ROADS TO HUMAN LEVEL AI?

biological—imitate humans. Even neural nets, should work eventually.

engineering—solve problems the world presents— presently ahead

direct programming, e.g. genetic algorithms

use logic, loftier objective. If it reaches human level, we will understand intelligence, and so will our robots. Logical AI has partly solved some inevitable problems that haven't even been noticed by physiological approaches.

The logic approach is the most awkward—except for all the others that have been tried.

Logic in AI

Features of the logic approach to AI—starting in 1958.

- Represent information by sentences in a logical language, e.g. first order logic, second order logic, modal logic.

- Auxiliary information in tables, programs, states, etc. is described by logical sentences.

- Inference is logical inference—deduction supplemented by calculation and some form of nonmonotonic inference—1980.

- Action takes place when the system infers that it should do the action.

- Observation of the environment results in sentences in memory.

Topics, methods and problems of logical AI

- deduction, nonmonotonic reasoning, theories of action, problem solving,
- The frame, qualification, and ramification problems have been partly solved.
- concepts as objects, contexts as objects, approximate objects
- Elaboration tolerance (educate without brain surgery)

THE COMMON SENSE INFORMATIC SITUATION

• Common sense: A structure composed of abilities and knowledge. "Programs with common sense"—1959.

• The *common sense* informatic situation, which differs from a *bounded* informatic situation, has been difficult to define precisely.

• Bounded informatic situations, e.g. chess positions, take specific facts into account. In common sense, there is no limitation on what objects and facts may become relevant.

• New facts may require revising conclusions, plans, and algorithms. Formal *nonmonotonic reasoning*, e.g. circumscription and default logic, are important tools for representing common sense reasoning in logic.

• Actions and other events often have only partly knowable effects. Often not even probabillistic models are available.

COMMON SENSE INFORMATIC SITUATION—2

• Specific theories, e.g. scientific theories, are embedded in common sense. Skills are also embedded in common sense.

• Common sense physics: When two objects collide, there is a noise. An object pushed off a table will fall to the floor.

• Common sense psychology: A person comes to dislike someone whom he thinks killed his fellow countrymen.

• The facts behind many human abilities are not ordinarily expressed in language but are often expressible in language or logic.

• Common sense abilities: Grasp object being touched, Recognize a surface of an object—the knife. Fumble for plastic surface.

• Common sense facts: $In(Rpocket, Knife, Now)$
$\wedge(\exists x)(Plastic\text{-}feel(x) \wedge Surface(x, Knife))$,
• Human-level common sense requires representing the up-to-now mental state as an object and reasoning about it.

EMBEDDING A SCIENTIFIC FACT IN SITUATION CALCULUS

• Scientific theories are embedded in common sense, and the formulas are embedded in natural language.

• Galileo's formula $d = \frac{1}{2}gt^2$ can be embedded in a simple common sense theory of situation calculus by

$$Falling(b,s) \wedge Velocity(b,s) = 0 \wedge Height(b,s) = h \wedge \wedge h - d > 0 \wedge$$
$$d = \tfrac{1}{2}gt^2$$
$$\rightarrow (\exists s')(time(s') = time(s) + t \wedge Height(b,s') = Height(b,s) - h). \quad (1)$$

• For controlling a robot (1) must be used in connection with facts about concurrent events.

• The situation calculus formula connects Galileo's formula to quantities that are defined in (mostly unobserved) situations to which the theory applies.

• Like other scientific formulas, Galileo's formula is used more in constructing theories than in planning action in a specific situation. Robots may use (1), i.e. $d = \frac{1}{2}gt^2$ expanded into situation calcculus, directly if they can immediately measure the physical quantities involved.

EMBEDDING A SKILL IN COMMON SENSE—a philosophical path

• Objects exist independent of perception.

• Machine learning research is mistaken in concentrating on classifying perception. Herbert Simon's Bacon method for scientific discovery is limited by its concentration on discovering relations among observables.

• A 3-d object is not a construct from 2-d views. We learn about objects from view and by other means. The blind live in the same world as the sighted.

• Draw an object you can only feel but can't see. A program that can get an object from a pocket is a good *Drosophila*.

EMBEDDING A SKILL IN SITUATION CALCULUS

The skill of finding an object in a pocket can be partly embedded in situation calculus.

$$In(Knife, RPocket, s) \rightarrow Holding(Knife,$$
$$Result(Move(RHand, Interior(RPocket)); FumbleFor(Knife);$$
$$Grasp(Knife); Remove(RHand, RPocket), s))$$

Alternatively,

$$In(Knife, RPocket, s) \rightarrow (\exists finger\ surface)$$
$$(surface \in Surfaces(Knife) \wedge finger \in Fingers(RHand)$$
$$\wedge(\lambda(s')(Touches(finger, surface, s') \wedge Observes(Touches(finger,$$
$$surface), s')))(Result($$
$$Move(RHand, Interior(RPocket));$$
$$FumbleFor(Knife)), s)).$$

Complications: Conscious guiding of the fumbling: fumble until object is found, very little detailed information is needed, and very little is available. For example, one doesn't need information about the other objects in the pocket.

Query: What do we know about the physics of pockets, and how is it represented in the human brain, and how should robots represent it?

"Keep trying a, and you will shortly achieve a situation s' such that $Holds(f, s')$." How should this be represented logically?

SPECIFIC ABILITIES IMBEDDED IN COMMON SENSE

- Skills like walking, playing tennis
- Scientific theories
- AI programs, e.g. Mycin
- A chess player and a chess program
- Make a decision based determining which of two actions leads to a better resulting situation. Humans only maybe.

INTERACTION OF SKILLS AND KNOWLEDGE

- Partial knowledge of the skills and of situations.

- Picking my knife from my pocket containing coins and keys.

- Interaction of observation with reasoning about action in logical AI, e.g. Filman, Reiter, Levesque, Shanahan, Sandewall, Doherty.

- It may be new to emphasize partial knowledge about effects of exercising a skill.

CURRENT PROJECT—DOMAIN DEPENDENT CONTROL FOR LOGICAL PROBLEM SOLVER

• General logical problem solvers without domain dependent control experience combinatorial explosion.
• There is a profusion of cut-down logics.
• STRIPS should be a strategy for a logical problem solver. Likewise DASL.
• Minsky proposed in 1956 that a geometry theorem prover should only try to prove sentences true in the diagram. Herbert Gelernter implemented it, but in 1959 IBM decided IBM should not be seen as making anything but data-processing machines.
• Selene Makarios works on domain dependent control. She has some results in reducing search in the blocks world.

EXAMPLES OF CONTROL

• When looking for feasible actions, don't subsitute into formulas of the form $Result(a, s)$. This is part of STRIPS.

• When trying to prove two triangles congruent and you have side a in one triangle equal to side a' in the other, try to prove the corresponding adjacent angles equal.

• In blocks world and heuristically similar problems look for moves to final position.

• Josefina Sierra-Ibañez and more recently Selene Makarios.

APPEARANCE AND REALITY

- The world is made of three-dimensional objects which are only partly observable.
- History is only partly observable or even knowable.
- Reality is more persistent than appearance.
- Appearance of a scene after an event depends on the reality of the scene and not just on what could be observed.
- Pattern recognition and scientific discovery research has not properly taken these facts into account.

ELABORATION TOLERANCE

• A collection of facts, e.g. a logical theory, is elaboration tolerant to the extent that it can be readily elaborated. www.formal.stanford.edu/jmc/elaboration.html contains theory and extensive examples.

• English language statements are very elaboration tolerant provided human common sense is available. Adding sentences will almost always work.

• Neural nets, connectionist systems, and present chess programs have almost no elaboration tolerance. Example: T. Sejnowski's Nettalk cannot be elaborated to include Pinyin pronunciations of the letters "x" and "q".

• Many elaborations of well constructed nonmonotonic logical theories can be accomplished just by adding sentences.

• Example from "Missionaries and Cannibals". There's an oar on each bank of the river, ...

• *Formalizing Elaboration Tolerance* by Aarati Parmar is a forthcoming Stanford dissertation.

FREE WILL IN A DETERMINIST WORLD

• We can make a situation calculus theory of a process more determinist by adding axioms asserting that certain events occur.

• Human free will may consist of using a non-determinist theory to decide deterministically on an action.

Here's a minimal example of using a non-determinist theory within a determinist rule.

$$
\begin{array}{l}
Occurs(Does(Joe, \\
\quad \textbf{if } \text{Better-for}(Joe, Result(Does(Joe, a1), s), \\
\qquad\qquad Result(Does(Joe, a2), s)) \\
\quad \textbf{then } a1 \\
\quad \textbf{else } a2 \\
\qquad), s).
\end{array}
\tag{2}
$$

• Here Better-for-Joe($s1, s2$) is to be understood as asserting that Joe thinks $s1$ is better for him than $s2$.

If we take the actor as understood, as is common in situation calculus stdies, we get a shorter formula

$$\begin{array}{l} Occurs(\textbf{if } \text{Better}(Result(a1, s), Result(a2), s) \\ \quad \textbf{then } a1 \\ \quad \textbf{else } a2 \\ \qquad), s). \end{array} \tag{3}$$

• Do animals, even apes, make decisions based on comparing anticipated consequences? If not, can apes be trained to do it? Chess programs do. According to Dan Dennett, some recent experiments suggest that apes sometimes consider the consequences of alternate actions. Jane Goodall (personal communication) assures me that chimpanzees do.

OTHER ASPECTS OF LOGICAL AI

- non-monotonic reasoning
- concepts as objects
- contexts as objects

REFERENCES

- Murray Shanahan, *The frame problem in artificial intelligence* M.I.T. Press, 1997

- "What is artificial intelligence"—John McCarthy http://www-formal.stanford.edu/jmc/whatisai.html.

- My AI articles are all on www-formal.stanford.edu/jmc/.

- These references are completely inadequate.

- This research was supported by DARPA.