Available online at www.sciencedirect.com
ScienceDirect
Artificial Intelligence 171 (2007) 1174–1182
Artificial Intelligence
www.elsevier.com/locate/artint

From here to human-level AI

John McCarthy

Computer Science Department, Stanford University, Stanford, CA 94305, USA

Available online 10 October 2007

Abstract

Human-level AI will be achieved, but new ideas are almost certainly needed, so a date cannot be reliably predicted—maybe five years, maybe five hundred years. I'd be inclined to bet on this 21st century.

It is not surprising that human-level AI has proved difficult and progress has been slow—though there has been important progress. The slowness and the demand to exploit what has been discovered has led many to mistakenly redefine AI, sometimes in ways that preclude human-level AI—by relegating to humans parts of the task that human-level computer programs would have to do. In the terminology of this paper, it amounts to settling for a *bounded informatic situation* instead of the more general *common sense informatic situation*.

Overcoming the "brittleness" of present AI systems and reaching human-level AI requires programs that deal with the *common sense informatic situation*—in which the phenomena to be taken into account in achieving a goal are not fixed in advance.

We discuss reaching human-level AI, emphasizing logical AI and especially emphasizing representation problems of information and of reasoning. Ideas for reasoning in the common sense informatic situation include nonmonotonic reasoning, approximate concepts, formalized contexts and introspection.

Keywords: Human-level AI; Elaboration tolerance

1. What is human-level AI?

The first scientific discussion of human level machine intelligence was apparently by Alan Turing in the lecture [35]. The notion was amplified as a goal in [34], but at least the latter paper did not say what would have to be done to achieve the goal.

Allen Newell and Herbert Simon in 1954 were the first people to make a start on programming computers for general intelligence. They were over-optimistic, because their idea of what has to be done to achieve human-level intelligence was inadequate. The *General Problem Solver* (GPS) took general problem solving to be the task of transforming one expression into another using an allowed set of transformations.

Many tasks that humans can do, humans cannot yet make computers do. There are two approaches to human-level AI, but each presents difficulties. It isn't a question of deciding between them, because each should eventually succeed; it is more a race.

E-mail address: jmc@cs.stanford.edu.
URL: http://www-formal.stanford.edu/jmc/.

0004-3702/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.artint.2007.10.009

1. If we understood enough about how the human intellect works, we could simulate it. However, we don't have sufficient ability to observe ourselves or others to understand directly how our intellects work. Understanding the human brain well enough to imitate its function therefore requires theoretical and experimental success in psychology and neurophysiology.[1] See [28] for the beginning of the information processing approach to psychology.
2. To the extent that we understand the problems achieving goals in the world presents to intelligence we can write intelligent programs. That's what this article is about.
   Much of the public recognition of AI has been for programs with *a little bit of AI and a lot of computing*. This succeeded for chess and checkers and has so far failed for the game of go. Go requires the identification of subpositions that are analyzed separately first and then in interaction with each other. Human chess players also do this, but the chess programs don't. The price of the much greater computation this makes necessary has been affordable in chess but not in go. Computer speed bypasses many other heuristics that save humans enormous computation.

What problems does the world present to intelligence? More narrowly, we consider the problems it would present to a human scale robot faced with the problems humans might be inclined to relegate to sufficiently intelligent robots. The physical world of a robot contains middle sized objects about which its sensory apparatus can obtain only partial information quite inadequate to fully determine the effects of its future actions. Its mental world includes its interactions with people and also meta-information about the information it has or can obtain.

Our approach is based on what we call the *common sense informatic situation*, which we contrast with the *bounded informatic situation* that characterizes both formal scientific theories and almost all (maybe all) experimental work in AI done so far.

A formal theory in the physical sciences deals with a *bounded informatic situation*. Scientists decide informally in advance what phenomena to take into account. For example, much celestial mechanics is done within the Newtonian gravitational theory and does not take into account possible additional effects such as outgassing from a comet or electromagnetic forces exerted by the solar wind. If more phenomena are to be considered, a person must make a new theory. Probabilistic and fuzzy uncertainties can still fit into a bounded informatic system; it is only necessary that the set of possibilities (sample space) be bounded.

Most AI formalisms also work only in a bounded informatic situation. What phenomena to take into account is decided by a person before the formal theory is constructed. With such restrictions, much of the reasoning can be monotonic, but such systems cannot reach human level ability. For that, the machine will have to decide for itself what information is relevant. When a *bounded informatic system* is appropriate, the system must construct or choose a limited *context* containing a suitable theory whose predicates and functions connect to the machine's inputs and outputs in an appropriate way.[2] The logical tool for bounding the informatic situation is *nonmonotonic* reasoning.

2. The common sense informatic situation

Contention: *The key to reaching human-level AI is making systems that operate successfully in the common sense informatic situation.*

In general a thinking human is in what we call the *common sense informatic situation* [13]. It is more general than any *bounded informatic situation*. The known facts are incomplete, and there is no *a priori* limitation on what facts are relevant. It may not even be decided in advance what phenomena are to be taken into account. The consequences of actions cannot be fully determined. The *common sense informatic situation* necessitates the use of *approximate concepts* that cannot be fully defined and the use of *approximate theories* involving them. It also requires *nonmonotonic* reasoning in reaching conclusions.

The common sense informatic situation also includes some knowledge about the system's mental state.

[1] Recent work with positron emission tomography has identified areas of the brain that consume more glucose when a person is doing mental arithmetic. This knowledge will help build AI systems only when it becomes possible to observe what is going on in these areas during mental arithmetic.

[2] The textbook [4] puts it this way. "To get human-level computational intelligence it must be the agent itself that decides how to divide up the world, and which relationships to reason about".

A nice example of the common sense informatic situation is illustrated by an article in the *American Journal of Physics* some years ago. It discussed grading answers to a physics problem. The exam problem is to find the height of a building using a barometer. The intended solution is to measure the air pressure at the top and bottom of the building and multiply the difference by the ratio of the density of mercury to the density of air.

However, other answers may be offered. (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 measure the length of the shadow of the building. (3) Rappel down the building using the barometer as a measuring rod. (4) Lower the barometer on a string till it reaches the ground and measure the string. (5) Offer the barometer to the janitor of the building in exchange for information about the height. (6) Ignore the barometer, count the stories of the building and multiply by ten feet.

Clearly it is not possible to bound in advance the common sense knowledge of the world that may be relevant to grading the problem. Grading some of the solutions requires knowledge of the formalisms of physics and the physical facts about the earth, e.g. the law of falling bodies or the variation of air pressure with altitude. However, in every case, the physics knowledge is embedded in common sense knowledge. Thus before one can use Galileo's law of falling bodies $s = \frac{1}{2}gt^2$, one needs common sense information about buildings, their shapes and their roofs.

Bounded informatic situations are obtained by nonmonotonically inferring that only the phenomena that somehow appear to be relevant are relevant. In the barometer example, the student was expected to infer that the barometer was only to be used in the conventional way for measuring air pressure. For example, a reasoning system might do this by applying circumscription to a predicate *relevant* in a formalism containing also metalinguistic information, e.g. that this was a problem assigned in a physics course. Formalizing relevance in a useful way promises to be more difficult than just using existing *relevance logics.*

Common sense facts and common sense reasoning are necessarily imprecise. The imprecision necessitated by the common sense informatic situation applies to computer programs as well as to people.

Some kinds of imprecision can be represented numerically and have been explored with the aid of Bayesian networks, fuzzy logic and similar formalisms. This is in addition to the study of approximation in numerical analysis and the physical sciences.

3. The use of mathematical logic

What about mathematical logical languages?

Mathematical logic was devised to formalize precise facts and correct reasoning. Its founders, Leibniz, Boole and Frege, hoped to use it for common sense facts and reasoning, not realizing that the imprecision of concepts used in common sense language was often a necessary feature and not always a bug. The biggest success of mathematical logic was in formalizing purely mathematical theories for which imprecise concepts are unneeded. Since the common sense informatic situation requires using imprecise facts and imprecise reasoning, the use of mathematical logic for common sense has had limited success. This has caused many people to give up. Others devise extended logical languages and even extended forms of mathematical logic.

It is necessary to distinguish between mathematical logic and particular mathematical logical languages. Particular logical languages are determined by a particular choice of concepts and the predicate and function symbols to represent them. Failure to make the distinction has led some people to conclude that logic is inadequate when all they have shown is that a particular language is inadequate for some purpose. Different concepts and different predicate and function symbols might still succeed. In the words of the drive-in movie critic of Grapevine, Texas, "I'm surprised I have to explain this stuff."

The pessimists about logic or some particular set of predicates might try to prove a theorem about its inadequacies for expressing common sense.[3]

Since it seems clear that humans don't use logic as a basic internal representation formalism, maybe something else will work better for AI. Researchers have been trying to find this something else since the 1950s but still haven't succeeded in getting anything that is ready to be applied to the common sense informatic situation. Maybe they will eventually succeed. However, I think the problems listed in the later sections of this article will arise in any approach to human-level AI.

[3] Gödel's theorem is not relevant to this, because the question is not one of decideability or of characterizing truth.

Mathematical logic has been concerned with how people ought to think rather than how people do think. We who use logic as a basic AI formalism make programs reason logically. However, we have to extend logic and extend the programs that use it in various ways.

One important extension was the development of modal logic starting in the 1920s and using it to treat modalities like knowledge, belief and obligation. Modalities can be treated either by using modal logic or by reifying concepts and sentences within the standard logic. My opinion is that reification in standard logic is more powerful and will work better.

A second extension was the formalization of nonmonotonic reasoning beginning in the late 1970s—with circumscription and default logic and their variants as the major proposals. Nonmonotonic logic has been studied both as pure mathematics and in application to AI problems, most prominently to the formalization of action and causality. Several variants of the major formalisms have been devised.

Success so far has been moderate, and it isn't clear whether greater success can be obtained by changing the concepts and their representation by predicate and function symbols or by varying the nonmonotonic formalism.[4]

We need to distinguish the actual use of logic from what Allen Newell, [26] and [27], calls the logic level and which was also proposed in [8].

4. Approximate concepts and approximate theories

Other kinds of imprecision are more fundamental for intelligence than numerical imprecision. Many phenomena in the world are appropriately described in terms of *approximate concepts*. Although the concepts are imprecise, many statements using them have precise truth values. We offer two examples: the concept of Mount Everest and the concept of the welfare of a chicken. The exact pieces of rock and ice that constitute Mount Everest are unclear. For many rocks, there is no *truth of the matter* as to whether it is part of Mount Everest. Nevertheless, it is true without qualification that Edmund Hillary and Tenzing Norgay climbed Mount Everest in 1953 and that John McCarthy never set foot on it.

The point of this example is that it is possible and even common to have a solid knowledge structure from which solid conclusions can be inferred based on a foundation built on the quicksand of approximate concepts without definite extensions.

As for the chicken, it is clear that feeding it helps it and wringing its neck harms it, but it is unclear what its welfare consists of over the course of the decade from the time of its hatching. Is it better off leading a life of poultry luxury and eventually being slaughtered or would it be better off escaping the chicken yard and taking its chances on starvation and foxes? There is no *truth of the matter* to be determined by careful investigation of chickens. *When a concept is inherently approximate, it is a waste of time to try to give it a precise definition.* Indeed different efforts to define such a concept precisely will lead to different results—if any.

Most human common sense knowledge involves approximate concepts, and reaching human-level AI requires a satisfactory way of representing information involving approximate concepts. McCarthy in [22] discusses how to do it in logic. A simple approach would involve giving necessary conditions and sufficient conditions for a rock to be part of Mount Everest but not to try for necessary and sufficient conditions.

5. Nonmonotonic reasoning

Common sense reasoning is also imprecise in that it draws conclusions that might not be made if there were more information. Thus common sense reasoning is *nonmonotonic*. I will not go into the details of any of the proposals for handling nonmonotonic reasoning.

In particular, getting from the common sense informatic situation to a bounded informatic situation needs nonmonotonic reasoning.

[4] One referee for KR96 foolishly and arrogantly proposed rejecting a paper on the grounds that the inadequacy of circumscription for representing action was known.

6. Elaboration tolerance

Human abilities in the common sense informatic situation also include what may be called *elaboration tolerance*—the ability to elaborate a statement of some facts without having to start all over. Thus when we begin to think about a problem, e.g. determining the height of a building, we form a bounded context and try to solve the problem within it. However, at any time more facts can be added, e.g. about the precision with which the time for the barometer to fall can be estimated using a stop watch and also the possibilities of acquiring a stop watch.

McCarthy in [21] discusses about 20 elaborations of the Missionaries and Cannibals problem—mostly informally. Lifschitz in [6] formalizes nine of them.

7. Formalization of context

A third extension of mathematical logic involves formalizing the notion of context [16]. Notice that when theories are used in human communication and study, the theory is used in a context which people can discuss from the outside. If computers are to have this facility and are to work within logic, then the "outer" logical language needs names for contexts and sentences giving their relations and a way of entering a context. Clearly human-level AI requires reasoning about context.

Human-level AI also requires the ability to *transcend* the outermost context the system has used so far. Besides in [17], this is also discussed in [20].

Further work includes [1] and [2].

8. Reasoning about events—especially actions

Reasoning about actions has been a major AI activity, but this paper will not discuss my or other people's current formalisms, concentrating instead on the long range problem of reaching human level capability. We regard actions as particular kinds of events and therefore propose subsuming reasoning about actions under the heading of reasoning about events.

Most reasoning about events has concerned determining the effects of an explicitly given sequence of actions by a single actor. Within this framework various problems have been studied.

- The frame problem concerns not having to state what does not change when an event occurs.
- The qualification problem concerns not having to state all the preconditions of an action or other event. The point is both to limit the set of preconditions and also to jump to the conclusion that unstated others will be fulfilled unless there is evidence to the contrary. For example, wearing clothes is a precondition for airline travel, but the travel agent will not tell his customer to be sure to wear clothes.
- The ramification problem concerns how to treat side-effects of events other than the principal effect mentioned in the event description.

Each of these involves elaboration tolerance, e.g. adding descriptions of the effects of additional events without having to change the descriptions of the events already described. When I wrote about applications of circumscription to formalizing common sense [11], I hoped that a *simple abnormality theory* would suffice for all of them. That didn't work out when I tried it, but I still think a common nonmonotonic reasoning mechanism will work. Costello in [3] argues that simple abnormality theories have the same expressive power as more elaborate nonmonotonic formalisms that have been proposed.

Human level intelligence requires reasoning about strategies of action, i.e. action programs. It also requires considering multiple actors and also concurrent events and continuous events. Clearly we have a long way to go.

Some of these points are discussed in [19] which concerns elaborating a simple narrative involving two actors.

9. Introspection and self-awareness

People have a limited ability to observe their own mental processes. For many intellectual tasks introspection is irrelevant. However, it is at least relevant for evaluating how one is using one's own thinking time. Human-level AI will require introspective ability. In fact programs can have more than humans do, because they can examine themselves, both in source and compiled form and also reason about the current values of the variables in the program. McCarthy in [20] discusses this in some detail.

That robots also need introspection is argued, and how to do it is also discussed in [20].

10. Heuristics

The largest qualitative gap between human performance and computer performance is in the area of heuristics, even though the gap is disguised in many applications by the millions-fold speed advantage of computers. The general purpose theorem proving programs run very slowly, and the special purpose programs are very specialized in their heuristics.

I think the problem lies in our present inability to give programs domain and problem dependent heuristic advice. In [7] I advertised that the Advice Taker would express its heuristics declaratively. For more than 40 years I did not know how to do it.

Recently Josefina Sierra-Santibanez made important theoretical and experimental progress in controlling planning with declaratively expressed heuristics. The work is reported in [32] and [33] which show the advantage of using declarative heuristics over other approaches, specifically in blocks world problems.

Here are some ideas I hope are relevant starting with a simple example and then introducing the more general idea but still limited idea of *postponable variables*.

McCarthy in [10] discusses coloring maps with four colors in terms of modifying a Prolog program for map coloring. It is simply exemplified in terms of coloring a map of the United States. Note that California (CA) has only three neighbors—Arizona (AZ), Nevada (NV), and Oregon (OR). Therefore, we *postpone* coloring CA the resulting *reduced map* has been colored. Then we can always find a color for CA, different from those used to color AZ, NV, and OR. The process of reducing the map can be continued, because once CA has been removed, AZ has only three neighbors. Eventually we get a *completely reduced map* in which every state has four or more neighbors. It turns out that the completely reduced map of the USA is empty. Postponement thereby eliminates tree search from the problem of coloring the map of the US.

In fact the completely reduced maps of Europe and Asia, South America, Africa, and the departments of France are also empty. I have not found any actual political maps on the earth that don't reduce completely, even though it is not difficult to make an artificial map in which each country has four or more neighbors.[5]

10.1. Postponable variables

Map coloring provides a nice example of the notion of *postponable variable* in constraint satisfaction problems.

Definition 1. A *constraint satisfaction problem CSP* consists of a set *VV* of variables and a set *C* of constraint relations among the variables that are required to be satisfied.

Definition 2. A variable v in *CSP* is postponable iff no matter how the constraints $C \setminus \mathit{Constraints\text{-}involving}(v)$ that do not involve the variable are satisfied, there is always a value for v that conjoined to the values of the variables satisfies all the constraints.

Here's a formula expressing the notion of a variable v being postponable in a constraint satisfaction problem *dcsp*. In (1) the variables of *csp* are represented by constants, e.g. in coloring the US. *CA* is a constant representing a variable *ca*. We use *vv* as a meta-variable ranging over the variables of *csp*.

$$\begin{aligned}&\mathit{Postponable}(vv, csp) := (\forall as)(\mathit{Satisfies}(as, \mathit{Remove}(vv, csp))\\&\quad \rightarrow (\exists x \in \mathit{Domain}(csp))\mathit{Satisfies}(as + \mathit{Set}(vv, x), csp)).\end{aligned} \tag{1}$$

[5] McCarthy [10] also discusses postponing countries with four or fewer neighbors and how to use the topology to fix the situation that arise when it is necessary to color a country all four of whose neighbors have been assigned distinct colors.

Here $Remove(vv, csp)$ is the constraint satisfaction problem obtained from csp by removing all constraints involving vv. $Satisfies(as, csp)$ says that the assignment as of values to variables satisfies the constraint problem csp. $Domain(csp)$ is the domain over which the variables of csp range. $as + Set(v, x)$ is the assignment of values to variables obtained from as by adding the assignment of x to the variable that is the value of vv.

Notice that the concepts involved in defining *Postponable* are all metamathematical with respect to the original constraint satisfaction problem. The ability to work at the metalevel will be essential to human-level AI.

I hope that the methods described by Sierra will be applicable to expressing the use of postponability in constraint satisfaction problems.

11. Psychological, social and political obstacles

The computer science world is still suffering from a 1990s fit of pseudo-practicality that is inimical to the solution of difficult scientific problems. Lip service is given to basic research, and a lot of basic research is done, but the initiation of ambitious research by young people is hampered by the now prevalent doctrine that "basic research" should be done in connection with applied problems that have been identified by the competent committees. I think that Newell and Minsky and I would have had a much harder time initiating AI research if the atmosphere of the 1950s had been like that of the 1990s.

Computer science suffers more than older fields from this disease, one of the main carriers of which was the Computer Science and Telecommunications Board of the National Research Council. Its worst sin was merging computer science with computer engineering in its harmful, narrow-minded report *Computing the Future* [30].

Nevertheless, the main problem in reaching human level AI is not the politics of science and technology but the intrinsic difficulty of the scientific problems.

11.1. Philosophical and ideological objections to AI

From the beginning, the idea of artificial intelligence has encountered philosophical and ideological objections. Turing [34] deals with many of them, properly dividing the objectors into those who think human-level performance is impossible and those who think that a machine wouldn't be *really intelligent* even if you couldn't distinguish its performance from that of a human.

Hubert Dreyfus [5] is in the first category, John Searle [31] is in the second. Roger Penrose [29] took an intermediate position, arguing from Gödel's incompleteness theorems that a computer built in accordance with the present laws of physics couldn't be intelligent, but appealing to ideas of quantum gravity that he couldn't specify that material systems, e.g. humans, could be intelligent. See [15] and [18] discuss Penrose's arguments and related questions.

Some of the objections to AI have been frankly religious. However, as with objectors to evolution, the religious objectors to AI use whatever material they can find from Dreyfus, Searle, Penrose and others. AI researchers should feel relieved that the attacks on AI have been nowhere near as fierce as those on evolution. Maybe it's because human-level AI is not discussed in high school texts on programming and computer science.

I don't want to go further into the objections in this article.

12. Conclusion

Between us and human-level intelligence lie many problems. They can be summarized as that of succeeding in the *common sense informatic situation*.

The problems include:

common sense knowledge of the world Many important aspects of what this knowledge is in and how it can be represented are still unsolved questions. This is particularly true of knowledge of the effects of actions and other events.

epistemologically adequate languages These are languages for expressing *what a person or robot can actually learn about the world*. McCarthy and Hayes [25] discusses epistemological adequacy, note that a language expressing facts in terms of positions and velocities of molecules is not adequate.

concepts regarded as objects Humans do it. McCarthy [9] formalizes it.

elaboration tolerance What a person knows can be elaborated without starting all over. See [21].
nonmonotonic reasoning Perhaps new systems are needed, but maybe it is only a question of circumscribing the right formulas, varying the right predicates and functions. See [12].
contexts as objects This subject is just beginning. See the references of Section 7.
introspection AI systems will need to examine their own internal states. See [20].
action The present puzzles of formalizing action should admit a uniform solution. A new version of situation calculus is discussed in [23].

I doubt that a human-level intelligent program needs structures corresponding to all these entities and to the others that might have been listed. A generally intelligent logical program probably needs only its monotonic and nonmonotonic reasoning mechanisms plus mechanisms for entering and leaving contexts. The rest are handled by particular functions and predicates.

To what extent will all these problems have to be faced explicitly by people working with neural nets and connectionist systems? The systems I know about are too primitive for the problems even to arise. However, more ambitious systems will inhabit the common sense informatic situation. They will have to be elaboration tolerant and will require some kind of mental model of the consequences of actions.

Many will find dismayingly large the list of tasks that must be accomplished in order to reach human-level logical intelligence. Perhaps fewer but more powerful ideas would simplify the list. Others will claim that a system that evolves intelligence as life does will be more straightforward to build. Maybe, but the advocates of that approach have been at it as long as we have and still aren't even close.

So it's a race.

It will be much more scientifically satisfying to understand human level artificial intelligence logically than just achieve it by a computerized evolutionary process that produced an intelligent but incomprehensible result. In fact, the logical approach would be worth pursuing even if the intellectually lazier evolutionary approach won the race.

12.1. What will humanity do with human-level AI?

Here is where science fiction has had more to say than science, but it fictionalizes the science for literary objectives and social agendas. These have mainly involved assuming human-like personalities including good and bad behavior and suffering persecution and/or psychological problems like humans. There is no need to give AI systems human-like personalities, and they should be designed so as to excite neither fear nor sympathy.

When human-level AI becomes available, it is unlikely to be the monopoly of any group or government. Many individuals and groups will use it to advance their goals.

As early as 1970 some wanted international regulation of AI. This was a bad idea then and will remain a bad idea as long as we have so little understanding of what form human-level AI will take.

I'd like to have a robot servant, and if I had access to a human-level system I'd ask it for help in understanding the consequences for humanity of various policies about AI.

Acknowledgements

In writing the 1996 version of this article, I got useful suggestions from Eyal Amir, Saša Buvač, and Tom Costello.
This work was partly supported by ARPA (ONR) grant N00014-94-1-0775.
Some additional relevant papers are in my book [14] and on my Web site http://www-formal.stanford.edu/jmc/.

References

[1] S. Buvač, Quantificational logic of context, in: Proceedings of the Thirteenth National Conference on Artificial Intelligence, 1996.
[2] S. Buvač, V. Buvač, I.A. Mason, Metamathematics of contexts, Fundamenta Informaticae 23 (3) (1995).
[3] T. Costello, The expressive power of circumscription, Artificial Intelligence 104 (1–2) (1998) 313–329.
[4] A.M. David Poole, R. Goebel, Computational Intelligence, Oxford, 1998.
[5] H. Dreyfus, What Computers Still Can't Do, MIT Press, 1992.
[6] V. Lifschitz, Missionaries and cannibals in the causal calculator, in: A.G. Cohn, F. Giunchiglia, B. Selman (Eds.), KR2000: Principles of Knowledge Representation and Reasoning, Proceedings of the Seventh International Conference, Morgan Kaufman, 2000, pp. 85–96.

[7] J. McCarthy, Programs with common sense, in: Mechanisation of Thought Processes, Proceedings of the Symposium of the National Physics Laboratory, Her Majesty's Stationery Office, London, UK, 1959, pp. 77–84, http://www-formal.stanford.edu/jmc/mcc59.html. Reprinted in [14].
[8] J. McCarthy, Ascribing mental qualities to machines, in: M. Ringle (Ed.), Philosophical Perspectives in Artificial Intelligence, Harvester Press, 1979, http://www-formal.stanford.edu/jmc/ascribing.html. Reprinted in [14].
[9] J. McCarthy, First order theories of individual concepts and propositions, in: D. Michie (Ed.), Machine Intelligence, vol. 9, Edinburgh University Press, Edinburgh, 1979, http://www-formal.stanford.edu/jmc/concepts.html. Reprinted in [14].
[10] J. McCarthy, Coloring maps and the Kowalski doctrine, http://www-formal.stanford.edu/jmc/coloring.html, Technical Report STAN-CS-82-903, Dept Computer Science, Stanford University, 1982, AIM-346. Reprinted in [14].
[11] J. McCarthy, Applications of circumscription to formalizing common sense knowledge, Artificial Intelligence 28 (1986) 89–116, http://www-formal.stanford.edu/jmc/applications.html. Reprinted in [14].
[12] J. McCarthy, Applications of circumscription to formalizing common sense knowledge, Artificial Intelligence 28 (1986) 89–116, http://www-formal.stanford.edu/jmc/applications.html. Reprinted in [14].
[13] J. McCarthy, Artificial intelligence, logic and formalizing common sense, in: R. Thomason (Ed.), Philosophical Logic and Artificial Intelligence, Kluwer Academic, 1989, http://www-formal.stanford.edu/jmc/ailogic.html.
[14] J. McCarthy, Formalizing Common Sense: Papers by John McCarthy, Ablex Publishing Corporation, 1990.
[15] J. McCarthy, Review of "The Emperor's New Mind" by Roger Penrose, Bulletin of the American Mathematical Society 23 (2) (1990) 606–616.
[16] J. McCarthy, Notes on formalizing context, in: IJCAI93, http://www-formal.stanford.edu/jmc/context.html, 1993.
[17] J. McCarthy, Notes on formalizing context, in: IJCAI-93, http://www-formal.stanford.edu/jmc/context.html, 1993.
[18] J. McCarthy, Review of "Shadows of the Mind" by Roger Penrose, http://www-formal.stanford.edu/jmc/penrose2.html, 1995, Psyche, an Electronic Journal at http://psyche.cs.monash.edu.au/.
[19] J. McCarthy, Situation calculus with concurrent events and narrative, http://www-formal.stanford.edu/jmc/narrative.html, 1995, web only, partly superseded by [24].
[20] J. McCarthy, Making robots conscious of their mental states, in: S. Muggleton (Ed.), Machine Intelligence, vol. 15, Oxford University Press, 1996, appeared in 2000, http://www-formal.stanford.edu/jmc/consciousness.html. The web version is improved from that presented at Machine Intelligence 15 in 1995.
[21] J. McCarthy, Elaboration tolerance, http://www-formal.stanford.edu/jmc/elaboration.html, 1999, web only for now.
[22] J. McCarthy, Approximate objects and approximate theories, in: A.G. Cohn, F. Giunchiglia, B. Selman (Eds.), KR2000: Principles of Knowledge Representation and Reasoning, Proceedings of the Seventh International conference, Morgan Kaufman, 2000, pp. 519–526, http://www.formal.stanford.edu/jmc/approximate.html.
[23] J. McCarthy, Actions and other events in situation calculus, in: A.G. Cohn, B. Selman (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Eighth International Conference (KR2002), Morgan Kaufmann, 2002, http://www-formal.stanford.edu/jmc/sitcalc.html.
[24] J. McCarthy, T. Costello, Combining narratives, in: Proceedings of Sixth Intl. Conference on Principles of Knowledge Representation and Reasoning, Morgan Kaufman, 1998, pp. 48–59.
[25] J. McCarthy, P.J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: B. Meltzer, D. Michie (Eds.), Machine Intelligence, vol. 4, Edinburgh University Press, 1969, pp. 463–502, http://www-formal.stanford.edu/jmc/mcchay69.html, reprinted in [14].
[26] A. Newell, The knowledge level, AI Magazine 2 (2) (1981) 1–20, originally delivered as the Presidential Address, American Association for Artificial Intelligence, AAAI80, Stanford, CA, August 1980.
[27] A. Newell, Reflections on the knowledge level, Artificial Intelligence 59 (1–2) (1993) 31–38.
[28] A. Newell, H.A. Simon, Human Problem Solving, Prentice–Hall, Englewood Cliffs, NJ, 1972.
[29] R. Penrose, The Emperor's New Mind: Concerning Computers, Minds, and the Law of Physics, Oxford Univ. Press, 1989, ACM CR 9008-0655.
[30] C. Science, N.R.C. Telecommunications Board, Computing the Future: A Broader Agenda for Computer Science and Engineering, National Academies Press, 1992.
[31] J.R. Searle, Minds, Brains, and Science, Harvard University Press, Cambridge, MA, 1984.
[32] M. Sierra, Declarative formalization of reasoning strategies: A case study on heuristic nonlinear planning, Annals of Mathematics and Artificial Intelligence 39 (1–2) (2003) 61–100.
[33] M. Sierra, Heuristic planning: A declarative approach based on strategies for action selection, Artificial Intelligence 153 (1–2) (2004) 307–337.
[34] A. Turing, Computing machinery and intelligence, Mind 21 (1950).
[35] A.M. Turing, Lecture to the London mathematical society, in: The Collected Works of A.M. Turing, vol. Mechanical Intelligence, North-Holland, 1947. This was apparently the first public introduction of AI, typescript in the King's College archive, the book is 1992.