National University of Mexico

October 18th 2002

Computational models of thinking

Phil Johnson-Laird
Princeton University

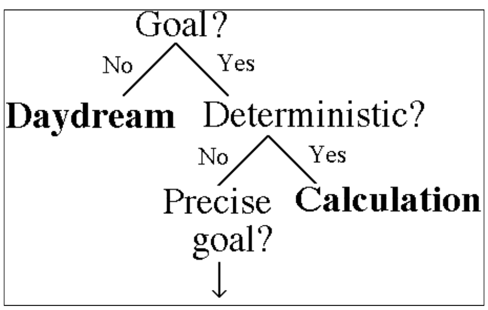
phil@princeton.edu
www.princeton.edu/~psych/PsychSite/~phil.html

GOALS OF THE TALK

- How do humans think and reason?
- Illustrate role of computer models in development of theories.
- Interdisciplinary nature of cognitive science: logic, linguistics, computer science & AI, experimental psychology, and neuroscience.

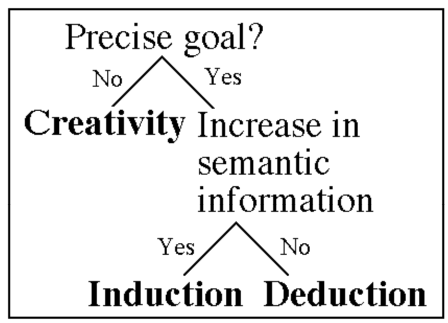
Professor Phil Johnson-Laird, Princeton University, 2002

A TAXONOMY OF THINKING



Professor Phil Johnson-Laird, Princeton University, 2002

TAXONOMY CONTINUED



Professor Phil Johnson-Laird, Princeton University, 2002

AN ALTERNATIVE TO TURING TEST

- Turing said: Don't ask can machines think, but ask can you distinguish between machine and human?
- Cognitive science: develop a theory of thinking, test it experimentally, and implement it in a computer model.
- *What* is computed?
  How is it computed?
  Where in the brain is it computed?

Professor Phil Johnson-Laird, Princeton University, 2002

AN INFERENCE

- A man: 'Does this train go to Ickenham?'
- Phil knew:

If the train goes to Uxbridge then it goes to Ickenham. [from Map]

The train goes to Uxbridge.

∴ The train goes to Ickenham.

- Phil: 'Yes!' [Doors closed.]

Professor Phil Johnson-Laird, Princeton University, 2002

ANOTHER INFERENCE

Some engineers knew:

- If the experiment is to continue then the turbines must be rotating fast enough
- The turbines are not rotating fast enough.

They went ahead with the experiment.

[Chernobyl disaster]

TWO FORMS OF INFERENCE

- If there's a square then there's a triangle.
  There's a square. What follows?
  Therefore, there's a triangle [17 out of 19]
- If there's a square then there's a triangle.
  There's not a triangle.
  Therefore, there's not a square [9 out of 19]
  "Nothing follows." [9 out of 19]

WHY LOGIC CANNOT BE A THEORY OF REASONING

- People draw their own conclusions from premises (with systematic biases).
- They say "nothing follows" from such premises: *It is raining. 3 is a prime.*
- From any set of premises, logic implies infinitely many valid conclusions, e.g.:

It is raining and 3 is a prime, or it is sunny.

No sane person makes such an inference.

THEORY OF MENTAL LOGIC

- Mind equipped with formal rules of inference (of which we are unaware).
- Example of proof:

If a square then a triangle.

Not a triangle.

1. Suppose: a square.
2. ∴ a triangle (modus ponens: if p then q; p; ∴ q)
3. ∴ a triangle & not a triangle (conjunction)
4. ∴ *Not* a square (reductio ad absurdum)

Example from Rips, L.J. (1994) *The Psychology of Proof.* MIT Press.
Complete computer implementation is not in public domain.

EFFECTS OF CONTENT

- *If Bill is in Rio then he is not in Sweden.*
  He is *in Sweden.* What follows?
  Enhances: ∴ *He is not in Rio.* (92%)
- *If Bill is in Brazil then he is not in Rio.*
  He is *in Rio.* What follows?
  Impedes: ∴ *He is not in Brazil.* (34%)
- Moral: mind does not use formal rules.

Results from J-L & Byrne, R.M.J. (2002) Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646-678.

THEORY OF MENTAL MODELS

- People envisage possibilities: each mental model represents a possibility.
- Mental models represent only what is true. The principle of 'truth'.
- An inference is **valid** *iff* it holds in all models of the premises.
- Counterexamples show that inference is **invalid**: a possibility in which premises hold but conclusion doesn't.

J-L (2001) Mental models and deduction. Trends in Cognitive Science, 5, 434-442.

COMPUTER IMPLEMENTATION

- COMMON LISP (lists, arrays)
- Stage 1: Parse sentence + compositional semantics.

  Stage 2: Check relation between sentence and current models: valid or inconsistent.

  Stage 3: Else update existing set of models.
- Four levels of expertise, from novice to AI.

J-L & Savary, F. (1999) Illusory inferences: a novel class of erroneous deductions. *Cognition*, 71, 191-229. Program in public domain.

COMPOSITIONAL SEMANTICS

- Meaning of an expression depends on meaning of its parts and their syntactic interrelations – Frege (1892), Tarski (1936).
- Each syntactic rule has a corresponding semantic rule (Montague, 1974); standard in design of compilers, but the programming language, CHLF-5, was an exception.
- As parser uses a rule to analyze the syntax of a constituent, it calls the corresponding semantic rule to build interpretation.

SENTENTIAL CONNECTIVES

Connective	Interpretation as set of possibilities		
A and B.	a	b	
A or B, not both.	a	¬ b	['¬' is negation]
	¬ a	b	
If A then B.	a	b	
	¬ a	b	
	¬ a	¬ b	

But, *A* and *B* can be sentences containing connectives. Expert level of program: AI.

SYNTACTIC RULES

- A "context-free" grammar with recursive rules of the following sort:-

Sentence → variable

Sentence → *not* sentence

Sentence → sentence *and* sentence

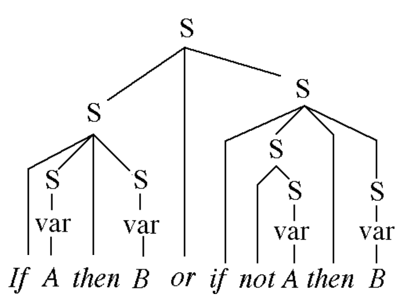
Sentence → sentence *or* sentence

Sentence → *if* sentence *then* sentence

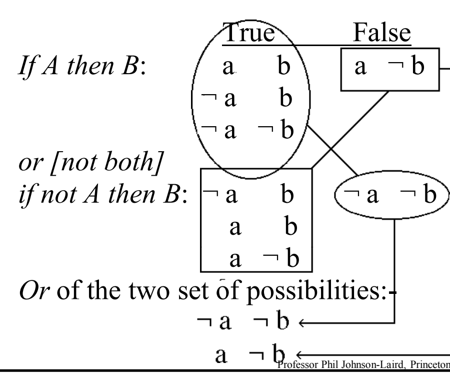
Variable → *A, B, C*, … [A lexicon]

- Parser needs 'stack-like' memory (Chomsky's 1959 proof).

A SHIFT-AND-REDUCE PARSER



If A then B or if not A then B



MENTAL MODELS

Connective	Mental models
A and B.	a b
A or B, not both.	a
	b
If A then B.	a b
	. . . ← implicit model

- Represent what is true, not what is false.

MENTAL MODELS FOR: If A then B or else if not A then B

- Program at lowest level of expertise:

	True
If A then B:	a b
	. . .
or [not both]	
if not A then B:	¬ a b
	. . .

Or of the two sets of possibilities:-

a b

¬ a b

AN ILLUSORY INFERENCE

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

There is an ace in the hand.

What follows?

Nearly everyone says: *there is a king.*

J-L & Savary, F. (1999) Illusory inferences: a novel class of erroneous deductions. *Cognition*, 71, 191-229. Program in public domain.

AN ANALYSIS

- *If ace then king or else if not ace then king.*
- Mental models / AI models

Mental models		AI models	
ace	king	ace	¬ king
¬ ace	king	¬ ace	¬ king

- *There is an ace.*
- Mental model / AI model

Mental model		AI model	
ace	king	ace	¬ king

A VARIANT

- Only one of the following two assertions is true about cards in a hand:
  - *If ace then king.*
  - *If not ace then king.*
- The following is definitely true:
  - *There's an ace.*
- What follows?

ANOTHER SORT□□□ OF ILLUSION

- Only one of the following assertions is true about a particular hand of cards:

  There is a king or an ace, or both.
  There is a queen or an ace, or both.
  There is a jack or a ten, or both.

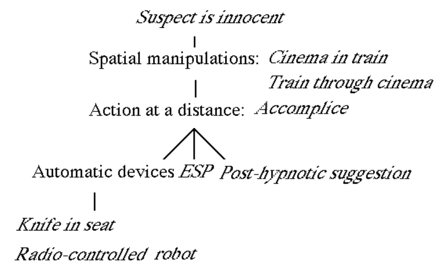
  Is it possible that there is an ace in the hand?
- 99% of Princeton students: "Yes".

From Goldvarg, Y., & J-L (2000) Illusions in modal reasoning. Memory & Cognition, 28, 282-294.

AN INFERENCE FOR A JURY

- The victim was stabbed to death in a cinema.
- The suspect was on an express train to New York city at the time of the stabbing.
- What conclusion would you draw?

INDUCTIVE INFERENCES



Professor Phil Johnson-Laird, Princeton University, 2002

AN INFERENCE

- More than half the people in the room speak Spanish.

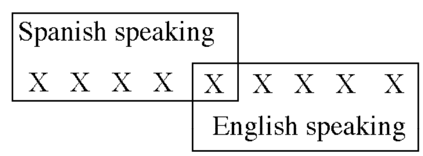
  More than half the people in the room speak English.

  Does it follow that more than half the people in the room speak Spanish and English?
- Even 7 year-olds say: "No".

Professor Phil Johnson-Laird, Princeton University, 2002

COUNTEREXAMPLES

- Reasoners use counterexamples, i.e., model or diagram of possibility consistent with premises but not with conclusion:



Professor Phil Johnson-Laird, Princeton University, 2002

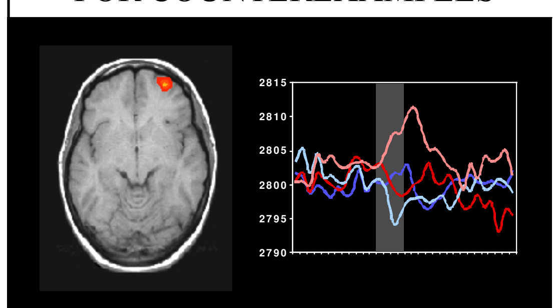
BRAIN IMAGING

- Functional magnetic resonance imaging (fMRI) reveals active regions of brain.
- Comprehension depends on left-hemisphere.
- Reasoning depends on right-hemisphere (non-linguistic regions, which may mediate spatial processes).
- Study compared reasoning with, and without, search for counterexamples & mental math.

Kroger, J., Cohen, J.D., and J-L (2002) A double dissociation between logic and mathematics. Under submission.

Professor Phil Johnson-Laird, Princeton University, 2002

BRAIN ACTIVITY IN SEARCH FOR COUNTEREXAMPLES



Professor Phil Johnson-Laird, Princeton University, 2002

WHAT, HOW, WHERE?

- What is computed in reasoning? conclusion that holds in set of possibilities consistent with premises.
- How? Compositional semantics yields *mental* models of premises. Induction uses general knowledge.
- Where? Comprehension in left hemisphere; significant right hemisphere activity for reasoning.

CONCLUSIONS

- Humans think and reason by manipulating mental models.
- Computer implementation of theories:
  prevents theorists assuming too much.
  can yield surprising predictions (illusions).
- Cognitive science is interdisciplinary: takes ideas from theory of computability, logic, compiler design, linguistics, and many other disciplines.