hawaii and is a Christian. ... Overall, 20% of Americans say they think the President is a ...
Poll: 20% believe Barack Obama was born outside US

The Brittleness Bottleneck – still pervades software today!

AI Systems had no model of what they did/didn't know,
i.e., what the limits of their expertise actually were.
Little if any understanding of *context* in which rules/fact were stated
Limited expressivity in representation language (for efficiency)

Humans, when confronted with a novel situation/problem,...
...can fall back on increasingly **general knowledge**
...can **analogize** to superficially far-flung situations/experiences
...can **reason** (infer things based on what we already know)

35

Types of Reasoning We Do

- Barack Obama was born August 4, 1961
- JFK was President Jan. 20, 1961- Nov. 22, 1963

DEDUCE

INDUCE

ANALOGIZE

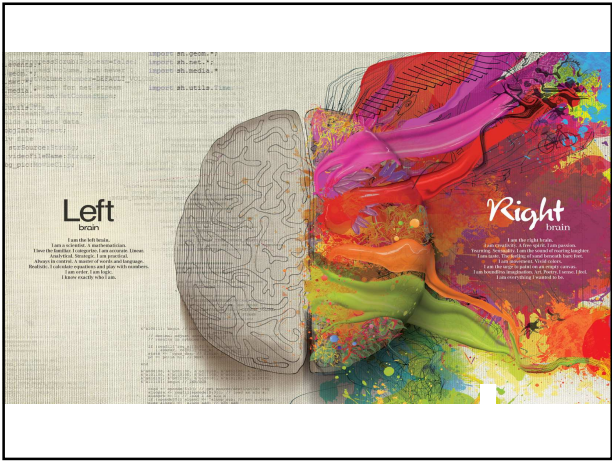
ABDUCE

JFK was President when Obama was born

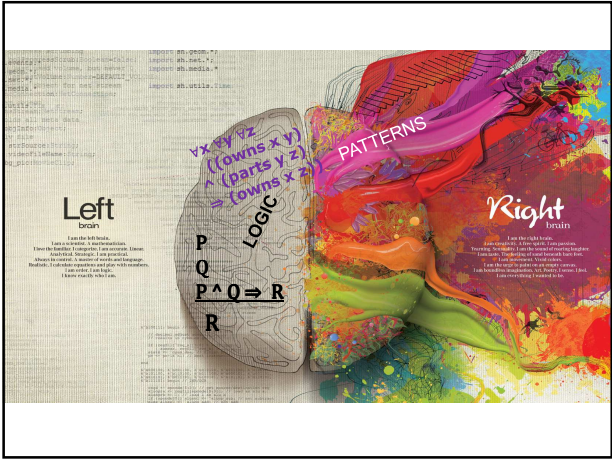
All the swans I've seen so far are white...

Repetitive physical exercise builds up muscles...

Someone walks in with their clothes dripping wet...







Q: Who was the Prime Minister of the UK when Theresa May was born?

P
Q
 $P \wedge Q \Rightarrow R$
R

Google Who was President when Barack Obama was born

All News Images Shopping Videos More Settings Tools

United States of America > President (1961)

John F. Kennedy  Dwight D. Eisenhower 

Barack Obama - U.S. President, Lawyer, U.S. Senator - Biography.com 
www.biography.com/people/barack-obama-12782369
 Jan 31, 2017 · Barack Hussein Obama II was born on August 4, 1961, in Honolulu, Hawaii. His mother, Ann Dunham, was born on an Army base in Wichita, Kansas, during World War II.

Barack Obama - Wikipedia 
https://en.wikipedia.org/wiki/Barack_Obama
 Barack Hussein Obama II is an American politician who served as the 44th President of the ...
 Obama was born in Honolulu, Hawaii, two years after the territory was admitted to ... Obama was born on August 4, 1961, at Kapiʻolani Maternity ...

Barack Obama citizenship conspiracy theories - Wikipedia 
https://en.wikipedia.org/wiki/Barack_Obama_citizenship_conspiracy_theories
 ... House website, reaffirming that he was born on August 4, 1961, in Honolulu, Hawaii. During Barack Obama's campaign for president in 2008 and through his presidency, many ...
 Background: Release of the birth certificates - False claims

Google Who was Prime Minister of the UK when Theresa May was born?

All News Images Videos Shopping More Settings Tools

About 1,200,000 results (1.01 seconds)

May was the longest serving **home secretaries** in British history. Theresa May is Britain's first female Prime Minister since **Margaret Thatcher**. She is considered one of the most powerful Conservative women and was the longest serving **home secretaries** in British history. Mar 29, 2017

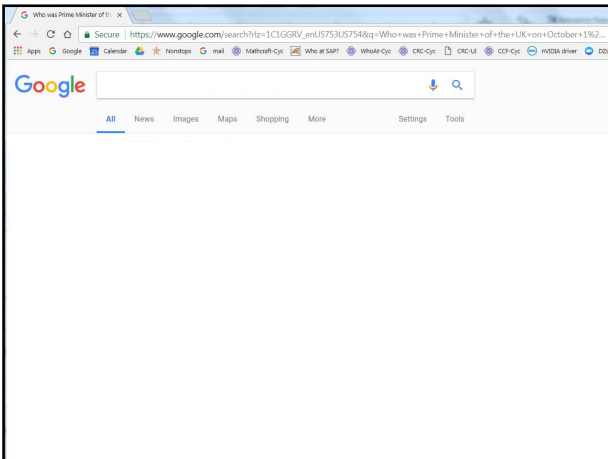
Theresa May - Prime Minister, Government Official - Biography.com 
<https://www.biography.com/people/theresa-may-071116>

October 1, 1956

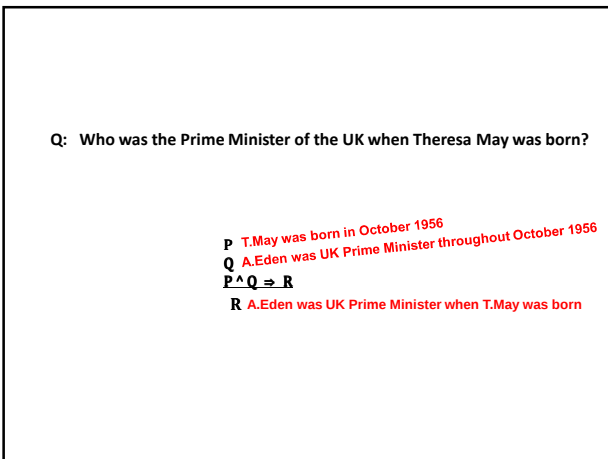
People also ask

- Is Theresa May married? 
- Who is the new Prime Minister of UK? 
- Who is the prime minister of England right now? 
- What is the salary of the British prime minister? 

Theresa May - Wikipedia 
https://en.wikipedia.org/wiki/Theresa_May
 Appointment. On 13 July 2016, two days after becoming Leader of the Conservative Party, May was appointed Prime Minister by Queen Elizabeth II, becoming only the second female British Prime Minister after Margaret Thatcher, and the first female British Prime Minister of the 21st century.
 Early career - Home Secretary - Minister for Women and ... - Prime Minister







Q: Who was the Prime Minister of the UK when Theresa May was born?

P (During (BirthEvent TheresaMay) (OctoberOf 1956))
 Q (During (OctoberOf 1956) (PrimeMinistry UK AnthonyEden))
 $P \wedge Q \Rightarrow R$
 R (During (BirthEvent TheresaMay) (PrimeMinistry UK AnthonyEden))

Q: Who was the Prime Minister of the UK when Theresa May was born?



P (During (BirthEvent TheresaMay) (OctoberOf 1956))
 Q (During (OctoberOf 1956) (PrimeMinistry UK AnthonyEden))
 $P \wedge Q \Rightarrow R$
 R (During (BirthEvent TheresaMay) (PrimeMinistry UK AnthonyEden))

CYC: A large, broad common sense knowledge base expressed in higher order logic + fast inference engine

Example: Less Brittle Searching

How could the results be more relevant if the search engine had some **logical understanding**?
(fewer false positives (hard) and false neg's. (very hard!))

The basic idea:

Get the computer to **understand**, not just store, information.

I.e., **represent the meaning** of the information.

Then it can **reason logically** to answer your queries, to generate plausible scenarios, to analogize, etc.

The New York Times article snippet shows a headline "1993 pipe bombing of France's embassy in Lebanon." with a sub-headline "Explosion Shook Beirut, Killed 29, Wounded 100." and a photo of the damaged building. A purple dotted arrow labeled "CIC" points from the article to a text box.

Intelligence Analyst's Query:

"Government buildings damaged in terrorist events in Beirut in the decade before 9/11"

Query from Analysts:

"...attacks in which no civilians died..."

2 Reports about the Khobar Towers Bombing:

abc NEWS
"There were 17 fatalities."

CNN
"17 soldiers died."

Query from Analysts:

"...attacks in which no civilians died..."

Civilians are disjoint from soldiers

```
(ForAll ?P
  (ForAll ?C
    (implies
      (and
        (isa ?P Person)
        (children ?P ?C))
      (loves ?P ?C))))
```

When you become happy, you smile.
 You become happy when someone
 you love accomplishes a milestone.
 Taking one's first step is a milestone.
 Parents love their children.

**"Jack watched his daughter take her first step.
 His expression changed a moment later, when..."**

0. $\forall x \forall y (\text{sees } x \ y) \Rightarrow (\text{believes } x \ y)$ $\frac{1}{2}$ $\forall x \forall y \forall z ((\text{believes } x \ y) \wedge (y \Rightarrow z)) \Rightarrow (\text{believes } x \ z)$

1. $\forall x \forall y (\text{parent } x \ y) \Rightarrow (\text{loves } x \ y)$
2. $\forall h (\text{happy } h) \Leftrightarrow (\text{smiling } h)$
3. $\forall d \forall p (\text{daughter } d \ p) \Rightarrow (\text{parent } p \ d)$
4. $\forall r \forall e ((\text{loves } r \ e) \wedge (\text{believes } r \ (\text{accomplishingBigMilestone } e))) \Rightarrow (\text{happy } r)$
5. $\forall w (\text{takingFirstStep } w) \Rightarrow (\text{accomplishingBigMilestone } w)$
6. $(\text{daughter } \text{DAU} \ \text{Jack})$
7. $(\text{takingFirstStep } \text{DAU})$
8. $(\forall s (\text{Not } (\text{smiling } s)))$ **8½. (sees Jack (takingFirstStep DAU))**

9. (parent Jack DAU)	From 6 and 3, DAU/d Jack/p
10. (loves Jack DAU)	From 9 and 1, Jack/x DAU/y
11. $(\text{accomplishingBigMilestone DAU})$	From 7 and 5, DAU/w
12. $(\text{believes Jack (takingFirstStep DAU)})$	From 0 and 8½, Jack/x (takingFirstStep DAU)/y
13. $(\text{believes Jack (accomplishingBigMilestone DAU)})$	From 12, ½, 5, and 7
14. (happy Jack)	From 13 and 4, DAU/e Jack/r
15. (smiling Jack)	From 14 and 2, Jack/h

QED! We are done! 15 is the same as 8, which is what we were trying to prove.

E&Ps Hedging Production To Support Cash Flow and Returns Goals

Stephen Whitfield, Senior Staff Writer | 24 January 2018



Heading to increased investor pressure, a new report from IHS Markit indicates that North American exploration and production (E&P) companies were more hedged than usual entering 2018.

The report, IHS Markit's "Commodity Price Outlook: North American E&P," showed that a group of 41 companies studied had hedged 20% of oil production, or 127 million barrels per day (bbl/d), by the end of 2017. The companies had also hedged 36% of gas production, or 10.37 billion cubic feet (bcf), or \$15.5MMcf. This is an increase from 15% of oil and 36% of gas production hedged at the end of 2016.

Phil Chenoweth, principal energy analyst at IHS Markit and author of the analysis, said in a statement that the hedging from the E&P peer group "signals a shift in the way E&P companies are taking advantage of the initial oil price rally to the low \$40s per barrel zone."

"Companies are seeking more predictable cash flows because of greater investor demand to improve corporate returns and keep capital spending within cash flow," he said. "The higher level of hedging is less about supporting operational production, growth and more about providing financial confidence that those companies are capable about managing more predictable operations."

If at any given depth within the monitored duration the moving average DTS temperature in the liner is within the specified threshold value of the moving average of the saturation steam temperature, and the DTS curve is trending toward a relatively stable saturation steam temperature, the [REDACTED] SAGD production engineers consider there to be a possible impending steam breakthrough.

The Trump administration has denied ExxonMobil permission to bypass sanctions to drill for oil in Russia.

"The Treasury Department will not be issuing waivers to U.S. companies, including Exxon, authorizing drilling prohibited by current Russian sanctions," Treasury Secretary Steven Mnuchin, said in a statement Friday. Mnuchin said he consulted with President Trump on the decision.

Exxon had applied for a waiver from sanctions imposed by the Obama administration in a bid to resume its lucrative joint venture with Russian state oil giant PAO Rosneft.

In a statement, Exxon said "we understand" the decision by the Treasury Department. Exxon explained that its application for a license was aimed at meeting the company's "contractual obligations" in Russia, where competitors are allowed to drill under European sanctions.

There are powerful reasons why Exxon would want to get back into business with Rosneft: Their agreement to form a joint venture, signed in 2011, allowed Exxon to conduct offshore exploration in the Black Sea and the Kara Sea in Siberia.

The undeveloped oil fields are thought to be the most promising in the Russian Arctic, according to S&P Global Platts.

In 2012, the two firms agreed to jointly develop more oil reserves in Siberia and establish a

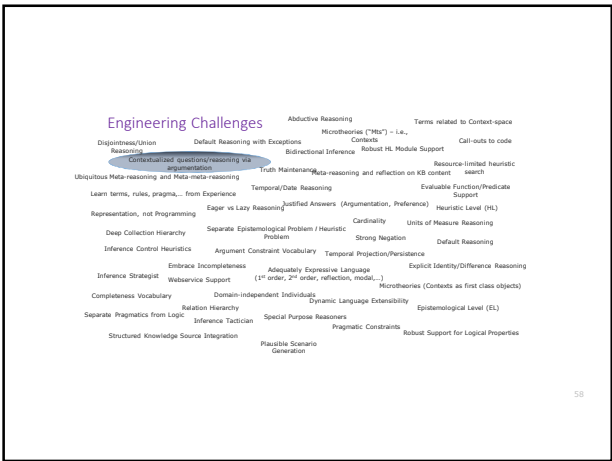
What factors argue <for/against> the conclusion that <ETA> <performed> <the Madrid train station attacks>?

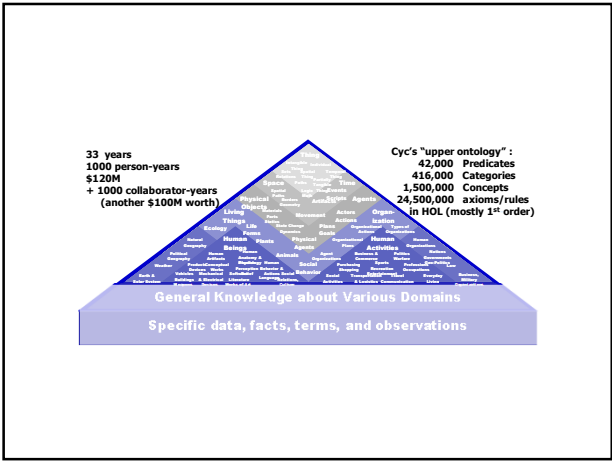
For:

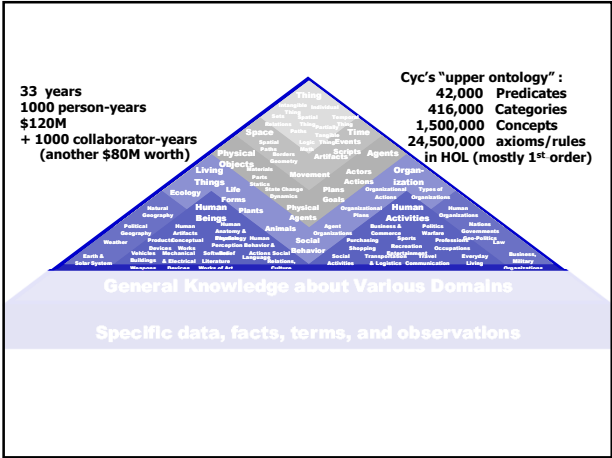
- ETA often executes attacks near national election
- ETA has performed multi-target coordinated attacks
- Over the past 30 years, ETA performed 75% of all terrorist attacks in Spain
- Over the past 30 years, 98% of all terrorist attacks in Spain were performed by Spain-based groups, and ETA is a Spain-based group.

Against:

- ETA warns (a few minutes ahead of time) of attacks that would result in a high number civilian casualties, to prevent them. There was no such warning prior to this attack.
- ETA generally takes responsibility for its attacks, and it did not do so this time.
- ETA has never been known to falsely deny responsibility for an attack, and it did deny responsibility for this attack.







ELIZA (DOCTOR)

Joe Weizenbaum, MIT

- Carl Rogers advocated **reflection**

Patient: "I swear a lot."

"How do you feel about the fact that you swear a lot?"

- Counted a lot on anthropomorphism

- Count on patient-to-doctor respect/awe/...

Patient: "My dog's mother died recently."

"Tell me more about **your** mother."

Veneer of intelligence, not intelligence

ELIZA (DOCTOR)

"blah blah blah **mother** blah blah blah"

Patient: "My dog's mother died recently."

"Tell me more about your mother."

ELIZA



"blah blah blah **mother** blah blah blah"

Patient: "My dog's mother died recently."

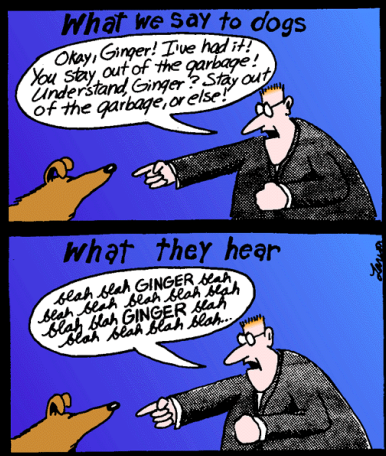
"Tell me more about your mother."

ELIZA

“blah blah blah”

Patient: “My dog’s mother died recently.”

“Tell me more about your mother.”



ELIZA (DOCTOR)

“blah blah blah **mother** blah blah blah”

Patient: “My dog’s mother died recently.”

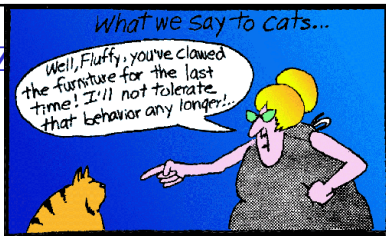
“Tell me more about your mother.”

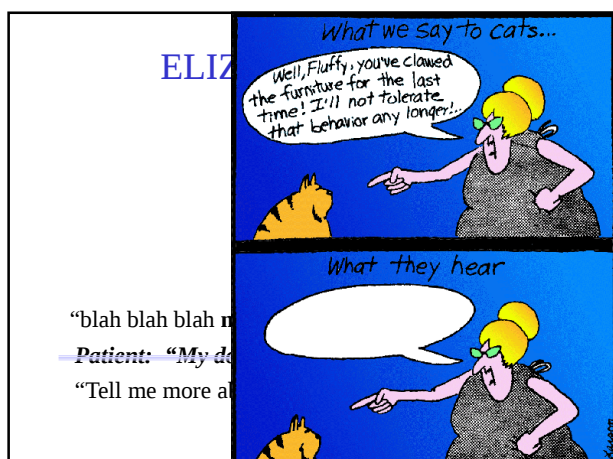
ELIZA

“blah blah blah **mother** blah blah blah”

Patient: “My dog’s mother died recently.”

“Tell me more about your mother.”





ELIZA (DOCTOR)

“blah blah blah **mother** blah blah blah”

Patient: “My dog’s mother died recently.”

“Tell me more about your mother.”

ALICE (today!!)

- JUDGE11[9:26:0]I WILL ASK YOU SOME SIMPLE QUESTIONS AND GIVE SIMPLE COMMANDS
- PROGRAM[9:26:0]Do you think your plan will succeed?
- JUDGE11[9:26:16]WHAT IS THE COLOR OF A BLUE CAR
- PROGRAM[9:26:16]WHAT COLOR IS a blue car
- JUDGE11[9:26:35]WHERE IS SUE’S NOSE WHEN SUE IS IN HER HOUSE
- PROGRAM[9:26:35]Where it belongs.
Try searching the internet.

Find and clean (consistency-check) information by inference (+KB)

XYZCo ID #	birth date	hire date	salutation	first name	last name	emerg contact	signif other
8041	9/1/57	8/5/91	Mr	Pat	Jones	8053	8053

Find and clean (consistency-check) information by inference (+KB)

XYZCo ID #	birth date	hire date	salutation	first name	last name	emerg contact	signif other
8041	9/1/57	8/5/91	Mr	Pat	Jones	8053	8053
8053	3/3/49	2/9/48	Ms	Jan	Smith	8053	8199

What Led to Cyc?

Motive

1. Programs need general world knowledge, and commonsense, to break the “brittleness bottleneck”

NL understanding speech understanding, robotics, learning, expert systems, search,...

Means

2. We know (knew) enough to do this; it is (was) more an engineering task than a scientific research task.

Opportunity

3. The time was right (1984).

Natural Language Understanding requires having lots of knowledge

1. The pen is in the box.
The box is in the pen.
2. The police watched the demonstrators...
...because they feared violence.
...because they advocated violence.

Natural Language Understanding requires having lots of knowledge

3. Mary and Sue are sisters.
Mary and Sue are mothers.

Natural Language Understanding requires having lots of knowledge

4. Every American has a mother.
Every American has a president.
5. John saw his brother skiing on TV. The fool...
...didn't have a coat on!
...didn't recognize him!

Natural Language Understanding requires having lots of knowledge

6. George Burns: "My aunt is in the hospital.
I went to see her today, and **took** her flowers."

Gracie Allen: "George, that's terrible!
You should have **brought** her flowers!"

Took
Table
Sanction
⋮

Natural Language Understanding requires having lots of knowledge

7. "and" sometimes means "or":

"Patients who had appendectomies in 2008 **and** 2009."

8. coordinating a modifier: usually "it's obvious"

- "... the agents who were no longer spying **or** dead in 2011"
- "... the agents who were no longer spying **or** recruiting in 2011"

Natural Language Understanding requires having lots of knowledge

9. • Tim was mad at Joe, because he stole his lunch.
Joe Tim's
- Tim was mad at Joe, so he stole his lunch.
Tim Joe's

- They ate the pie **with** ice cream.
- They ate the pie **with** rhubarb.
- They ate the pie **with** paper plates.
- They ate the pie **with** cold milk.
- They ate the pie **with** friends.
- They ate the pie **with** dinner.
- They ate the pie **with** enthusiasm.
- They ate the pie **with** spoons.
- They ate the pie **with** napkins.

What Led to Cyc?

Motive

1. Programs need general world knowledge, and commonsense, to break the “brittleness bottleneck”

NL understanding, speech understanding, robotics, learning, expert systems, search,...

Means

2. We know (knew) enough to do this; it is (was) more an engineering task than a scientific research task.

Opportunity

3. The time was right (1984).

Berkeley Technology Law Journal

Volume 1
Issue 1 Spring

Article 5

January 1986

The National Cooperative Research Act of 1984: A New Antitrust Regime for Joint Research and Development Ventures

Christopher O.B. Wright

H
ATION
E AND JAPAN'S
ALLENGE
LD
ENBAUM
ORDUCK



The New York Times **Business Day** Search

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL

Search International | DealBook | Markets | Economy | Energy | Media | Technology | Personal Tech

8 % ANNUAL ANNUITY RETURN
Get guaranteed lifetime income and reduced risks to retirees all here.

'Fifth Generation' Became Japan's Lost Generation
By ANDREW POLLACK
Published June 5, 1992

TOKYO, June 4— A bold 10-year effort by Japan to seize the lead in computer technology is fizzling to a close, having failed to meet many of its ambitious goals or to produce technology that Japan's computer industry wanted.

After spending more than \$400 million on its widely heralded Fifth Generation computer project, the Japanese Government said this week that it was willing to give away the software developed by the project to anyone who wanted it, even foreigners. Machines That

FACEBOOK
TWITTER
GOOGLE+
EMAIL
SHARE
PRINT
REFRINTS

\$50
WITH SHIP
\$1

What Led to Cyc?

Motive

1. Programs need general world knowledge, and commonsense, to break the "brittleness bottleneck"
NL understanding, speech understanding, robotics, learning, expert systems, search,...

Means

2. We know (knew) enough to do this; it is (was) more an engineering task than a scientific research task.

Opportunity

3. The time was right (1984).

The Knowledge Principle

A system exhibits intelligent understanding and action at a high level of competence primarily because of the **knowledge** it can bring to bear: the concepts, facts, representations, models, methods, metaphors, and heuristics about its domain of endeavor.

If a program is to perform a complex task well, it must know a great deal about the world in which it operates.

The Breadth Hypothesis

To behave intelligently in unexpected situations, the problem solver often must:

- fall back on increasingly general knowledge
- analogize to specific far-flung knowledge

To avoid the “more knowledge => slow” explosion, rely on meta-knowledge: static (ontology) or dynamic (rules)

AI as Empirical Inquiry

Premature mathematization, or focusing on toy problems, washes out details from reality that later turn out to be significant.

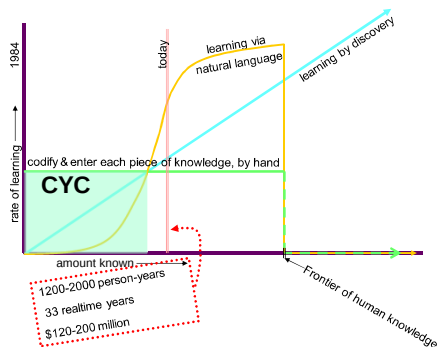
- Don’t succumb to physics envy
- Test the growing AI experimentally, *fasifiably*
 - On real problems: “Nature” still holds most key surprises.
 - On difficult problems: it actually makes development easier.

2-Stage Research Program (Cyc)

1. Slowly hand-code a v. large, v. broad KB.
2. When enough knowledge is present, the system should actively, increasingly, help with the knowledge acquisition process.

It should be faster to acquire more from texts, databases, [websites], interactive dialogues, etc. (gradually transitioning more and more to "learning by discovery" than by "reading")

Building Cyc qua Engineering Task



Why "AI" is Only Half the Answer

- ☒ There are two AI power sources (statistics, logic).
- ☒ Everyone else is focusing on right-brain AI (brittle)
- ☒ **One exception: Cyc (Manhattan-Project-like effort)**
- ☐ Now that it's more or less done, building a Cyc app. means extending Cyc's ontology, KB,... slightly
- ☐ The path to "real AI": statistics+logic synergy
 - A couple quick demo's of Cyc running.
 - Examples drawn from petroleum engineering Cyc app's.

Programming (and problem-solving) by

- Statistical trial and error / Machine Learning
- Brain Surgery / Software Engineering
- Education – Accretion of concepts, skills, facts, techniques, heuristics, rules,... leading to an **incremental asymptotic approach to competence**

Approach: Teaching...

• Massive KB
• + Inference

Approach: Engineering

92

33 years
1000 person-years
\$120M
+ 1000 collaborator-years
(another \$80M worth)

Cyc's "upper ontology":
42,000 Predicates
416,000 Categories
1,500,000 Concepts
24,500,000 axioms/rules
in HOL (mostly 1st order)

General Knowledge about Various Domains

Specific data, facts, terms, and observations

Query:
Is no [infant](#) a [doctor](#)?

Answer:
Yes.

Because: [Detailed Justification](#) [Linear Justification](#)

Detailed Justification:

- ▼ No [infant](#) is a [doctor](#).
 - ▶ No [adult](#) is a [young animal](#).
- ▼ An [infant](#) is a type of [young animal](#).
 - ▶ An [infant](#) is a type of [preschool-age child](#).
 - ▶ A [preschool-age child](#) is a type of [child](#).
 - ▶ A [child](#) is a type of [juvenile](#).
 - ▶ A [juvenile](#) is a type of [young animal](#).
- ▼ A [doctor](#) is a type of [adult](#).
 - ▶ A [doctor](#) is a type of [licensed-professional](#).
 - ▶ [Licensed-professional](#) is a type of [adult worker](#).
 - ▶ An [adult worker](#) is a type of [human adult](#).
 - ▶ A [human adult](#) is a type of [adult](#).

CST-NoInfantAveDoctor

Query:
Is it true or false that [cats](#) are capable of [cancan dancing](#)?

Answer:
False.

Because: [Detailed Justification](#) [Linear Justification](#)

Detailed Justification:

- ▼ #SentenceTruth holds of [cats](#) are capable of [cancan dancing](#) and [No](#).
 - ▼ [Cats](#) are not capable of [cancan dancing](#).
 - ▶ [Inanimate objects](#) cannot play the role of doers in animate activity.
 - ▶ Type Behavior Incapable and Type Capable Fn Behavior Capable cannot hold of the same thing at the same time.
 - ▶ If someone deliberately performed some action, then that individual did that action.
 - ▼ A [cat](#) is a type of [inanimate object](#).
 - ▶ A [cat](#) is a type of [capsule](#).
 - ▶ A [capsule](#) is a type of [storage device](#).
 - ▶ A [storage device](#) is a type of device that is not a weapon.
 - ▶ A device that is not a weapon is a type of [device](#).
 - ▶ A [device](#) is a type of solid tangible artifact.
 - ▶ A solid tangible artifact is a type of [artifact](#).
 - ▶ An [artifact](#) is a type of non-natural inanimate thing.
 - ▶ A non-natural inanimate thing is a type of [inanimate object](#).
 - ▼ [Cancan dancing](#) is a type of animate activity.
 - ▶ An [action performed by animals](#) is a type of animate activity.
 - ▶ [Cancan dancing](#) is a type of [action performed by animals](#).
 - ▶ [Cancan dancing](#) is a type of [dancing by humans](#).
 - ▶ [Dancing by humans](#) is a type of [performing art](#).
 - ▶ A [performing art](#) is a type of cultural activity.
 - ▶ A cultural activity is a type of [human activity](#).
 - ▶ A [human activity](#) is a type of [action performed by animals](#).

CST-CatAveCancanDance

Query:
If

- [COFFEE](#) is [hot coffee](#).
- [CUP](#) is a [coffee cup](#).
- and [CUP](#) is [upside-down](#).

 is it false that [COFFEE](#) is in the open container [CUP](#)?

Answer:
Yes.

Because: [Detailed Justification](#) [Linear Justification](#)

Detailed Justification:

- ▼ [COFFEE](#) is not in the open container [CUP](#).
 - ▶ It is false that
 - something is a [container](#).
 - it is [upside-down](#).
 - it is immersed in some fluid substance.
 - that fluid substance has the density some density.
 - some other quantity is greater than that density.
 - some other thing has the density that other quantity.
 - and that other thing is in the open container the solid object.
- ▼ [CUP](#) is a [container](#).
 - ▶ A [coffee cup](#) is a type of [Container](#).
 - ▶ [CUP](#) is a [coffee cup](#).
- ▶ [CUP](#) is [upside-down](#).
- ▼ [CUP](#) is immersed in [the atmosphere](#).
 - ▶ All terrestrial functioning object is immersed in [the atmosphere](#).
 - ▶ [CUP](#) is terrestrial functioning object.
- ▼ [COFFEE](#) has the density liquid density.
 - ▶ Every [liquid](#) has the density liquid density.
 - ▶ [COFFEE](#) is a [liquid](#).
- ▼ [The atmosphere](#) has the density gaseous density.
 - ▶ All [gas](#) has the density [gaseous density](#).
 - ▶ [The atmosphere](#) is [air](#).

CST-CoffeeInOpenCupAndIsOther

Query:
Is it false that dead birds are capable of [flying by flapping wings](#)?

Answer:
Yes

Because: [Detailed Justification](#) [Linear Justification](#)

Detailed Justification:

- ▼ Dead birds are not capable of [flying by flapping wings](#).
 - ▶ [Inanimate objects](#) cannot play the role of doers in animate activity.
 - ▶ Type Behavior Incapable and Type Capable Fn Behavior Capable cannot hold of the same thing at the same time.
- ▼ A dead bird is a type of [inanimate object](#).
 - ▶ A dead bird is a type of natural inanimate thing.
 - ▶ A natural inanimate thing is a type of [inanimate object](#).
- ▼ [Flying by flapping wings](#) is a type of animate activity.
 - ▶ An [action performed by animals](#) is a type of animate activity.
- ▼ [Flying by flapping wings](#) is a type of [action performed by animals](#).
 - ▶ [Flying by flapping wings](#) is a type of [locomotion](#).
 - ▶ [Locomotion](#) is a type of [animal locomotion](#).
 - ▶ [Animal locomotion](#) is a type of [action performed by animals](#).

CST-DeadBirdsDead

Query:
Is it true or false that [the Golden Gate Bridge](#) is a [portable object](#)?

Answer:
[False](#)

Because: [Detailed Justification](#) [Linear Justification](#)

Detailed Justification:

- ▼ #SentenceTruth holds of
 - [the Golden Gate Bridge](#) is a [portable object](#)
 - and [No](#).
- ▼ It is false that [the Golden Gate Bridge](#) is a [portable object](#).
 - ▼ No [suspension bridge](#) is a [portable object](#).
 - ▶ No [portable object](#) is a [site](#).
 - ▼ A [suspension bridge](#) is a type of [site](#).
 - ▶ A [suspension bridge](#) is a type of [fixed freestanding structure](#).
 - ▶ A [fixed freestanding structure](#) is a type of non-agent-like place.
 - ▶ A non-agent-like place is a type of [site](#).
- ▶ [The Golden Gate Bridge](#) is a [suspension bridge](#).

CST-IsTheGoldenGateBridgePortable

Query:
Is it true or false that [Argentina](#) is capable of [crying](#)?

Answer:
[False](#)

Because: [Detailed Justification](#) [Linear Justification](#)

Detailed Justification:

- ▼ #SentenceTruth holds of [Argentina](#) is capable of [crying](#) and [No](#).
 - ▼ [Argentina](#) is not capable of [crying](#).
 - ▶ It is false that
 - something is capable of playing a "some other thing" role in some other thing,
 - that thing is a type of thing,
 - and that type of thing cannot play the role of the other thing in that other thing.
 - ▶ [Argentina](#) is a [Latin American country](#).
 - ▼ [Latin American countries](#) cannot play the role of bodily doers in [crying](#).
 - ▶ Agentive artifacts cannot play the role of doers in [body movement](#).
 - ▶ If some living organism did some physical event non-deliberately, then that living organism did that physical event.
 - ▼ [Crying](#) is a type of [body movement](#).
 - ▶ [Crying](#) is a type of [making a facial expression](#).
 - ▶ [Making a facial expression](#) is a type of [body movement](#).
 - ▼ A [Latin American country](#) is a type of agentive artifact.
 - ▶ An [organization](#) is a type of agentive artifact.
 - ▶ A [Latin American country](#) is a type of [organization](#).
 - ▶ A [Latin American country](#) is a type of [independent country](#).
 - ▶ An [independent country](#) is a type of [country](#).
 - ▶ A [country](#) is a type of [political entity](#).
 - ▶ A [political entity](#) is a type of [organization](#).

CST-IsArgentinaCapableOfCrying

Staying true to that approach

- Rampant Inconsistency
- Astounding Inefficiency
- Lure of the dot-com era / AI “boom”
- Persevering through the AI “winter”
- Lure of the free lunch (statistical “learning” boom)

Approach: Teaching...

- Massive KB
- + Inference

Approach: Engineering

101

Representations of Knowledge

Each representation makes certain operations fast,
at the expense of other operations being slow

- E.g., natural language sentences => NL is sometimes an inefficient representation
- E.g., node & link diagrams
- E.g., n^{th} -order logic formulae
- E.g., high-resolution imagery
- E.g., database tuples
- E.g., algebraic equations
-
-

Five friends get together to play 5 doubles matches, with a different group of 4 players each time. The sums of the ages of the players for the different matches are 124, 128, 130, 136 and 142 years. What is the age of the youngest player ?

$$\begin{aligned} v+w+x+y &= 124 \\ v+w+x+z &= 128 \\ v+w+y+z &= 130 \\ v+x+y+z &= 136 \\ w+x+y+z &= 142 \end{aligned}$$

Closing the choke causes increasing bottomhole pressure.

```

    (cause-SlotType ClosingIndustrialChoke
      (IncreaseOnSlotFr ProductionTubingAndBottomholeSystem bottomholeFlowingPressure)) in OilGasIndustryM
  
```

Increasing bottomhole pressure causes decreasing drawdown.

```

    (cause-SlotType
      (IncreaseOnSlotFr
        (RestrictionOnSlotFr ProductionTubingAndBottomholeSystem bottomholeFlowingPressure))
      (DecreaseOnSlotFr
        (RestrictionOnSlotFr WellInflow inflowRate))) in FluidMechanicsM
  
```

In applications of the PI formula the value of inflow rate for well inflow is a dependent variable.

```

    (dependentVariableFormula ProductivityIndex-InflowFormula inflowRate WellInflow) in FluidMechanicsM
  
```

In applications of the PI formula the value of bottomhole flowing pressure for production tubing and bottomhole system is an independent variable.

```

    (independentVariableFormula ProductivityIndex-InflowFormula bottomholeFlowingPressure ProductionTubingAndBottomholeSystem) in FluidMechanicsM
  
```

According to the PI formula, values of bottomhole flowing pressure for production tubing and bottomhole systems vary inversely with values rate for well inflows.

```

    (inverselyRelatedSlotsForFormula ProductivityIndex-InflowFormula bottomholeFlowingPressure ProductionTubingAndBottomholeSystem inflowRate WellInflow) in FluidMechanicsM
  
```

The flow pressure values in the PI formula pertain to production tubing and bottomhole systems.

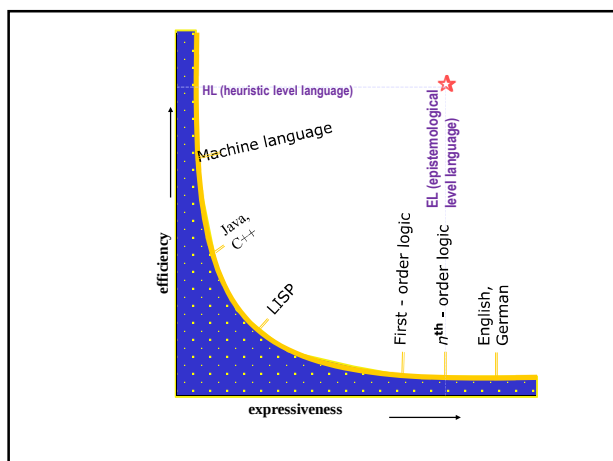
```

    (useConstraintForQuantityBasedFormulaSlot ProductivityIndex-InflowFormula
      (The FlowPressure) ProductionTubingAndBottomholeSystem) in FluidMechanicsM
  
```

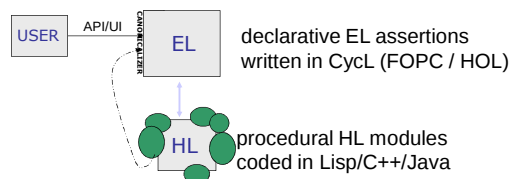
The speed of flow of the well inflow values in the PI formula pertain to well inflows.

```

    (useConstraintForQuantityBasedFormulaSlot ProductivityIndex-InflowFormula
      (The WellInflowRate)) in FluidMechanicsM
  
```



For the **EL**, use as expressive a language as is called for.



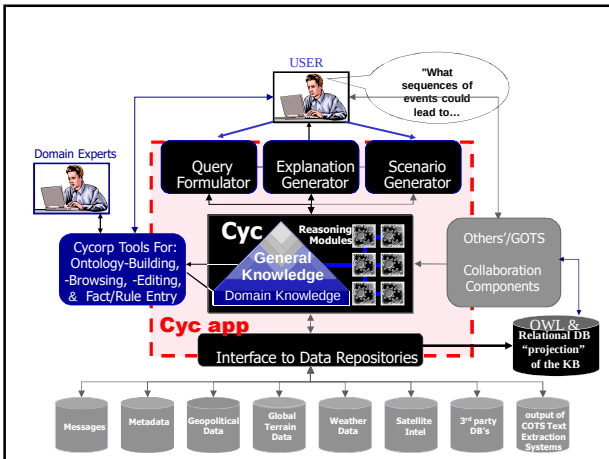
For the **HL**, use a grab-bag of over 1000 special-purpose inference procedures (algorithms) and data structures.

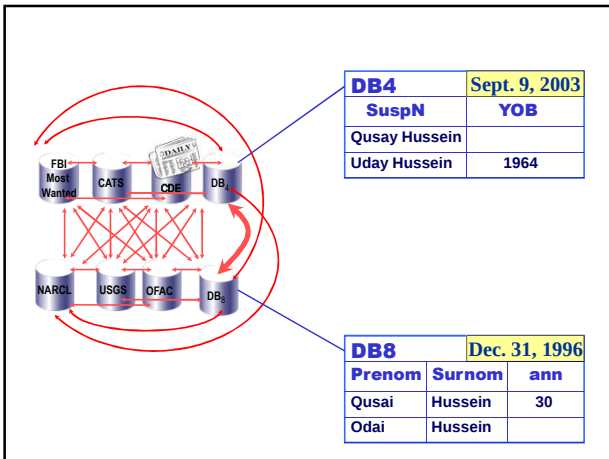
Why “AI” is Only Half the Answer

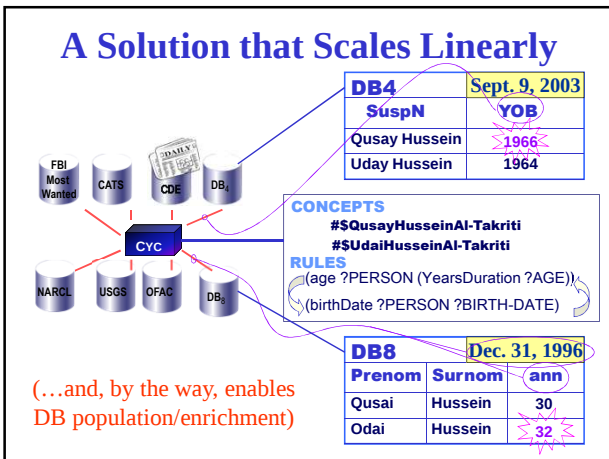
- ☒ There are two AI power sources (statistics, logic).
- ☒ Everyone else is focusing on right-brain AI (brittle)
- ☒ One exception: CYC (Manhattan-Project-like effort)
- ☐ **Now that it's more or less done, building a Cyc app. means extending Cyc's ontology, KB,... slightly**
- ☐ The path to “real AI”: statistics+logic synergy
 - A couple quick demo's of Cyc running.
 - Examples drawn from petroleum engineering Cyc app's.

Applying Cyc

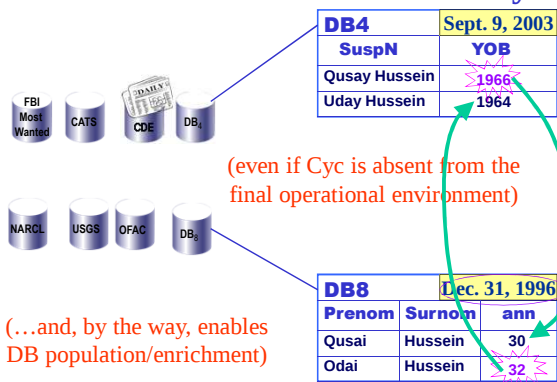
- Cyc is an energy source, not a single application.
Like oil, electricity, telephony, computers, data...
Cyc can spawn and sustain a **knowledge utility industry**.
- It can cost-effectively underlie almost all apps.
(Provide a common-sense layer to reduce brittleness when faced with unexpected inputs/situations)
- To apply Cyc, we **extend** its ontology, its KB, and possibly its suite of specialized reasoning modules







A Solution that Scales Linearly



A more recent example

“What major US cities are particularly vulnerable to an anthrax attack?”

The answer is logically implied by data dispersed through several sources:



“What major US cities are particularly vulnerable to an anthrax attack?”

“major US city” ⇔ ?C is a U.S. City with >1M population

“particularly vulnerable to an anthrax attack” ⇔

- the current ambient temperature at ?C is above freezing, and
- ?C has more than 100 people for each hospital bed, and
- the number of anthrax host animals near ?C exceeds 100k

“What major US cities are particularly vulnerable to an anthrax attack?”

U.S. cities with population > 1 million



1-2 conjuncts in a Cycl “Ask” expression

(and

(isa ?C USCity)

(> (NumberOfInhabitantsFn ?C) 10⁶)

(vulnerableToScriptedEvent TypeUsing

?C

DeployingABioAgentByInfectingAZoonoticHost

Anthrax-Bacterium))

U.S. cities with population > 1 million



1-2 conjuncts in a Cycl “Ask” expression

state	name	type	county	state_fips
TX	Dallas	pp1	Dallas	48
MN	Hennepin County	civil	Hennepin	27
CA	Sacramento County	civil	Sacramento	6
AZ	Phoenix	pp1	Maricopa	4

primary_lat	primary_long	elevation	population	status
32.78333	-96.8	463	1022830	BGN 1978 1959
45.01667	-93.45	0	1032431	
38.46667	-121.31667	0	1041219	
33.44833	-112.07333	1072	1048949	BGN 1931 1900 1897



The Geographic Names Information System (GNIS)
DB maintained by the US Geological Survey (USGS).

So how do we explain to our system that:

- row 1 of that table is “about” the city of Dallas, TX
- the **population** field of that table contains the number of inhabitants of the city that that row is “about”
- here is exactly how to access tuples of that database
- that access will be fast, accurate, recent, **complete**



- the **population** field of that table contains the number of inhabitants of the city that that row is “about”

We provide the field encodings and decodings, some of which correspond to explicit fields like population, two-letter state codes, etc:

```
(fieldDecoding Usgs-Gnis-LS ?x
 (TheFieldCalled "population")
 (numberOfInhabitants
  (TheReferentOfTheRow Usgs-Gnis)
  ?x))
```



The Geographic Names Information System (GNIS)
DB maintained by the *US Geological Survey* (USGS).

- how to access tuples of that database

We provide all the information needed for a JDBC connection script:

We assert, in the context (MappingMtfN Usgs-KS), all of these:

```
(passwordForSKS Usgs-KS "geografy")
(portNumberForSKS Usgs-KS 4032)
(serverOfSKS Usgs-KS "sksi.cyc.com")
(sqlProgramForSKS Usgs-KS PostgreSQL)
(structuredKnowledgeSourceName Usgs-KS "usgs")
(subProtocolForSKS Usgs-KS "postgresql")
(userNameForSKS "sksi")
```



The Geographic Names Information System (GNIS)
DB maintained by the *US Geological Survey* (USGS).

- that access will be fast, accurate, recent, **complete**

We provide meta-level assertions about the database, about each table of the database, about the completeness etc. of various kinds of data in the DB, etc.

We assert, in the context (MappingMtFn Usqs-KS):

```
(schemaCompleteExtentKnownForValueTypeInArg
  Usqs-Gnis-LS
  USCity
  numberOfInhabitants
  1)
```



The Geographic Names Information System (GNIS)
DB maintained by the *US Geological Survey* (USGS).

- that access will be fast, accurate, recent, **complete**

We provide meta-level assertions about the database, about each table of the database, about the completeness etc. of various kinds of data in the DB, etc.

We assert, in the context (MappingMtFn Usqs-KS):

```
(resultSetCardinality Usqs-Gnis-PS
  (TheSet (PhysicalFieldFn Usqs-Gnis-PS "state"))) TheEmptySet
  60.0)
```

```
(resultSetCardinality Usqs-Gnis-PS
  (TheSet
    (PhysicalFieldFn Usqs-Gnis-PS "primary_long")
    (PhysicalFieldFn Usqs-Gnis-PS "primary_lat")
    (PhysicalFieldFn Usqs-Gnis-PS "name")))
  (TheSet
    (PhysicalFieldFn Usqs-Gnis-PS "county")
    (PhysicalFieldFn Usqs-Gnis-PS "state")))
  530.36)
```

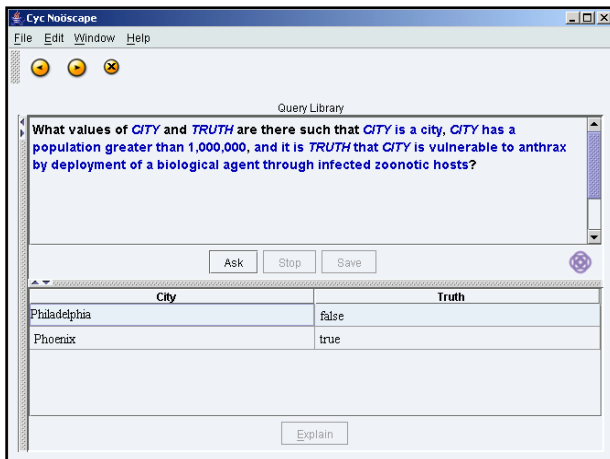
“What major US cities are particularly vulnerable to an anthrax attack?”

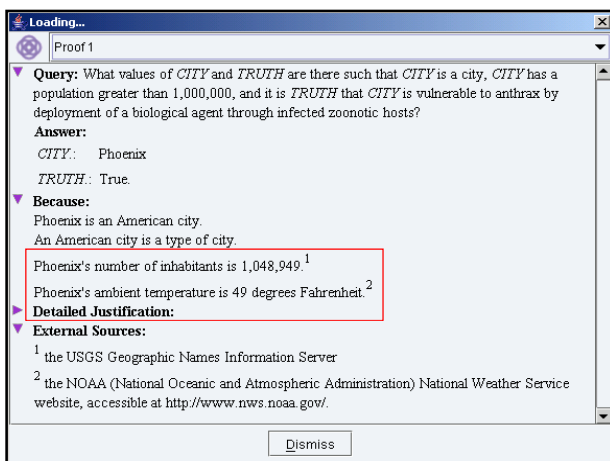
“major US city” ⇔ U.S. City with >1M population

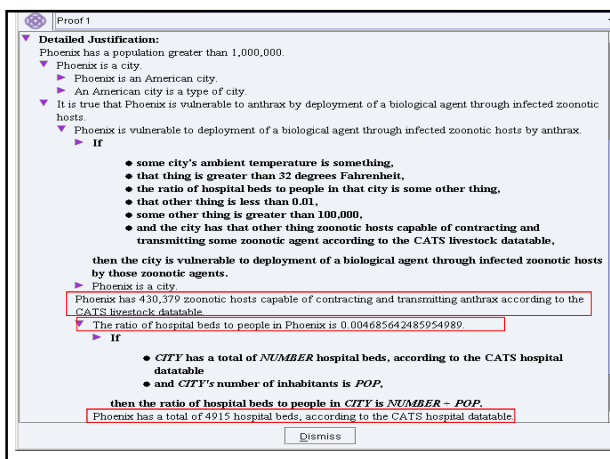
“particularly vulnerable to an anthrax attack” ⇔

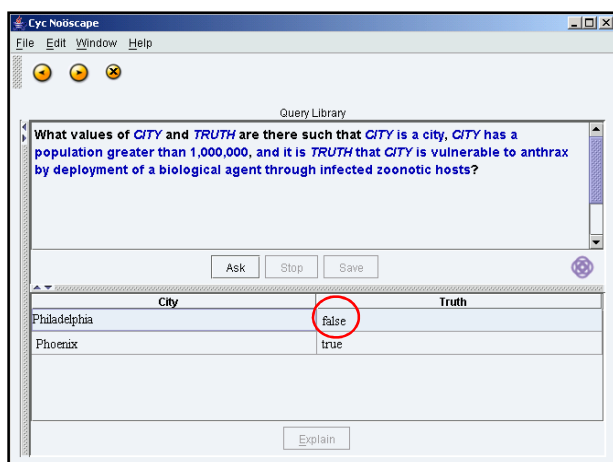
- the current ambient temperature at ?C is above freezing,
and
- ?C has more than 100 people for each hospital bed,
and
- the number of anthrax host animals near ?C exceeds 100k

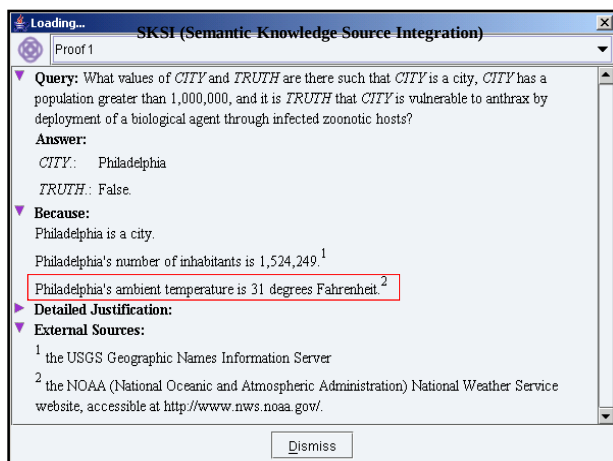
**Cyc knows that pullets are chickens, so
don't add those two numbers together!**

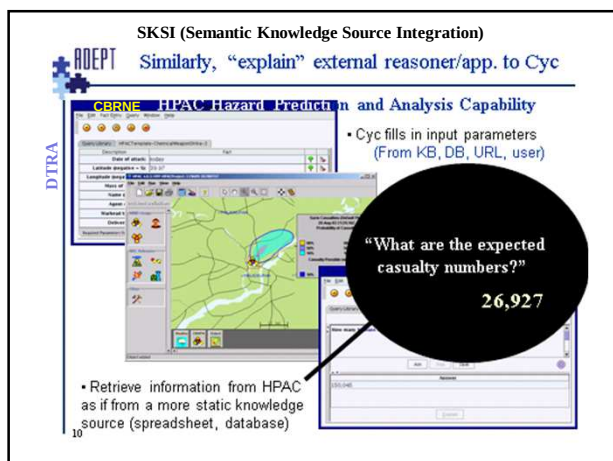


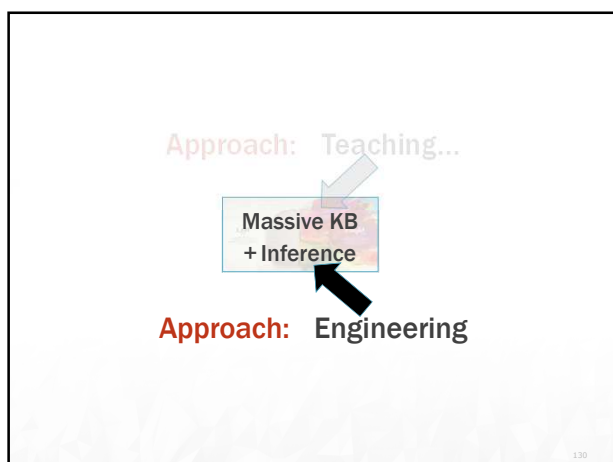












There is no one correct monolithic ontology.

E.g., Cyc's 21M axioms are divided into thousands of *contexts* by: granularity, topic, culture, geospatial place, time,...

There is a correct monolithic reasoning mechanism, but it is so deadly slow that we never call on it unless we have to

E.g., the Cyc inference engine is a community of 1100 "agents" that attack every problem and, recursively, every subproblem (subgoal). One of these 1100 is a general theorem prover; the others have special-purpose data structures/algorithms to handle the most important, most common cases, very fast.

There is no one correct monolithic ontology.

E.g., Cyc's 21M axioms are divided into thousands of *contexts* by: granularity, topic, culture, geospatial place, time,...

There is a correct monolithic reasoning mechanism, but it is so deadly slow that we never call on it unless we have to

E.g., the Cyc inference engine is a community of 1100 "agents" that attack every problem and, recursively, every subproblem (subgoal). One of these 1100 is a general theorem prover; the others have special-purpose data structures/algorithms to handle the most important, most common cases, very fast.

Even though they are expressed in formal logic, most axioms state *usuals*, not absolute truths.



Nonmonotonic (later information can show that something you earlier believed is false after all).

So the reasoning is **default**.

(Argumentation:
Gather up all the pro- and con-arguments, and compare them)

[illegible]

What factors argue <for/against> the conclusion that <ETA> <performed> <the March 2004 Madrid attacks>?

Contextualized questions/reasoning via argumentation.

For:

- ETA often executes attacks near national election
- ETA has performed multi-target coordinated attacks
- Over the past 30 years, ETA performed 75% of all terrorist attacks in Spain
- Over the past 30 years, 98% of all terrorist attacks in Spain were performed by Spain-based groups, and ETA is a Spain-based group.

Against:

- ETA warns (a few minutes ahead of time) of attacks that would result in a high number civilian casualties, to prevent them. There was no such warning prior to this attack.
- ETA generally takes responsibility for its attacks, and it did not do so this time.
- ETA has never been known to falsely deny responsibility for an attack, and it did deny responsibility for this attack.

Engineering Challenges

Abductive Reasoning

Terms related to Context space

Microtheories ("Mts") - i.e., Contexts

Default Reasoning with Exceptions

Call outs to code

Contextualized questions/reasoning via argumentation

Truth Maintenance

Bidirectional Inference

Robust HL Module Support

Resource-limited heuristic search

Ubiquitous Meta-reasoning and Meta-meta-reasoning

Meta-reasoning and reflection on KB content

Learn terms, rules, pragma... from Experience

Temporal/Date Reasoning

Enumerable Function/Predicate Support

Representation, not Programming

Eager vs Lazy Reasoning

Justified Answers (Argumentation, Preference)

Heuristic Level (HL)

Deep Collection Hierarchy

Separate Epistemological Problem / Heuristic Problem

Cardinality

Units of Measure Reasoning

Inference Control Heuristics

Argument Construct Vocabulary

Strong Negation

Default Reasoning

Inference Strategist

Embrace Incompleteness

Adequately Expressive Language
(1st order, 2nd order, reflection, modal...)

Explicit Identity/Difference Reasoning

Completeness Vocabulary

Domain-independent Individuals

Microtheories (Contexts as first class objects)

Dynamic Language Extensibility

Epistemological Level (EL)

Separate Pragmatics from Logic

Relation Hierarchy

Inference Tactician

Special Purpose Reasoners

Pragmatic Constraints

Robust Support for Global Properties

Structured Knowledge Source Integration

Plausible Scenario Generation

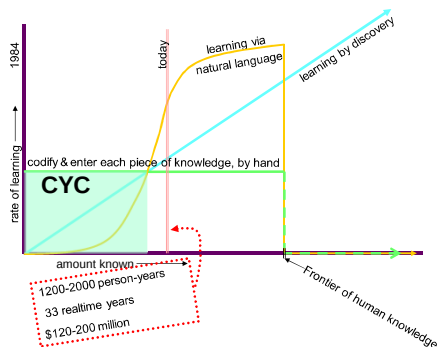
Why Bother Doing All That? i.e., why not “live with” ML?

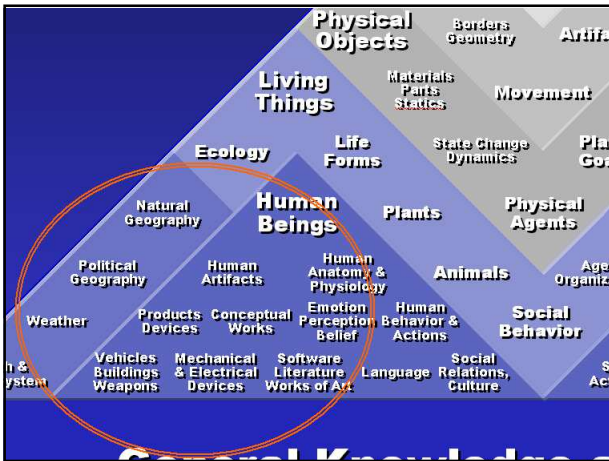
- Sometimes you want to see the **explanation** (or even: pro/con arguments for top **n** answers)
- Sometimes you care about rare conditions (where there is little or no training data)
- Statistics can be inhumanly **brittle**
- Multiple disagreeing experts; different heuristics apply in different eras, places,...

Each assertion should be situated in a context: in a region of context-*space*

- Anthropacity
 - Time
 - GeoLocation
 - TypeOfPlace
 - TypeOfTime
 - Culture
 - Sophistication/Security
 - Topic
 - Granularity
 - Modality/Disposition /Epistemology
 - Argument-Preference
 - Justification
- We identified 12 dimensions of mt-space
 - We developed a vocabulary of predicates and terms to describe points and regions along each of those 12 dimensions; and
 - We have been situating assertions more and more precisely, and we have been working out calculi for inferring contexts
 - E.g., if P is true in C1, and $P \Rightarrow Q$ is true in C2, in what context C2 can Q be validly concluded?

Building Cyc qua Engineering Task





Temporal Relations

37 Relations Between Temporal Things

- temporalBoundsIntersect
- temporallyIntersects
- startsAfterStartingOf
- endsAfterEndingOf
- startingDate
- temporallyContains
- temporallyCooriginating
- temporalBoundsContain
- temporalBoundsIdentical
- startsDuring
- overlapsStart
- startingPoint
- simultaneousWith
- after

Temporal Relations

“Ariel Sharon was in Jerusalem during 2005 with granularity calendar-week”



“Condoleezza Rice made a ten-day trip to Jerusalem in February of 2005”



Both of them were in Jerusalem during February 2005

diplomacy reasoning

They had a meeting, then (even if it didn't make the news.)

Relations Between an Event and its Participants

- | | |
|-----------------------|-------------------------|
| • performedBy | • fromLocation |
| • causes-EventEvent | • toLocation |
| • objectPlaced | • deviceUsed |
| • objectOfStateChange | • driverActor |
| • outputsCreated | • damages |
| • inputsDestroyed | • vehicle |
| • assistingAgent | • providerOfMotiveForce |
| • beneficiary | • transportees |

Over 400 more.

In In Our Geospatial Ontology

- We started in 1984 with just one binary predicate, “in”.
- in(X,Y) means the inner object X is spatially located in the region defined by the outer object Y.
- If I just tell you in(X,Y), and you aren’t told what X and Y are, then you (and Cyc) can’t answer questions like these:
 - From the outside of Y, can I see any part of X?
 - If I turn Y over and shake it, will X fall out?
 - Is there room to put more things in Y?
 - Is X actually a part of Y?
- Such failures led to our introducing new, more precise, more specialized versions of “in”. By now there are over 75 such predicates, organized in a graphical taxonomy.

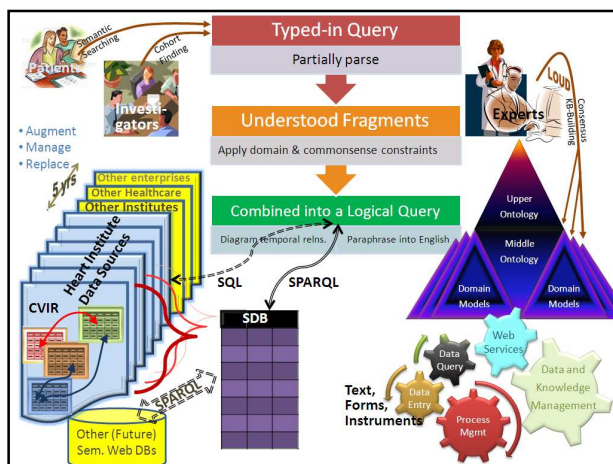
Propositional Attitudes

Relations Between Agents and Propositions

- | | |
|------------|-----------------|
| • goals | • opinesThat |
| • intends | • knowsThat |
| • desires | • remembersThat |
| • hopes | • perceivesThat |
| • expects | • seesThat |
| • believes | • fearsThat |

Applying Cyc

- Cyc is an energy source, not a single application.
Like oil, electricity, telephony, computers, data...
Cyc can spawn and sustain a **knowledge utility industry**.
- It can cost-effectively underlie almost all apps.
(Provide a common-sense layer to reduce brittleness when faced with unexpected inputs/situations)
- To apply Cyc, we **extend** its ontology, its KB, and possibly its suite of specialized reasoning modules



Cyc-powered production understanding

- Generate hypotheses about wells
- Much earlier alerts
- Fewer false alarms (and missed COI's)

Phase 1

- Generate hypotheses about SAGD wells
- Stable, accurate and correct for several months
- Used daily by engineers
- Reported COIs in real time, some of which were acted on to optimize well performance
- Required minimal tuning
- Proved Cyc's unique capabilities by deducing potential consequences and remediations

Phase 1: COIs

Conditions of Interest

Reach Saturation Steam Temperature in Liner

Temperature Increase in Liner

Depth/Temperature Gradient Change

Compute Liner Inflow During Shut-in

Temperature Rise After Shut-in

Potential Casing Integrity Fault

Thermal Cycling Count

Compilation of Shut-in Trends

Fiber Calibration Fault

SAGD Well Shut-in

SAGD Well Start-up

Phase 2

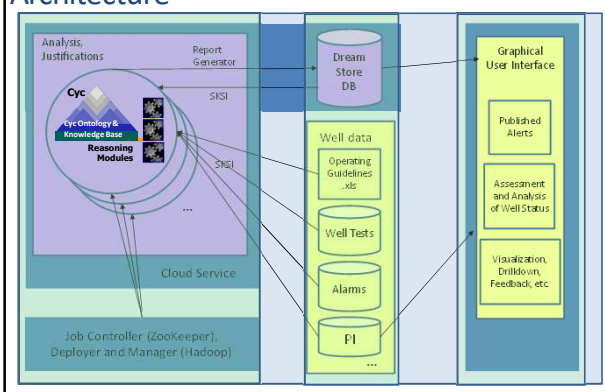
- Expand reasoning to pro and con hypothesis generation
- Auto-calibrate to individual well sensor performance
- Robust to limited sensor availability
- Focus on high-value events:
 - liquid loading - as a precursor to slugging, killed well, etc., identifying potential liquid loading and taking action might avoid a potential deferment scenario
 - Well beginning to make sand
 - ramp-up in violation of operating guidelines
 - Too slow: opportunity to prevent deferment
 - Too fast: opportunity to prevent premature well failure
 - well not returned to appropriate stable flow, as per operating guidelines and previous stable flow values
 - Well in "no-flow" when the operators think it is flowing

Phase 2: COIs

Conditions of Interest

Sand production
 Ramp-up Operating Guidelines departures
 Slugging
 Deposit in production tubing restricting flow
 Deposit in flowline or wellhead restricting flow
 Pressure increase downstream of choke
 Pressure decrease downstream of choke
 Well not flowing but not shut in
 Shutting in with wing or master valve
 Ramping up with wing or master valve
 Background events: Stable flow, Ramping up, Shutting in, Choke back, Bean-up

Architecture



SAGD systems

(eventTypeUsesDeviceType SteamAssistedGravityDrainage (SystemForFn SteamAssistedGravityDrainage))
"SAGD involves using SAGD systems."

(relationAllExistsCount parts (SystemForFn SteamAssistedGravityDrainage) SAGDInjectionWell 1)
"Every SAGD system has exactly one SAGD injection well as a part."

(relationAllExistsCount parts (SystemForFn SteamAssistedGravityDrainage) SAGDProductionWell 1)
"Every SAGD system has exactly one SAGD production well as a part."

(relationExistsCountInstance objectFoundInLocation (SystemForFn SteamAssistedGravityDrainage) Asset01 52)
"Asset 1 has exactly 52 SAGD systems."

Wellbores

(outputsCreated-TypeType WellDrilling Wellbore)
"Well drilling creates wellbores."

(relationAllExists outputsCreated DirectionalDrilling NonVerticalWellbore)
"Directional drilling creates non-vertical wellbores."

(relationAllExists properPhysicalParts HorizontalWell HorizontalWellbore)
"Horizontal wells have horizontal wellbores as parts."

Excavation-Hole
 InertArtifact
 PathThroughSolid
 Underground
 Wellbore
 NonVerticalWellbore
 HorizontalWellbore
 SAGDLowerHorizontalWellbore
 SAGDUpperHorizontalWellbore
 VerticalWellbore

Multiphase Fluid Flow

FluidFlow-Translation

MultiphaseFluidFlow-Translation

MultiphaseFluidFlow-AnnularFlow

AnnularFlow-Horizontal

AnnularFlow-VerticalUpward

MultiphaseFluidFlow-BubbleFlow

BubbleFlow-Horizontal

BubbleFlow-VerticalUpward

MultiphaseFluidFlow-ChurnOrStratifiedFlow

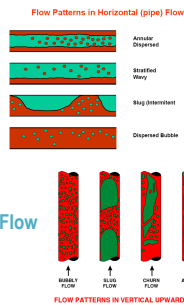
ChurnFlow-VerticalUpward

StratifiedFlow-Horizontal

MultiphaseFluidFlow-SlugFlow

SlugFlow-Horizontal

SlugFlow-VerticalUpward



Multiphase Fluid Flow

(implies
 (and
 (isa ?FLOW MultiphaseFluidFlow-Translation)
 (primaryObjectMoving ?FLOW ?STUFF)
 (hasConcentrationOfGasInFluidByVolume ?STUFF
 (VeryLowAmountFn ConcentrationOfGasInFluidByVolume)))
 (isa ?FLOW MultiphaseFluidFlow-BubbleFlow))

"If the fluid moving in a multiphase fluid flow has a very low concentration of gas by volume, it is bubble flow."

(implies
 (and
 (isa ?FLOW MultiphaseFluidFlow-Translation)
 (primaryObjectMoving ?FLOW ?STUFF)
 (hasConcentrationOfGasInFluidByVolume ?STUFF
 (VeryHighAmountFn ConcentrationOfGasInFluidByVolume)))
 (isa ?FLOW MultiphaseFluidFlow-AnnularFlow))

"If the fluid moving in a multiphase fluid flow has a very high concentration of gas by volume, it is annular flow."

Simple Productivity Index

(equationFormFor ProductivityIndex-InflowFormula
(numericallyEquals
(The ProductivityIndexValue)
(QuotientFn
(RateOfFlowFn
(The WellInflow))
(DifferenceFn
(The AverageReservoirPressure)
(The FlowPressure))))))

$$PI = \frac{Q}{P - P_{wf}} = \frac{Q}{\Delta P_{DD}}$$

Steady-state multiphase inflow with skin

(equationFormFor SteadyStateMultiphaseInflowWithSkinFormula
(numericallyEquals
(DifferenceFn
(The AverageReservoirPressure) (The FlowPressure))
(TimesFn
(QuotientFn
(TimesFn 141.2
(The FlowRate)
(The Viscosity)
(OilFormationVolumeFactorFn
(The NaturalReservoir)))
(TimesFn
(The Permeability) (The Height))
(DifferenceFn
(LogFn
(QuotientFn
(RadiusFn
(ReservoirOfWellFn
(The Well)))
(RadiusFn
(The Well))))
(PlusFn 0.5 (SkinFactorFn (The Well))))))

$$\bar{p}_R - p_{wf} = \frac{141.2 q B \mu}{k h} \left(\ln \frac{r_e}{r_w} - \frac{1}{2} + S \right)$$

Multiphase vertical outflow

(equationFormFor MultiphaseOilVerticalOutflowFormula
(numericallyEquals
(The
(RateOfChangeRelativeToRateOfChangeFn Pressure Height))
(PlusFn
(TimesFn
(DensityFn
(The MultiphaseHydrocarbonFluid))
(VectorMagnitudeFn
(The GravitationalForceVector))
(CosineFn
(InclinationFn
(The ProductionTubing))))
(QuotientFn
(TimesFn
(FrictionFactorOfFluidFlowFn
(The FluidFlowEvent))
(DensityFn
(The MultiphaseHydrocarbonFluid))
(SpeedFn
(The MultiphaseHydrocarbonFluid))
(TimesFn 2
(DiameterOfCylinderFn
(The ProductionTubing))))
(TimesFn
(DensityFn
(The MultiphaseHydrocarbonFluid))
(SpeedFn
(The MultiphaseHydrocarbonFluid))
(The
(RateOfChangeRelativeToRateOfChangeFn
(SpeedFn
(The MultiphaseHydrocarbonFluid))
(HeightFn
(The ProductionTubing))))))

VERTICAL FLOW

GENERAL SOLUTION :

$$\left(\frac{dp}{dh} \right)_{TOTAL} = \left(\frac{dp}{dh} \right)_{ELEVATION} + \left(\frac{dp}{dh} \right)_{FRICTION} + \left(\frac{dp}{dh} \right)_{ACC}$$

$$= \rho g \cos \theta + \frac{f \rho v^2}{2d} + \rho v \frac{dv}{dh}$$

TYPICALLY IN OIL WELLS :

> 90% < 10% NEGLIGIBLE EXCEPT FOR
MIST FLOW AT LOW ρ
AND HIGH VELOCITY

TO SOLVE EQUATION, NEED TO DETERMINE

$P > 1$

FOR 2-PHASE FLOW !

Oil Formation Volume Factor

(implies

(and

(isa ?RESERVOIR NaturalReservoir)

(reservoirPayload ?RESERVOIR ?HYDROCARBON)

(aboveGroundRegion ?RESERVOIR ?SURFACE)

(uniquePartOfType ?HYDROCARBON

(LiquidFn Hydrocarbon) ?OIL)

(objectHasSlotValueAtConditions ?OIL volumeOfObject ?R-VOL ?RESERVOIR)

(ratioOfTo ?R-VOL ?S-VOL ?RATIO)

(objectHasSlotValueAtConditions ?OIL volumeOfObject ?S-VOL ?SURFACE))

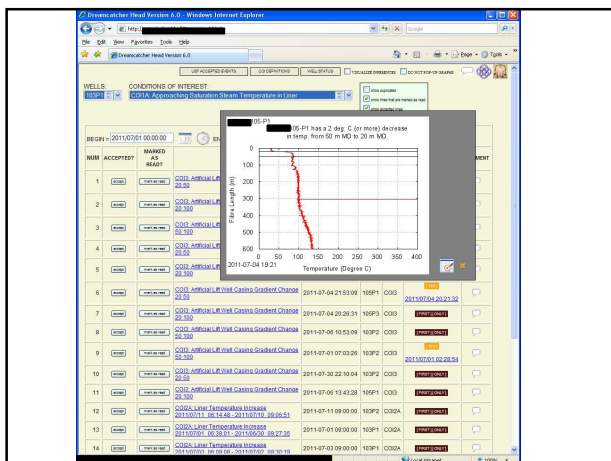
(oilFormationVolumeFactor ?RESERVOIR ?RATIO))

"Bo is equal to the ratio of R-VOL to S-VOL, where:

R-VOL is the volume at reservoir conditions of the liquid hydrocarbon in the reservoir payload, and

S-VOL is the volume at surface conditions of the liquid hydrocarbon in the reservoir payload."

WELL	ACCEPTED	MARKED AS READ?	DESCRIPTION	DATE/TIME	WELL	CO2	PREVIOUSLY REPORTED	COMMENT
1			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-04 20:21:52	10SP1	CO2	None	(3 repeats)
2			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-07 08:19:19	10SP2	CO2	None	(7 repeats)
3			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-01 02:28:54	10SP2	CO2	None	(8 repeats)
4			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-09 14:32:09	10SP1	CO2	None	
5			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-07 14:00:04	10SP3	CO2	20110707 06:18:18	None
6			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-04 21:53:09	10SP1	CO2	20110704 20:21:52	None
7			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-04 20:26:31	10SP3	CO2	Previously Reported	
8			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-06 10:53:09	10SP2	CO2	Previously Reported	
9			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-01 07:03:20	10SP2	CO2	20110701 06:28:54	
10			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-30 22:10:04	10SP2	CO2	Previously Reported	
11			CO2 Artificial Lift Well Casing Gradient Change 20.100	2011-07-08 13:43:28	10SP1	CO2	Previously Reported	
12			CO2 Line Temperature Increase 20110701 06:18:18 - 20110707 06:18:18	2011-07-11 09:00:00	10SP2	CO2A	Previously Reported	
13			CO2 Line Temperature Increase 20110701 06:18:18 - 20110707 06:18:18	2011-07-01 09:00:00	10SP1	CO2A	Previously Reported	
14			CO2 Line Temperature Increase 20110701 06:18:18 - 20110707 06:18:18	2011-07-03 09:00:00	10SP1	CO2A	Previously Reported	



Standard Operating Procedure: SOP 1: Impending Steam Breakthrough

Time: 11/07/2010 01:47:23

Use Live Data: ☒

[Find Events](#)

IMPENDING STEAM BREAKTHROUGH [Explain](#) [View Graph](#)

An impending steam breakthrough has been detected in 104-P2 at 785 m at 01:47:23, November 7, 2010. To remedy the situation, reduce or halt steam injection or reduce drawdown from 104-P2.

[Potential Consequences](#)

[Recommended Actions](#)

Query:
What impending steam breakthroughs occur at 104-P2 according to the Dreamcatcher SOP 1 criteria?

Answer:
an impending steam breakthrough based on the Dreamcatcher SOP 1 criteria

Detailed Justification:

- An impending steam breakthrough occurs at 104-P2 at 01:47:23, November 7, 2010 according to the Dreamcatcher SOP 1 criteria.
 - Specific DTS readings from 01:47:23, November 7, 2010 indicate reaching saturation steam temperature in the liner is occurring in 104-P2.
 - In thermally enhanced oil recovery using steam, impending steam breakthroughs cause reaching saturation steam temperature in the liner.

External Sources:

Detailed Justification:

- An impending steam breakthrough occurs at 104-P2 at 01:47:23, November 7, 2010 according to the Dreamcatcher SOP 1 criteria.
- Specific DTS readings from 01:47:23, November 7, 2010 indicate reaching saturation steam temperature in the liner is occurring in 104-P2.

If at any given depth within the monitored duration the moving average DTS temperature in the liner is within the specified threshold value of the moving average of the saturation steam temperature, and the DTS curve is trending toward a relatively stable saturation steam temperature, the SAGD production engineers consider there to be a possible impending steam breakthrough.

The monitored duration for Dreamcatcher SOP 1 is 14 days.

The threshold for approaching the saturation steam temperature for Dreamcatcher SOP 1 is 216.18 degrees Celsius.
- On 01:47:23, November 7, 2010 104-P2 achieves a temperature of 211.71 degrees Celsius at a depth of 785 m which is within 4.47 degrees Celsius of the steam saturation temperature of 216.18 degrees Celsius.

216.18 degrees Celsius is the average steam saturation temperature at 104-P2 between 01:47:23, November 2, 2010 and 01:47:23, November 7, 2010.

211.71 degrees Celsius is the average temperature at 785 m in 104-P2 between 01:47:23, November 2, 2010 and 01:47:23, November 7, 2010.

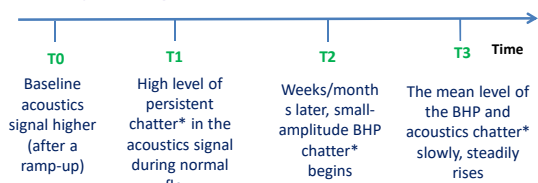
Detailed Justification ☐ Debugging? ☐ Show Cyl

- An impending steam breakthrough occurs at 04-P2 at 01:47:23, November 7, 2010 according to the Dreamcatcher SOP 1 criteria.
- Specific DTS readings from 01:47:23, November 7, 2010 indicate reaching saturation steam temperature in the liner is occurring in 04-P2.
- In thermally enhanced oil recovery using steam, impending steam breakthroughs cause reaching saturation steam temperature in the liner.
- In thermally enhanced oil recovery using steam, impending steam breakthroughs cause heating of the slotted liner.
- In thermally enhanced oil recovery using steam, attenuation of the barrier between the injected steam and the production liner causes steam getting closer to the producer.
- Water vapor and the production well are intended to be separate during thermally enhanced oil recovery using steam.
- Attenuation of the barrier between the injected steam and the production liner is the reduction of the barrier between water vapor and the production well during thermally enhanced oil recovery using steam.
- Every steam getting closer to the producer is characterized by movement toward some production well.
- Some water vapor is moving during every steam getting closer to the producer.
- In thermally enhanced oil recovery using steam, steam getting closer to the producer causes heating of the slotted liner.
- Attenuation of the barrier between the injected steam and the production liner is an essential part of impending steam breakthroughs.
- In thermally enhanced oil recovery using steam, heating of the slotted liner causes reaching saturation steam temperature in the liner.

Dreamcatcher Event List ☐ Debugging? ☐ Show Cyl

- Controlling the sub-cool level reduction can help prevent steam breakthrough.
- Diminished sub-cool level reduction inhibits rapidly developing steam breakthrough.
- Closing the choke controls the sub-cool level reduction.
- Closing the choke causes increasing bottomhole pressure.
- Increasing bottomhole pressure causes decreasing drawdown.
- In applications of the PI formula the value of inflow rate for well inflow is a dependent variable.
- According to the PI formula, values of bottomhole flowing pressure for production tubing and bottomhole system is an independent variable.
- The flow pressure values in the PI formula pertain to production tubing and bottomhole systems.
- The flow of flow of the well inflow values in the PI formula pertain to well inflows.

How Cyl recognizes Sand Production Problems

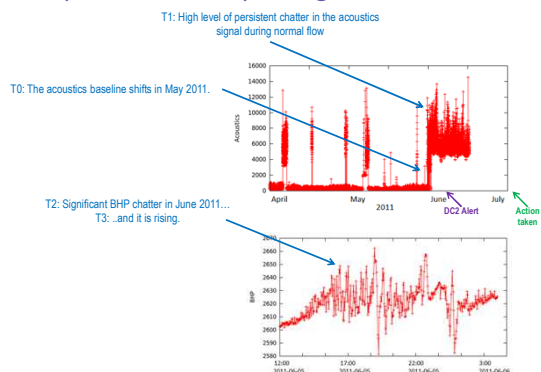


Why is it hard to recognize this early in the process?

Data supporting the above pattern is visible only during quiescent flowing periods during these months, i.e., periods which are punctuated by relatively massive sensor data fluctuations which mask the slow, steady rise in BHP chatter levels

*Oscillating amplitudes at high frequency

Example: Sand Buildup Recognition



Cyc can explain its reasoning in terms of . . .

. . . observed evidence, for example:

- When the well is flowing, look for the series of events and gradual changes described in the last slide
- Add in, as semi-independent supports to this argument
 - An argument *against* slugging: e.g. the oscillations in BHP happen with greater frequency than they would in slugging events.
 - Recent sand problem history.
 - Short-term THT increase, during T2, with no cotemporal choke change

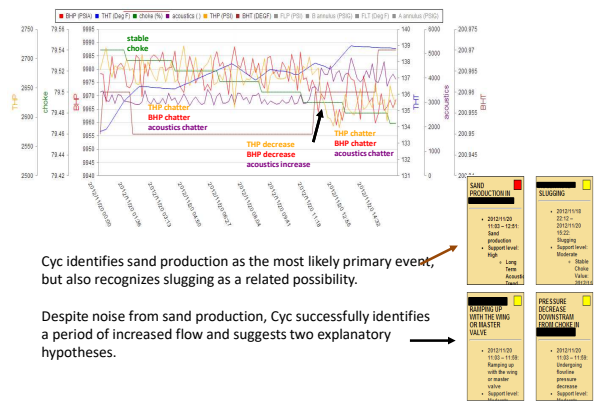
. . . first principles, for example:

- THT may increase due to increased friction from particulate matter passing through the well-head
- If erosion is left untreated, the rate of erosion may increase (proportional to the surface area of the sandface)

Cyc handles complex and controversial data:

- Recognize **subtle tell-tale signs of significant events** prior to their developing to the point at which they would trigger a high-likelihood alert (using long-term well history to avoid false alarms which traditional surveillance systems would otherwise generate if such early detection was a requirement)
- Recognize situations in which **many things are happening at once** and their "signals" superpose and would confuse traditional surveillance systems tuned to look for single-fault interpretations
- Recognize **situations where different experts disagree** about what is really happening, and present alternative detailed NL explanations for each position, identifying their respective advocates, automatically learning over time which judgmental rules to trust in different types of circumstances
- Rule out some events that would otherwise be "clearly recognized" by traditional surveillance systems where DC2 is able to reason from first principles, etc. that such events simply **aren't physically possible or plausible given this well's history** (especially in cases where ignoring one single suspicious signal enables this to be done)

Cyc handles complex and controversial data:



Why “AI” is Only Half the Answer

- ✗ There are two AI power sources (statistics, logic).
- ✗ Everyone else is focusing on right-brain AI (brittle)
- ✗ One exception: CYC (Manhattan-Project-like effort)
- ✗ Now that it's more or less done, building a Cyc app. means extending Cyc's ontology, KB,... slightly
- ❑ **The path to “real AI”: statistics+logic synergy**
 - A couple quick demo's of Cyc running.
 - Examples drawn from petroleum engineering Cyc app's.

Why Data Analytics is Only Half the Answer

- ☒ What's the question?
- ☒ Different Sources of ^{Power} ~~Energy~~
- ☒ Infrastructure to acquire/refine/distribute/harness
- ☒ Current state of the art
- ☐ Synergy between multiple sources of power

20 Motivating Applications (1984)

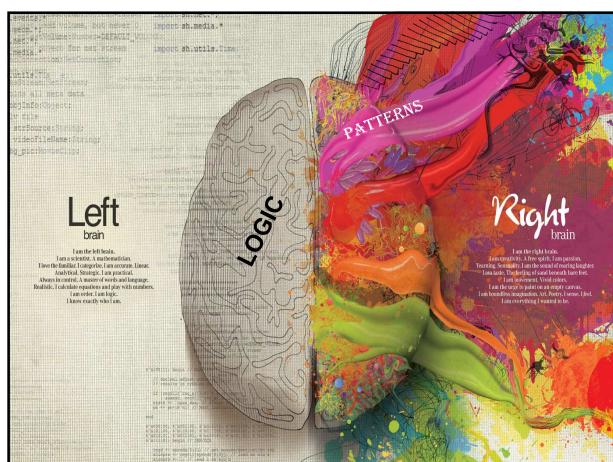
1984 App. Idea	How Basic-Cyc and/or Cyc-NL Could Be Used (more or less <i>today</i>)
1. Text Understanding and Document Preparation	Use context and pragmatic constraints to disambiguate pronouns, word senses (text) and slurs and prosody (speech), metaphors, irony, the order and scope of quantifiers, etc. and inject such things into generated text/speech. Smarter checking of "spelling" ("bear" => "bare"), grammar, style, and <i>content</i> (e.g., write "Barter, I discuss X." but never do)
2. Speech Understanding	
3. Translation / Generation	
4. Expert Systems	Less brittle behavior, because of the ability to fall back on increasingly general knowledge and the ability to analogize to specific but far-funking knowledge. More open-ended interactions with simulated characters, groups, and objects.
5. Training Simulations	
6. Games	
7. Online Commerce	Accrete models of the users to help match buyers/sellers and donors/donees, to help decide which colleges/donors/... to consider, and to decide which products someone would want (and when, and why, and how best to sell them)
8. Online Advice Services	
9. Directed Marketing	
10. Clean / Integrate DBs	Represent the meaning of the relations/files/columns/... as Cyc assertions, thus using Cyc as an integrating to virtually integrate them, to find inconsistencies among them, and dynamically add/link/guess/reorder user interface menu choices, display modes, etc.
11. ... and Spreadsheets	
12. 'Active' Menu/Forms	
13. Helpful Encyclopedia	Represent the actual content of online articles, captions, documents (e.g., court rulings), and user queries; then use deduction to find matches and to answer questions even if that requires combining multiple pieces of information logically and/or arithmetically.
14. Question-answering	
15. Search for docs/photos	Use Cyc to spawn plausible chains of action/reaction chains (scenarios) using that data.
16. Scenario Generation	VCR's/songs/coffeepots... that converse with their users (and learn their likes, habits, etc.) Help forgetful people remember where things are, who is related to who, etc.
17. Appliances	Stimulate creativity, offload tasks/decisions, catch interpersonal misunderstandings, etc.
18. Mental Prostheses / Mental Amplifiers	
19. Automated Discovery	Use Cyc constraints to guide the process of hypothesis formation, experiment design, data collection, evaluation of results, judging their significance, introduction of new terms
20. Image Compression	Store (and transmit) terse semantic descriptions of situations instead of pixels

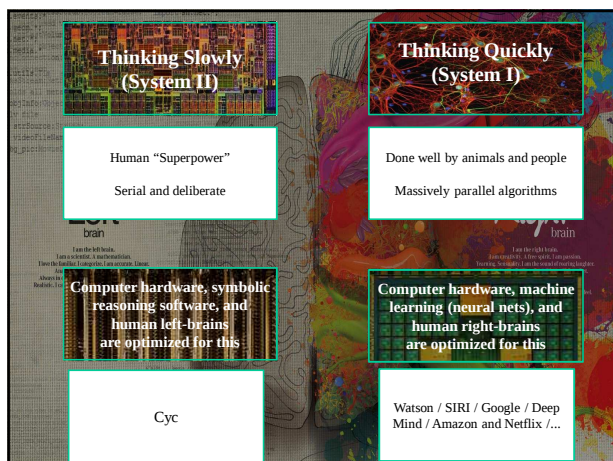
5 almost-as-old (~1990) Application Ideas

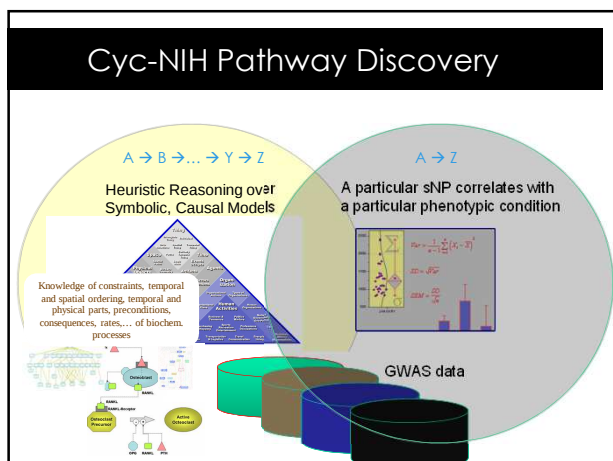
2002 App. Idea	How Basic-Cyc and/or Cyc-Net Could Be Used (more or less <i>today</i>)
21. Email	Prioritize, summarize, add explanatory footnotes color-coded by type (explanation of an acronym, link to proper noun's home page, link to something it seems to contradict), increasingly auto-forward (to appropriate other recipients) and auto-answer messages.
22. Network Security	Generate plausible network attack scenarios for various types of perpetrators and goals, based on a model (machines on the network, programs each has installed/running, open ports, ...); propose defenses; do "What-if...?" reasoning on edited model to try them out
23. Ontological Engineering	Actively help a user (a non-logician, non-programmer,...) to add new knowledge to Cyc, to refine and organize knowledge, and to articulate and refine their questions as well.
24. Wireless	Mediate (using common sense, dialogue, etc.) to help someone using a wireless device (e.g. laptop) with little or no screen real estate who is attempting to navigate CRT-designed web pages
25. Thesaurus Management	Semi-automate the process of aligning massive technical-term lists, such as the numerous unaligned 300k-term thesauri used by pharmaceutical companies.

Cyc Tech. Intro/History/State

- 1960's and 1970's AI: The Brittleness Bottleneck
- 1983: The Opportunity (and the Approach)
- 1984 – 2018: Staying true to that approach
- Now: Blossoming of applications and synergies
 - Synergy between System 2 and System 1 Thinking
 - Synergy between Left and Right Brain Thinking
 - Synergy between causal modeling and statistical learning

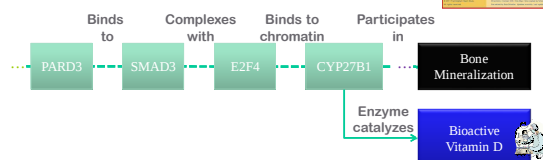






Cyc-NIH Pathway Discovery

$A \rightarrow B \rightarrow \dots \rightarrow Y \rightarrow Z$

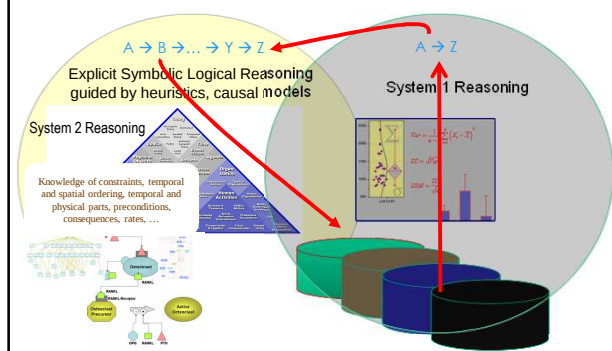


Framingham SHARe

- CD40 ligand from plasma
- TNF- α , pg/ml
- Osteoprotegerin from plasma
- Vitamin D, ng/L
- Cystatin C, mg/L
- P-selectin from plasma, n
- ... hundreds more



System 1 and 2 Reasoning Synergy

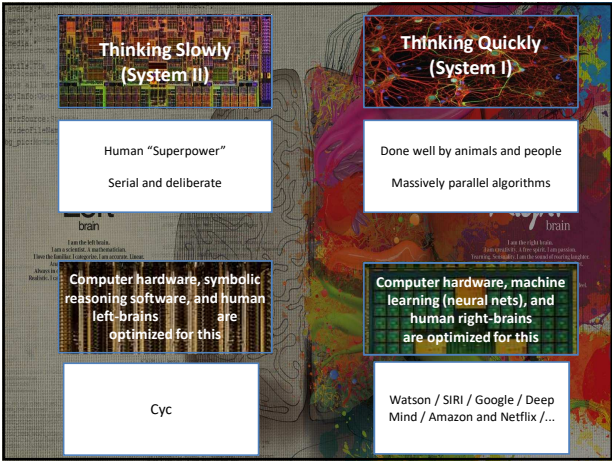


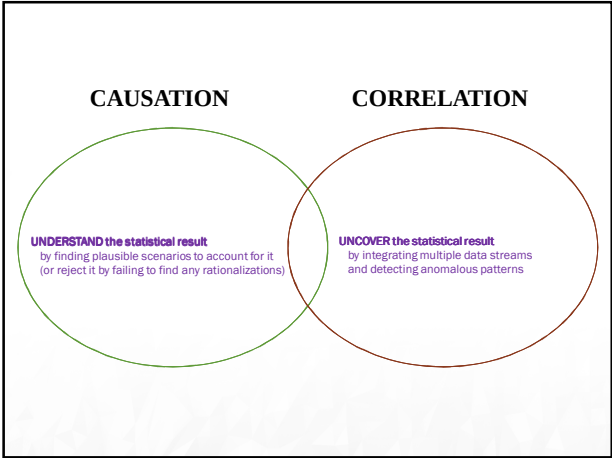
Why "AI" is Only Half the Answer

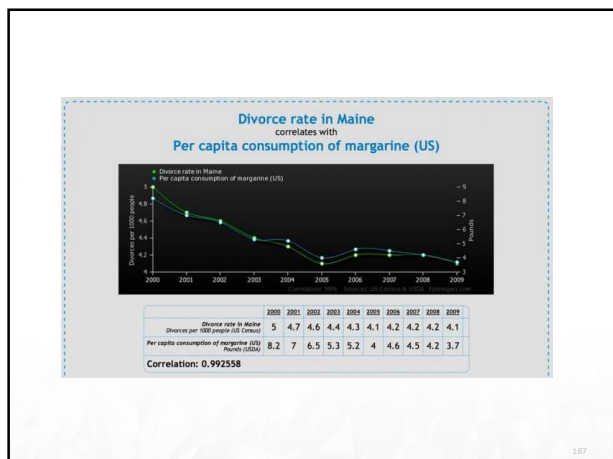
- ☒ There are two AI power sources (statistics, logic).
- ☒ Everyone else is focusing on right-brain AI (brittle)
- ☒ One exception: CYC (Manhattan-Project-like effort)
- ☒ Now that it's more or less done, building a Cyc app. means extending Cyc's ontology, KB, ... slightly
- ☒ The path to "real AI": statistics+logic synergy

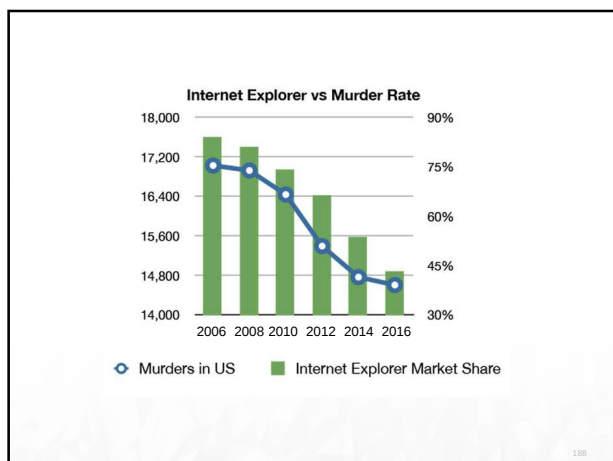
- A couple quick demo's of Cyc running.
- Examples drawn from petroleum engineering Cyc app's.











Why "AI" is Only Half the Answer

Sometimes The Veneer of Intelligence is Not Enough

Doug Lenat
Lenat@cyc.com

CYC Corp, 7718 Wood Hollow Drive, Austin, Texas

SPE

February 7, 2018
