The attribution problem in Cognitive Science

- We can't see the processes we care the most about, so we must infer them from observable behavior.
- But how can we infer the invisible?
- Well, …what is visible?
- Reason-respecting behavior.
- How can we account for that?

Reason respecting behavior

- Info processing psychology – knowledge, goals, plans, means
- Language (formal linguistics)
- Theorem proving
- Chess and other games

Newell's BIG News

- THE central questions in cognitive science are these:
  - How can the phenomena of mind exist in the physical world?
  - How can the physical phenomena of mind be explained?
- Now, after 2000 years of asking, we know the answers!
- Physical Symbol Systems

Formal Systems

- We know of another system that produces reason-respecting sequences.
- It's LOGIC
- FORMal, get it?
- Strings of symbols
- Rules for manipulating strings of symbols
- *"If you take care of the syntax, the semantics will take care of itself."* (Haugeland, 1981)

The classical view of computing and cognition (PSSH)

- Symbols and expressions (designation and interpretation)
- Meanings are composed of meaning elements
- Formal operations transform expressions
- Three distinct levels
  - knowledge/computational
    - (what does it do?)
  - symbol/representational
    - (how is the doing organized?)
  - biology/implementation
    - (what stuff does it?)

Formal Systems have a history

Early Accounting Systems

Six ovoid tokens representing an account of six units of oil

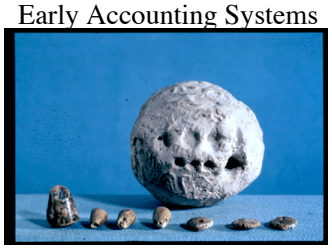
Early Accounting Systems

Plain tokens. Mesopotamia, 4000 B.C.

Early Accounting Systems

An envelope and its contents representing 7 units of oil

Early Accounting Systems



An envelope, its contents of tokens, and corresponding markings. 3300 B.C.

Early Accounting Systems

Complex tokens. Sheep, oil, metal, garment. 3300 B.C.

Early Accounting Systems

Impressed tablet showing an account of grain. 3100 B.C.

Early Accounting Systems

Pictographic tablet showing 33 units of oil. 3100 B.C.

Properties of Spoken and Written Language

- Spoken
  - Ephemeral
  - Dynamic
  - Auditory (sound)
  - Structure in time
- Written
  - Semi-permanent
  - Static
  - Visual (sight)
  - Structure in space

decontextualization

decontextualization

weave or surrounds

with

relation to

not

to make or take

the act

The Secret of Our Success

The world of things and events

Encoding →

Representations of the world of things and events

Formal operations ↓

New representations of the world

← Decoding

The Secret of Our Success

The world of things and events

Encoding →

$t = 0$
$x = 0$

$a = g = 9.8 m/sec^2$
$v = gt$
$x = 1/2gt^2$ ↓

Falling object

Falls 490m in 10 seconds

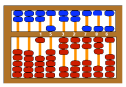
← Decoding

$t = 10$ sec.
$x = 1/2\ g\ 100$
$= 50g$ meters

Getting symbols to behave in a way that fits the world

- Why does this turn out to be possible?
- Why is the world a place that can be modeled by mathematics?
- No one knows, but it DOES work!

Consider the abacus



Here patterns of beads represent numbers. People learn rules for transforming these patterns of beads in such a way that the semantic interpretation of before-and after pairs corresponds to a useful mathematical function. But there is nothing instrinsically mathematical about the rules themselves: they are just rules for moving beads around. What makes the rules useful for doing mathematics is that we are assured of a certain continuing correspondence between the formal or syntactic patterns of beads and mathematical objects (such as numbers).

The Turing Machine

- An imaginary (theoretical) device.
- It works by manipulating meaningless symbols.
- It can compute the answer to any sufficiently well-specified problem.
- Digital computers are not imaginary, and they can be equivalent to a Turing machine.

The three big pieces of early Cognitive Science

1. Formal Systems
2. Meaningful computation by mindlessly following rules
3. Mechanized symbol manipulation

Mindware as reason-respecting Software

- In a formal system, state plus operator implies a new state,
- States of mind could lead to other states of mind in ways that follow rules.
- It is the program that matters.
- The machine it runs on is “a mere implementational detail.”

in physical symbol systems:

- Symbols as abstract types that express the identity of multiple tokens.
- Expressions as structures containing symbol tokens.
- Designation as a relation between a symbol and the entities it symbolizes.
- Interpretation as realizing the designations of expressions.
- Operations of assigning symbols, and copying, reading, and writing expressions.

PSS Hypothesis

- The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system.
- Necessary means that any physical system that exhibits general intelligence will be an instance of a physical symbol system.
- Sufficient means that any physical symbol system can be organized further to exhibit general intelligent action.
- General intelligent action means the same scope of intelligence seen in human action: that in real situations behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some physical limits.

Physical Symbol System Hypothesis

- Symbolic Codes
- Cognition happens at the level of deliberative thought – symbol or representational level
- Intelligence, wherever it is found including human intelligence, will be found to be a physical symbol system.

Reverse Engineering

- Pick something that people do that is smart.
- Figure out how to do that same thing on a computer.
- Then look at the program in the computer. It should tell you something about the nature of the task and the things the person must do in order to perform the task.

And it works!

- This is exciting!
- Just document the I/O relations,
- build a program that can do the job (the ***sufficiency*** criterion),
- and then look inside to see how it does it.
- Wow!
- This is Artificial Intelligence.
- It is Newell's big news.

Gloating

(possible in 1980, not so easy now)

- These advances far outstrip what has been accomplished by other attempts to build intelligent mechanisms, such as the work in building robots driven directly by circuits; the work in neural nets, or the engineering attempts at pattern recognition using direct circuitry and analogue computation. (Newell, 1980: 171-2)

Can we get symbols to work in a way that fits the brain?

- An empirical hypothesis: Physical symbol system hypothesis.

Wait a minute!

- How was it decided that the PSS must reside in the brain?
- Is that even the correct question?
- Examine Newell's article.

Sliding inside the head

- Although little can be said about exact boundaries, some interior milieu must exist within which the symbol system can freely and successfully interpret expressions. (Newell, 1980: 158)
- These capabilities must exist in some interior system, and thus can be illustrated there, without involving the interaction with the external world. (Newell, 1980: 167)

Make the world go away

- A novel feature of physical symbol systems is the approach to symbolic function, not just by processing, but by *internal* symbolic processing. …
- The prototype symbolic relation is that of access from a symbol to an expression, not that of naming an external object. (Newell, 1980: 169)

Fully committed to internal symbolic processes

- There must exist a neural organization that is an architecture that supports a symbol structure.

Where is the boundary between symbolic and sub-symbolic levels of operation?

- The neural system is not in fact irrelevant- its operation supports the symbolic level. But it does so in a way that normally hides most of its properties, realizing instead a symbol system with properties of its own. (p. 175)

Thinking Meat?!

How can we get Reason-respecting behavior out of a lump of flesh?

Could the brain be a meat computer?

- The program (the rules) run by the brain must be a formal system
- Brain states must correspond to symbols or propositions in a formal language.
- Functional equivalence of your brain states to mine – Not identity.
- Brain states must cause other brain states in just the right “reason-respecting” way.

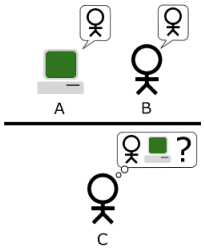
The meat wouldn't matter

- If we knew the program, we could run it on any suitable computer.
- Then we would have an artificial human mind!
- And we already know how to do this.

Ok, But….

- The time course of real-world action
- There are many levels of software. Are there also levels of mindware?
- Are games good representatives of cognitive tasks?
- Is the Turing test a good representative?

The Turing Test



OK, But…(continued)

- Engineering A/I vs Research A/I
- Consciousness (the C word) and qualia
- Language and Searle's Chinese Room

Searle's Chinese Room

- A person (you?) in a room with a slot in the door.
  - Book of rules
  - Box of symbols
- Chinese people outside push strings of symbols through the slot
- You use the rules to make new strings and push them out the slot.
- Do you speak chinese? Does the room?

Why not formal symbols?

- What symbolic computers do well and what people do well.
- Nature doesn't work like that.
- The metaphor is vague.
- Every device, considered at different levels, could be a model of many different things.

Questions remain

- If the PSSH does not describe what the brain does, what explanation do you offer instead? What's the alternative?
- If the brain doesn't do PSSH,
  - What does the brain do?
  - And what does PSSH?
    - (because banishing PSSH from the brain doesn't make it go away.)

Is the Meat Magic?

- The answer depends on whether or not we think consciousness is simply a matter of information processing.

What is Cognitive Science?

- "Cognitive science has been viewed as the study of the natural domain of cognition…
- where the latter includes prototypical phenomena of perception, problem-solving, reasoning, learning, memory, and so on."

How to make a field in science

1. Round up some phenomena that seem related.
2. Find some principles that unify the phenomena or that explain patterns and regularities
3. Use those principles to put boundaries on the domain of inquiry. Specify what is in and what is out.
4. Figure out how the stuff in the domain works

Limiting the scope of phenomena to be accounted for

- Thus, our situation is one of defining a symbol system to be a universal machine, and then taking as a hypothesis that this notion of symbol system will prove adequate to all of the symbolic activity this physical universe of ours can exhibit, and in particular all the symbolic activities of the human mind. (Newell, 1980:155)

Two aspects of mindfulness ?

- Reason respecting flow of thoughts (symbolic activity)
- Everything else
  - Qualia
  - Affect/emotion
  - Embodiment

Are these really different aspects?

Examining assumptions

- What was modeled? "Every such operation consists of some change of the physical system consisting of the **computer** and **his** tape." Turing, 1937
- Are the levels really separable?
- Are embodied meanings compositional?
- Explore: is an abacus a computer? No.
- Is the navigation team a computer? Yes. Notice how well the account works for a socio-cultural system.

Still doing science – bounding the field

- Once upon a time, PSSH was the only game in town (if we all worked on PSSH, how would an alternative ever come along?)
- Co-constitution of theory and the set of phenomena for which the theory will be responsible. Eg. Turing test

Why the UCSD department of cognitive science does not study cognitive science.

It would be both surprising and troublesome if too many of what we pretheoretically took to be clear cases of cognition ended up being omitted in the process. But it would also not be entirely surprising if some of our favorite candidate "cognitive" phenomena got left out. For example, it could turn out that consciousness is not something that can be given a computational account. Similarly, certain kinds of statistical learning, aspects of ontogenetic development, the effect of moods and emotions, and many other important and interesting phenomena could simply end up not being amenable to a computational account. (Pylyshyn p. 40)

Newell's constraints on general intelligence

1. Behave as an (almost) arbitrary function of the environment (universality).
2. Operate in real time.
3. Exhibit rational, i.e., effective adaptive behavior.
4. Use vast amounts of knowledge about the environment.
5. Behave robustly in the face of error, the unexpected, and the unknown.
6. Use symbols (and abstractions).

Newell's constraints (continued)

7. Use (natural) language.
8. Exhibit self-awareness and a sense of self.
9. Learn from its environment.
10. Acquire its capabilities through development.
11. Arise through evolution.
12. Be realizable within the brain as a physical system.
13. Be realizable as a physical system.