HUMAN-TYPE COMMON SENSE NEEDS EXTENSIONS TO LOGIC

John McCarthy, Stanford University

• Logical AI (artificial intelligence) is based on programs that represent facts about the world in languages of mathematical logic and decide what actions will achieve goals by logical reasoning. A lot has been accomplished with logic as is.

• This was Leibniz's goal, and I think we'll eventually achieve it. When he wrote Let us calculate, maybe he imagined that the AI problem would be solved and not just that of a logical language for expressing common sense facts. We can have a language adequate for expressing common sense facts and reasoning before we have the ideas needed for human-level AI.

• It's a disgrace that logicians have forgotten Leibniz's goal, but there's an excuse. Nonmonotonic reasoning is needed for common sense, but it can yield conclusions that aren't true in all models of the premises—just the preferred models.

• Almost 50 years work has gone into logical AI and its rival, AI based on imitating neurophysiology. Both have achieved some success, but neither is close to human-level intelligence.

• The common sense informatic situation, in contrast to bounded informatic situations, is key to human-level AI.

• First order languages will do, especially if a heavy duty axiomatic set theory is included, e.g. $A \times B$, A^B, list operations, and recursive definition are directly included. To make reasoning as concise as human informal set-theoretic reasoning, many theorems of set theory need to be taken as axioms.

THREE KINDS OF EXTENSION: more may be needed.

- Non-monotonic reasoning. Gödel's completeness theorem tells us that logical deduction cannot be extended if we demand truth in all interpretations of the premises. Non-monotonic reasoning is relative to a variety of notions of *preferred interpretation.*

- Approximate objects. Many entities with which commonsense reasoning deals do not admit if-and-only-if definitions. Attempts to give them if-and-only-if definitions lead to confusion.

- Extensive reification. Contrary to some philosophical opinion, common sense requires lots of reification, e.g. of actions, attitudes, beliefs, concepts, contexts, intentions, hopes, and even whole theories. Modal logic is insufficient.

THE COMMON SENSE INFORMATIC SITUATION

- By the *informatic situation* of an animal, person or computer program, I mean the kinds of information and reasoning methods available to it.

- The *common sense informatic situation* is that of a human with ordinary abilities to observe, ordinary innate knowledge, and ordinary ability to reason, especially about the consequences of events that might occur including the consequences of actions it might take.

- Specialized information, like science and about human institutions such as law, can be learned and embedded in a person's common sense information.

- Scientific theories and almost all AI common sense theories are based on bounded information situations in which the entities and the information about them are limited by their human designers.

- When such a scientific theory or an AI common sense theory of the kinds that have been developed proves inadequate, its designers examine it from the outside and make a better theory. For a human's common sense as a whole there is no outside. AI common sense also has to be extendable from within.

- This problem is unsolved in general, and one purpose of this lecture is to propose some ideas for extending common sense knowledge from within. The key point is that in the common sense informatic situation, any set of facts is subject to elaboration.

THE COMMON SENSE INFORMATIC SITUATION (2)

The common sense informatic situation has at least the following features.

- In contrast to bounded informatic situations, it is open to new information. Thus a person in a supermarket for steaks for dinner may phone an airline to find whether a guest will arrive in time for dinner and will need a steak.

- Common sense knowledge and reasoning often involves ill-defined entities. Thus the concepts of my obligations or my beliefs, though important, are ill-defined. Leibniz might have needed to express logically, “If Marlborough wins at Blenheim, Louis XIV won’t be able to make his grandson king of Spain.” The concepts used and their relations to previously known entities can take arbitrary forms.

• Much common sense knowledge has been learned by evolution, e.g. the semi-permanence of three dimensional objects and is available to young babies [**?**].

• Our knowledge of the effects of actions and other events that permits planning has an incomplete form.

• We do much of our common sense thinking in bounded contexts in which ill-defined concepts become more precise. A story about a physics exam problem provides a nice example.

COMMON SENSE INFORMATIC SITUATION—PHYSICS EXAMPLE

A nice example of what happens when a student doesn't do the nonmonotonic reasoning that puts a problem in its intended bounded context was discussed in the *American Journal of Physics*. Problem: find the height of a building using a barometer.

- Intended answer: Multiply the difference in pressures by the ratio of densities of mercury and air.

- In the bounded context intended by the examiner, the above is the only correct answer, but in the common sense informatic situation, there are others. The article worried about this but involved no explicit notion of nonmonotonic reasoning or of context. Computers solving the problem will need explicit nonmonotonic reasoning to identify the intended context.

UNINTENDED COMMON SENSE ANSWERS

(1) Drop the barometer from the top of the building and measure the time before it hits the ground.

(2) Measure the height and length of the shadow of the barometer and the shadow of the building.

(3) Rappel down the building with the barometer as a yardstick.

(4) Lower the barometer on a string till it reaches the ground and measure the string.

(5) Sit on the barometer and multiply the stories by ten feet.

(6) Tell the janitor, “I’ll give you this fine barometer if you’ll tell me the height of the building.”

(7) Sell the barometer and buy a GPS.

- The limited theory intended by the examiners requires elaboration to admit the new solutions, and these elaborations are not just adding sentences.

- We consider two common sense theories that have been developed (the first now and the second if there's time). Imbedding them properly in the common sense informatic situation will require some extensions to logic—at least nonmonotonic reasoning.

A WELL-KNOWN COMMON SENSE THEORY

Here's the main axiom of the *blocks world*, a favorite domain for logical AI research.

$$Clear(x,s) \wedge Clear(y,s) \rightarrow On(x,y,Result(M$$

with the definition

$$Clear(x,s) \equiv (\forall z)\neg On(z,x) \vee x = Table. \tag{1}$$

Only one block can be on another. A version that reifies relevant fluents and in which the variable l ranges over locations is

$$\begin{aligned} &Holds(Clear(Top(x)),s) \wedge Holds(Clear(l),s) \\ &\quad \rightarrow Holds(At(x,l),Result(Move(x,l)),s). \end{aligned} \tag{2}$$

This reified version permits quantification over the first argument of $Holds$. More axioms than there is time to present are needed in order to permit inferring in a particular initial situation that a certain plan will achieve a goal, e.g. to infer

$$\begin{aligned} &On(Block1, Block2, \\ &Result(Move(Block2, Top(Block2, Result(Move(Block3 \end{aligned}$$

where we have $On(Block3, Block1, S0)$ and therefore $Block3$ has to be moved before $Block1$ can be moved.

More elaborate versions of the blocks world have been studied, and there are applications (Reiter and Levesque) to the control of robots. However, each version is designed by a human and can be extended only by a human.

We'll discuss the well known example of the stuffy room if there's time.

NEED FOR NON-MONOTONICITY

• Human-level common sense theories and the programs that use them must elaborate themselves. For this extensions to logic are needed, but Gödel showed that first order logic is complete. New conclusions given by extended inference rules would be false in some interpretations—but not in preferred interpretations.

• We humans do nonmonotonic reasoning in many circumstances. 1 The only blocks on the table are those mentioned. 2 A bird may be assumed to fly. 3 The meeting may be assumed to be on Wednesday. 4 The only things wrong with the boat are those that may be inferred from the facts you know. 5 In planning one's day, one doesn't even think about getting hit by a meteorite.

• Deduction is monotonic in the following sense. Let A be a set of sentences, p a sentence such that $A \vdash p$, and B a set of sentences such that $A \subset B$, then we will also have $B \vdash p$. Increasing the set of premises can never reduce the set of deductive conclusions.

If we nonmonotonically conclude that $B1$ and $B2$ are the only blocks on the table and now want to mention another block $B3$, we must do the nonmonotonic reasoning all over again. Thus nonmonotonic reasoning is applied to the whole set of facts—not to a subset.

- The word but in English blocks certain nonmonotonic reasoning. "The meeting is on Wednesday but not at the usual time."

- Nonmonotonic reasoning is not subsumed under probabilistic reasoning—neither in theory nor in practice. Often it's the reverse.

- Many formalizations of nonmonotonic reasoning have been studied, including circumscription, default logic, negation as failure in logic programming. We'll discuss circumscription, which involves minimization in logical AI and so is analogous to minimization in other sciences.

- There are also general theories of nonmonotonic reasoning [**?**]. Unfortunately, the ones I have seen are not oriented towards common sense.

CIRCUMSCRIPTION

Circumscription is a form of minimization in logic, perhaps a logical analog of calculus of variations. We minimize a predicate P with one or more arguments. We are allowed to vary Z, a vector of predicates or domain elements with respect to an ordering $P < P'$. We use a notation proposed by Vladimir Lifschitz where $CIRC[A; P; Z]$ is defined by

$$A(P, Z) \wedge \neg(\exists p\ z)[A(p, z) \wedge p < P]$$

Any unmentioned symbols are thus assumed constant for the purposes of circumscription. Often $p \leq p' \equiv (\forall x)(p(x) \rightarrow p'(x))$.

I don't know whether circumscription admits anything analogous to Lagrange multipliers.

AN EXAMPLE OF CIRCUMSCRIPTION

Let A be the following axiom concerning objects that fly.

$$\begin{array}{l}
\neg Ab1(x) \rightarrow \neg Flies(x)\\
Bird(x) \rightarrow Ab1(x)\\
Bird(x) \wedge \neg Ab2(x) \rightarrow Flies(x)\\
Penguin(x) \rightarrow Bird(x)\\
Penguin(x) \rightarrow Ab2(x)\\
Penguin(x) \wedge \neg Ab3(x) \rightarrow \neg Flies(x).
\end{array}$$

$Circ[A; (Ab1, Ab2, Ab3); (Flies)]$ lets us infer that the flying objects are the birds that aren't penguins.

Now add to A the assertions $Bat(x) \rightarrow Ab1(x)$ and $Bat(x) \wedge \neg Ab2(x) \rightarrow Flies(x)$ and do the circumscription again. The flying objects are now bats and the birds that are not penguins.

REIFYING PROPOSITIONS AND INDIVIDUAL CONCEPTS

...it seems that hardly anybody proposes to use different variables for propositions and for truth-values, or different variables for individuals and individual concepts.—(Carnap 1956, p. 113) [Church 1951, perhaps?]

- It is customary to assert the necessity, truth or knowledge of propositions in some form of modal logic, but modal logic is weaker than ordinary language which can treat concepts as objects.

- We propose abstract spaces of concepts to provide flexibility. Thus we can have $pp\ AAnd\ qq = qq\ AAnd\ pp$ when convenient. Expressions denoting concepts have doubled initial letters.

- Also human-level common sense needs functions from things to concepts of them. Here's an example.

$$Denot(NNumber(PPlanets)) = Number(Planets),$$
$$\neg Knew(Kepler, CComposite(NNumber(PPlanets))),$$
$$Knew(Kepler, CComposite(CConcept1(Denot(NNumbe$$

- Here $Denot(xx)$ is the thing xx denotes.

- We can also define $Exists(xx) \equiv (\exists x)Denotes(xx, x)$ so that $\neg Exists(PPegasus)$ asserts that Pegasus doesn't exist.

- $NNumber(PPlanets)$ is the concept of the number of planets, and $CConcept1(number)$ is a standard concept of that number.

APPROXIMATE OR PARTLY DEFINED CONCEPTS

- Humans language expresses and humans often think in terms of concepts that are only partly defined. Examples: the snow and rocks that constitute Mount Everest, the wants of the United States. $Wants(U.S., DDemocratic(IIraq))$. Mathematical concepts are an exception.

- Syntactically, approximate concepts are handled by weak axioms, e.g.

$$\ldots \rightarrow Wants(U.S., pp)$$

and

$$\ldots \rightarrow \neg Wants(U.S., pp).$$

- In general, there is no fact of the matter, even undiscovered, exactly characterizing $Wants(U.S., p$

- The semantic situation seems similar. In some interpretations $Wants(U.S., pp)$ is true, and in others it is false, but these needn't match up, although they shouldn't be contradictory. Defining the semantics of approximate concepts seems puzzling.

- A concept that is approximate in general, can be precise in a limited context. The barometer problem shows that.

CONTEXTS AS OBJECTS—1

- Everything a person says, writes, or thinks is in a context, and the meanings of what one says is relative to the context. Attempts to define terms free of context are usually incompletely successful outside mathematics.

- People switch from one context to another rather automatically. I propose contexts as objects—members of suitable abstract spaces.

- The are two main formulas. $Ist(c, pp)$ asserts that the proposition pp is true in the context c. $Value(c, tterm)$ gives the value of the individual concept $tterm$ in the context c. Using $Value$ requires that there be a domain associated with c.

CONTEXTS AS OBJECTS—2

- Ist can be compounded. We can have $Ist(c1, IIst(c2,$ and $Ist(c1, VValue(c2, term)\ Equals\ a)$.

- An alternative notation to $Ist(c, p)$ is

$$c : pp,$$

and likewise

$$c1 : cc2 : pp.$$

EXAMPLES

- Here are some semi-formal examples.

$Ist(Conan\ Doyle, DDetective(HHolmes))$
$Ist(U.S.medicalhistory, DDoctor(HHolmes))$
$Ist(U.S.literature, PPoet(HHolmes))$
$Value(U.S.literature, HHolmes) = Value(U.S.medicalh$
$Ist(U.S.legalhistory, JJudge(HHolmes))$
$Value(U.S.literature, HHolmes) = Father(Value(U.S.le$

- Here's an example of *lifting* a theory in which the predicates On and $Above$ have two arguments to a situation calculus theory in which they have three arguments. [An application of abstract group theory would provide bigger examples.]

On AND *Above* WITH TWO AND THREE ARGUMENTS

To describe the two argument $Above\text{-}theory$, we write

$Above\text{-}theory$:

$$(\forall xy)(On(x,y) \rightarrow Above(x,y)),$$

$$(\forall xyz)(Above(x,y) \wedge Above(y,z) \rightarrow Above(x,z)),$$

etc.

which stands for

$C0 : Ist(Above\text{-}theory, (\forall xy)(On(x,y) \rightarrow Above(x,y)))$

etc.

LIFTING *Above-theory*

We want to apply *Above-theory* in a context C in which On and $Above$ have a third argument denoting a situation. We have

C : $\qquad (\forall x\ y\ s)(On(x,y,s) \equiv Ist(C1(s), On(x,y))$

thus associating a context $C1(s)$ with each situation s. We also need

$C0$: $\qquad Ist(C, (\forall p\ \ s)(Ist(Above\text{-}theory, p) \rightarrow Ist$

which abbreviates to

C : $\qquad (\forall p\ s)(Ist(Above\text{-}theory, p) \rightarrow Ist(C1(s), p$

giving finally

C: $\qquad On(x,y,s) \rightarrow Above(x,y,s)$

APPROXIMATE ENTITIES CAN BE PRECISE IN LIMITED CONTEXTS

- Owning, buying and selling, e.g. of a house or a business, are such complicated concepts in general that a complete axiomatic theory is out of reach. However, reasonably complete theories are possible and used in limited contexts, e.g. while shopping in a supermarket.

$$InMarket(s) \wedge Ist(C(Market), Owns(x, Result(Buys(x$$
$$\rightarrow Owns(x, Result(Buys(x), s))$$

- The AI *drosophila* theories sampled above are also valid in limited contexts.

- The lifting relations between the sentences true in limited contexts and those valid in more general contexts need to be explored.

CONSCIOUSNESS AND SELF AWARENESS

- Much self awareness is simple enough not to require any extensions to logic, e.g. sensations of hunger. or of the positions of ones limbs.

- However, knowledge and belief, especially assertions of non-knowledge involve formulas analogous to reflexion principles. asserting truth.

- In discussing what self awareness a robot requires, I found it helpful to reify hopes, fears, promises, beliefs, what one thinks a concept denotes, intentions, prohibitions, likes and dislikes, its own abilities and those of others, and many more. The doctrine, common among philosophers and mathematicians, advocating minimizing the set of concepts, seems to me to be mistaken.

- When a human or robot needs to refer to the whole of its knowledge, the situation becomes more complicated, and there are possibilities for paradox, e.g. with

- $\neg Know(I, SSitting(PPresidentBush, NNow))$.

- $\neg Know(Putin, SSitting(PPresidentBush, NNow))$.

- Kraus, Perlis, and Horty treated formulas like the above expressing non-knowledge.

- One way of avoiding paradox may be to allow reference to ones knowledge up to the present time. This is analogous to the restricted comprehension principle.

COMMON SENSE IN MATHEMATICS

- In mathematical writing, the text between the formulas is essential to understanding the formulas. Introductions often contain no formulas. Understanding this text is mathematical common sense. Its formal expression should be more straightforward than the common sense of (say) history.

- —from Gödel, *Collected Works*, p. 147, we have

Similarly, proofs, from a formal point of view, are nothing but finite sequences of formulas (with certain specifiable properties). Of course, for metamathematical considerations, it does not matter what objects are chosen for primitive signs, and we shall assign natural numbers to this use. Consequently, a formula will be a finite sequence of natural numbers, and a proof array a finite sequence of finite sequences of natural numbers. The metamathematical notions (propositions) thus become notions (propositions) about natural numbers of sequences of them; therefore they can (at least in part) be expressed by the symbols of PM itself. In particular, it can be shown that the notions "formula", "proof", and "provable formula" can be defined in the system PM; that is, we can find a formula $F(v)$ with one free variable v (of the type of a number sequence) such that $F(v)$, interpreted according to the meaning of the terms of PM, says: v is a provable formula. We now construct an undecidable proposition of the system PM, that is, a proposition A for which neither A nor *not*-A is provable, in the following manner.

- False mathematical counterfactual: If $2^{2^5}+1$ were prime, twice it would be prime.

- "the notion that the continuum hypothesis is analogous to the parallel axiom", "Gödel's incompleteness theorems demolished Hilbert's program.", "Russell's first reaction to the paradox, which he discovered on reading Frege's work, was the 'vicious circle principle' which declared . . . meaningless",

AI RESEARCH ON COMMON SENSE IN LOGIC

- Much has been done to express common sense knowledge and reasoning in logic. However, present axiomatic AI theories require human modification whenever they are to be elaborated. Human-level AI systems must modify their own theories.

- There are biennial conferences on knowledge representation and also triennial workshops on common sense. CYC is a mostly proprietary database of more than a million common sense facts. expressed a syntactically sugared mathematical logic. Its reasoning facilities have proved difficult to use.

- Automatic theorem proving and interactive theorem proving have had considerable success in *bounded* mathematical and AI domains.

DIFFICULTIES AND CONCLUSIONS

- We will eventually have human-level logical AI.

- Sooner if we have help from logicians in devising ways of representing common sense theories and extending them.

- The above are almost surely not the only kinds of extensions to logic needed for dealing with common sense knowledge and reasoning.

- We discussed the following kinds of extensions. (1) Formal non-monotonic reasoning, (2) Reification, especially of concepts and contexts—and even theories, (3) Approximate entities without if-and-only-if definitions.

- There is a particular difficulty in extending a theory defined in a limited context to a more general context if the theory requires nonmonotonic reasoning, e.g. if the set of blocks is to be minimized.

- Human-level logical AI will also require language for expressing facts about methods effective in reasoning about particular subjects.

- Articles discussing these questions are available in http://www-formal.stanford.edu/jmc/.

STUFFY ROOM AXIOMS

Effect axioms:

$$
\begin{array}{l}
Blocked1(Result(Block1, s)) \\
Blocked2(Result(Block2, s)) \\
\neg Blocked1(Result(Unblock1, s)) \\
\neg Blocked2(Result(Unblock2, s)) \\
Stuffy(Result(Getstuffy, s)) \\
\neg Stuffy(Result(Ungetstuffy, s))
\end{array}
$$

Occurrence axioms:

$$
\begin{array}{l}
Blocked1(s) \wedge Blocked2(s) \wedge \neg Stuffy(s) \\
\rightarrow Occurs(Getstuffy, s) \\
(\neg Blocked1(s) \vee \neg Blocked2(s)) \wedge Stuffy(s) \\
\rightarrow Occurs(Ungetstuffy, s)
\end{array}
$$

AN ELABORATION GIVING OSCILLATING STUFFINESS

Suppose Bob is unhappy when the room is stuffy, but Alice is unhappy when the room is cold. The stuffy room axioms tolerate adding the following axioms which makes Vent1 oscillate between open and closed.

$$\begin{gathered}
Stuffy(s) \rightarrow Occurs(Does(Bob, Unblock1), s) \\
Unblocked1(s) \rightarrow Occurs(Getcold(Alice), s), \\
Cold(Alice, (Result(Getcold, s)), \\
Cold(Alice, s) \rightarrow Occurs(Does(Alice, Block1), s).
\end{gathered}$$

Alas, these axioms hold in a bounded domain. Common sense requires logic in which they inhabit an extendable context.