ELSEVIER
Artificial Intelligence 77 (1995) 371-392
Artificial Intelligence

Book Review

An epistemological science of common sense ☆

Fausto Giunchiglia [a,b,*]
[a] *Mechanized Reasoning Group, IRST, Povo 38100, Trento, Italy*
[b] *University of Trento, Trento, Italy*

Received February 1995

I started with the goal of writing a review of the book *Formalizing Common Sense: Papers by John McCarthy*, a collection of seventeen papers, listed in the bibliography from [2] to [18] in the same order and numbering. (This collection contains also an introductory overview by Vladimir Lifschitz, referenced as [1].) This goal seemed quite challenging for at least two reasons. First, this collection contains McCarthy's most influential papers in Artificial Intelligence (AI). McCarthy is one of the founders of AI and some of the papers in the collection, e.g. [2,3,7,9], have shaped the field and have been a source of inspiration for many researchers. Second, the papers in the collection have been written over a long period of time, i.e. between 1959 and 1988. Providing a uniform and global (as opposed to paper by paper) perspective on this work was something I was not sure would be possible.

The second worry turned out to be misplaced. A careful reading soon revealed the many connections between the different papers and the fact that all of them are contributions towards the same research project, i.e. the development of an *advice taker* [2]. It also revealed the distinct flavour that all these papers have, and how much they have influenced AI. Finally, the organization of the papers in the collection, which is chronological, clearly highlighted the development and refinement over time of such a project, starting from the 1959 paper "Programs with Common Sense" [2]. These discoveries shed a new light on McCarthy's work. They convinced me that the identification and development of such a research project is an even more important contribution than the individual papers themselves; and that this had to be the topic of the review.

This review attempts to describe the advice taker project as it emerges from the papers of the collection: its evolution over time, what has been achieved, what remains to be done, what in my opinion are its strengths and its limitations. My criticisms are

☆ A review of Vladimir Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990); vi + 256 pages, hardback, ISBN 0-89391-535-1 (Library of Congress: Q335.M38 1989).

* E-mail: fausto@irst.itc.it.

SSDI 0004-3702(95)00066-6

not meant to be exhaustive; they are more the natural result of my taste and know how. This review hardly reports the many controversial discussions about some basic assumptions underlying McCarthy's research. I have decided to avoid them as I am mostly in agreement with McCarthy's perspective. I am therefore not the right person for raising such issues. Also, and more important, I thought that an organic presentation and discussion of the project would be more interesting, to write and to read, than a metadiscussion about why it should fail or succeed. Finally, as a matter of style, the reader will notice that this review makes extensive use of quotations. The quotations explicitly separated from the text report concepts which are crucial for the development of the review (and, in my opinion, of McCarthy's research project).

1. An epistemology of common sense

1.1. Common sense

McCarthy clearly states his long term research goal in his 1959 paper [2]. Indeed, the title itself, "Programs with Common Sense", describes it. The following citation identifies what characterizes a program with common sense.

> The *advice taker* is a proposed program for solving problems by manipulating sentences in formal languages. ...
>
> The main advantages that we expect the *advice taker* to have is that its behavior will be improvable merely by making statements to it, telling it about its symbolic environment and what is wanted from it. To make these statements will require little if any knowledge of the program or the previous knowledge of the *advice taker*. One will be able to assume that the *advice taker* will have available to it a fairly wide class of immediate consequences of anything it is told and its previous knowledge. This property is expected to have much in common with what makes us describe certain humans as having *common sense*. We shall therefore say that *a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows*. [2, p. 9] (The page numbers are always those of the collection.)

The above quotation introduces many important ideas. First, it suggests that a program or person has common sense if it is able to perform in the same way and as effectively as (certain) humans. This might seem to imply that the need for common sense is a truism, as the usefulness of a program without common sense becomes unclear. McCarthy discusses this issue and further qualifies his research project in [15]. Here he points out that many useful programs have been and can be written which lack (some forms of) common sense. For instance, as we can infer from [15], a program which reasons about other people's beliefs does not necessarily need common sense about the world's evolution in time. Less obviously, a program which reasons about the world's evolution in time does not necessarily need common sense about the world's evolution in time. Common sense is needed only for particular domains, and only in order to achieve advice taker level performance. From [15], two examples of useful programs without common sense are chess programs and MYCIN [19,20]. However, as McCarthy points out, this is not without consequences. MYCIN lacks common sense knowledge about the world's evolution in time, a form of common sense which is necessary for its correct use. For instance, MYCIN will allow a patient to die of cholera while trying to kill the bacteria. To make things worse, MYCIN is unable to understand its own limitations, and very hard to extend beyond the goals for which it had been designed. In my interpretation of McCarthy's position, programs without enough common sense, e.g. MYCIN, can still be successfully used if the missing common sense is provided by the user, with MYCIN, by the physician. It is interesting to notice that the absence of common sense does not seem to create problems with chess programs. My explanation is that, though these programs do not have common sense, they have enough of it. We have very little or no common sense about the chess world (commonsensical consequences are something which should be "easy" to compute); chess programs have as much common sense as we do.

Second, the above quotation suggests that a person or program has common sense if it is said to be so by some outside observer (in the quotation, us). This leads to a subjectivistic notion of common sense. Different people at different points in time have different, although perhaps related, common sense. Even the same person in different points in time can have different common sense. Consider for instance the chess world. Suppose that we were chess masters and that we were able to compute almost effortlessly moves beyond the capabilities of a chess program. Then, once we accepted McCarthy's notion of common sense, we would be ready to say that the chess program is without common sense (about the chess world). Vice versa, anybody who does not know "enough" about cholera will not think that MYCIN does not have common sense. This subjectivistic view of common sense might endanger McCarthy's program as it is unclear how it is possible to reproduce what is impossible to define. McCarthy never discusses this problem; instead, his attitude has been that of stating what common sense is and of proposing ways to formalize it.

Finally, the above quotation implicitly characterizes the three basic properties of common sense. A program or person has common sense if

Property 1: it knows a sufficiently large set of facts of the environment where it lives;

Property 2: it can increment what it knows by automatically deducing a sufficiently large set of immediate consequences from it;

Property 3: it can increment what it knows by being told.

These three properties, which are also discussed at length in [3,6,7], are at the core of McCarthy's research. About property 3 McCarthy writes in [14], "What we can learn from natural language is not how to express in English what we already know how to express in computerese. Rather, we must study those ideas expressible in natural language that no one knows how to represent at all in a computer" [p. 188]. This issue is taken on in [12,5]. In [12] McCarthy describes some of the desiderata for the "common business communication language" (CBCL), a common language for expressing business communications, e.g. requests for price quotations, inquiries about the status of delayed orders and references to standard commercial legal agreements. The main message is that the most difficult problems in defining a language like CBCL, or even natural language front ends to computer programs, involve deciding what is to be sayable. In [5] he discusses the problems which are involved in understanding a real-world piece of text, an excerpt from the New York Times about a salesman robbed by three unknown persons. He gives a list of questions that an intelligent person or program should be able to answer, based on the information provided by the New York Times article. In order to achieve these objectives, he introduces ANL, an Artificial Natural Language, which is discussed with an approach similar to that used in [12] to describe CBCL.[1] He then gives hints about how to express in ANL the knowledge stated in the article with the further complication, with respect to CBCL, of considering how to reason about it (i.e. how to achieve at the same time properties 2 and 3).

Properties 1 and 2 distinguish between what the advice taker knows and what allows it to deduce new facts from what it knows. This is explicitly stated and discussed in [15] as a distinction between *common sense knowledge* and *common sense reasoning*. A very similar distinction is also discussed in [3], even if for what concerns intelligence, as a distinction between the epistemological and the heuristic part of intelligence.[2] In particular, McCarthy writes: "The epistemological part is the representation of the world

[1] Another early example of natural language understanding based on the use of an internal (conceptual) language can be found in [21]. It is not clear whether there has been any, possibly mutual, influence between the two proposals.

[2] McCarthy sometimes talks of common sense, e.g. in [2], and sometimes of intelligence, e.g. in [3]. This bouncing back and forth between the two notions is pervasive. In the papers of the collection McCarthy never tries to study the relation between intelligence and common sense. However in [3], where he talks of intelligence for the first time, he writes

> The idea of an intelligent machine is old, but serious work on the artificial intelligence problem, or even serious understanding of what the problem is, awaited the stored program computer. ...
> We shall, therefore, be interested in an intelligent entity that is equipped with a representation or model of the world. On the basis of this representation a certain class of internally posed questions can be answered, not always correctly. ...
> ... we shall say that an entity is intelligent if it has an adequate model of the world (including the intellectual world of mathematics, understanding of its own goals, and other mental processes), if it is clever enough to answer a wide variety of questions on the basis of the model, if it can get additional information from the external world when required, and can perform such tasks in the external world as its goals demand and its physical abilities permit. [p. 23]

A careful comparison between the quotation from p. 9 and that from p. 23 suggests that the property of being intelligent can be characterized by the same three properties introduced above for common sense. This perspective, whose correctness is not an issue in this paper, motivates McCarthy's attitude (and that taken by us here) of using the notions defined for common sense for intelligence and vice versa. It is important to notice that the two notions are not collapsed. The main difference seems that intelligence requires that the associated capabilities be "good enough" to achieve a goal. Thus the first requirement for having common sense is to know about the surrounding environment, while being intelligent requires that such a representation be adequate (property 1). The second requirement for having common sense is being able to derive consequences from what is known, most often without a strong reasoning capability and possibly without involving any intelligence [2]. Intelligence requires instead the capability of answering "a wide variety of questions", possibly on difficult topics, e.g. mathematics and people's mental processes (property 2). Finally, the third requirement for common sense is the possibility to provide data to the advice taker without knowing of its actual internal state and functioning. An intelligent entity can instead actively use its capabilities to get information from the outside and to act as needed to satisfy a goal (property 3). However, as far as I know, these differences have not played any role in McCarthy's research.

in such a form that the solution of problems follows from the facts expressed in the representation. The heuristic part is the mechanism that on the basis of the information solves the problem and decides what to do" [3, pp. 23–24].

Most of the research in AI has actually concentrated on the heuristic part of intelligence and has considered the epistemological part only for what is needed in order to successfully achieve problem solving. McCarthy's main interests and most important contributions are instead in the epistemological part only, independently of the problem solving which needs to be done. The reasons for this choice are explained in [6]. In particular McCarthy writes:

> Considering epistemological problems separately has the following advantages:
> (1) The same problems of what information is available to an observer and what conclusions can be drawn from information arise in connection with a variety of problem-solving tasks.
> (2) A single solution of the epistemological problems can support a wide variety of heuristic approaches to a problem.
> (3) AI is a very difficult scientific problem, so there are great advantages in finding parts of the problem that can be separated out and separately attacked.
> (4) ... it is quite difficult to formalize the facts of common knowledge. Existing programs that manipulate facts in some of the domains are confined to special cases and don't face the difficulties that must be overcome to achieve intelligent behavior. [6, p. 77]

1.2. Epistemological adequacy

McCarthy's first step in the study of the epistemological part of intelligence is to classify the representation formalisms according to three notions of adequacy. The quotation below reports their informal definition, as first given in [3].

> A representation is called *metaphysically adequate* if the world could have that form without contradicting the facts of the aspect of reality that interests us. ...
>
> A representation is called *epistemologically adequate* for a person or machine if it can be used practically to express the facts that one actually has about the aspect of the world. ...
>
> A representation is called *heuristically adequate* if the reasoning processes actually gone through in solving a problem are expressible in the language. [pp. 27–28]

In McCarthy's view, a full treatment of common sense and its implementation inside the advice taker is crucially based on the ideas of epistemological adequacy and heuristic adequacy (see [3,6,7]). McCarthy does not further elaborate the idea of heuristic adequacy. In my opinion, the requirement of heuristic adequacy really amounts to requiring epistemologically adequate representations of reasoning. It is a derived notion, and therefore less important than the notion of epistemological adequacy, though crucial to the implementation of the advice taker. He further discusses the idea of metaphysical adequacy in [7], but only to point out that it is a necessary requirement for scientific theories, and that it does not seem enough if one wants to implement common sense inside an artifact. As he writes, "The skeptic who doubts whether there is anything to say about the world apart from the particular sciences should try to write a computer program that can figure out how to get to Timbuktoo, taking into account, not only the facts about travel in general, but also the facts about what people and documents have what information, and what information will be required at different stages of the trip and when and how it is to be obtained. He will rapidly discover that he ... will be unable to express and build into a program 'what everybody knows'." [7, pp. 114–115].

McCarthy's main interests are in fact concentrated on epistemological adequacy. Most of McCarthy's technical work, e.g. the work described in [5–9,16,17], aims at developing epistemologically adequate theories of common sense. He presents and discusses the idea of epistemological adequacy in various other places, see for instance [3, p. 30], [6, p. 70], [7, p. 93] and [13, p. 181]. Interestingly, all these discussions have a slightly different emphasis. However it appears quite clear that epistemological adequacy poses two important constraints on representations, both constraints being implicit in the word *practically* in the quotation from pp. 27–28. The first is ontological:[3] the representation language must represent the *common sense informatic situation*, i.e. it must represent information at the level of abstraction used by people during their everyday activities (see [4] for more detail about McCarthy's notion of information).[4] This requirement seems appropriate as it allows us to build machines able, for instance, to interact meaningfully with humans (e.g. to get information by being told) without relying on ad hoc or hardwired mechanisms. We cannot teach a program what it does not know how to represent. It is not easy to see how a representation of the world as a collection of particles interacting by forces between each pair of particles, or as a giant quantum-mechanical wave function, can be used to implement common sense inside a machine (as stated in [3], both these representations are metaphysically but not epistemologically adequate). The second constraint is theoretical: the theories (axioms and inference rules) we develop must have a form which makes it possible to acquire the "necessary" (common sense) knowledge and to perform the "necessary" (common sense) reasoning (whatever we take "necessary" to mean, see Section 1.1). Thus for instance an epistemologically adequate representation must allow the computer to represent "the information actually available to a subject under given circumstances" [7, p. 93], while an epistemologically inadequate formalism would not allow it to "achieve enough goals requiring general intelligence no matter how fast [it was] allowed to run" [6, p. 79]. Not only must a representation schema be based on a correct ontology. It must also use theories of the world which actually represent the available knowledge and in a way to allow the advice taker to compute the facts needed to achieve a goal.

In my opinion, epistemological adequacy is McCarthy's most important defined notion, and also the most crucial to his research program. The requirement for epistemological

[3] In this paper we use the word "ontology" with a meaning different from that used by philosophers. According to philosophical usage, ontology is about what kinds of things there are [22]. Here we use this word to refer to what kinds of things there are in the common sense world. This shift of meaning is consistent with our goals, and also justifiable, given that we are not interested in the philosophers' question.

[4] The term "common sense informatic situation" is first introduced in [23].

adequacy implicitly fixes the kind of formalisms and theories of common sense which must be developed and, at the same time, rules out a lot of possible options. For instance, it distinguishes AI from neurophysiology or other biology level approaches. (This view has been and is still very controversial, see for instance [24,25].) In particular, in [4, p. 66] McCarthy points out that AI and neurophysiology will fruitfully interact (but not collapse!) once the neurophysiologists will be able to compare the AI epistemological models with physiological data. The notion of epistemological adequacy is more compatible with psychology and in [4] McCarthy describes the many connections between the two disciplines. The distinction rises because of the need for generality implicit in the notion of epistemological adequacy. In [18] McCarthy points out that in psychology the subject of study is the realization of certain phenomena in humans, while AI is concerned with methods for achieving goals in general, independently of "whether the problem solver is a human, a Martian, or a computer program" [p. 246]. The requirement for epistemological adequacy yields (and, historically, has yielded) the development of an epistemic level theory of common sense (or using a similar notion introduced in [26], of a knowledge level theory of common sense). In my opinion, an appropriate name for such a theory of common sense is an *epistemology of common sense*.

2. A science of common sense

McCarthy's interest is in a *science of common sense*, that is, in a systematic study of how common sense appears in nature, and on the formulation of laws which describe it in general terms. This emerges pretty clearly in many papers, e.g. in [5,7,14], and in particular in [4]. This last paper is a reaction to Sir James Lighthill's report on AI, commissioned by the Science Research Council of Great Britain. As reported in [4], Lighthill classified the research in AI in three categories: category [A] was labeled as *advanced automation* or *applications*, category [C] comprised the studies of the central nervous system, while category [B] was defined as *building robots* and *bridging* between the other two categories. Lighthill's classification implied that the research in category [B], where most of the research in AI was, was worthwhile only for its contribution to the other two categories. McCarthy strongly objects to Lighthill's view of AI. As he writes,

> If we take this categorization seriously, then most AI researchers lose intellectual contact with Lighthill immediately, because his three categories have no place for what is or should be our main scientific activity—*studying the structure of information and the structure of problem solving processes independently of applications and independently of their realization in animals or humans.* [4, p. 64]

McCarthy discusses in many papers the "kind of science" AI should develop into. We have discussed above the connections and differences between AI and psychology and neurophysiology. However, McCarthy mostly concentrates on the connections and differences with logic and philosophy, the two disciplines which, in his opinion, have most in common with AI. This will be the topic of the following two subsections.

2.1. The role of philosophy

Philosophy and its connections with AI are discussed lengthily and technically in [3,6,7,18]. In [3], for instance, McCarthy describes some of the technical work developed in the philosophical literature (e.g. modal logics) and points out its relevance to AI. McCarthy's position can be synthetically described in three steps.

First, there is a common interest in epistemology. Thus, for instance, in [7] McCarthy writes:

> Philosophy and artificial intelligence. These fields overlap in the following way: In order to make a computer program behave intelligently, its designer must build into it a view of the world in general, apart from what they include about particular sciences [p. 114].

Second, despite the common interest, there are many differences. Thus, for instance, in [6], McCarthy writes:

> The word *epistemology* is used in this paper substantially as many philosophers use it, but the problems considered here have a different emphasis. Philosophers emphasize what is potentially knowable with maximal opportunities to observe and compute, whereas AI must take into account what is knowable with available observational and computational facilities. Even so, many of the same formalizations have both philosophical and AI interest. [p. 79]

(Notice how the word epistemology in the above quotation is given a meaning which is very reminiscent of the notion of epistemological adequacy.)

Third, the many existing differences have more far reaching consequences than one might expect. In particular, antithetic philosophical systems are sometimes equivalent in their ability to explain how the world manifests itself in the common sense informatic situation. This makes many important philosophical debates and alternatives irrelevant for AI. For instance in [3] McCarthy writes: "Since the philosophers have not really come to an agreement in 2500 years, it might seem that artificial intelligence is in a rather hopeless state if it is to depend on getting concrete enough information out of philosophy to write computer programs. Fortunately, merely undertaking to embody the philosophy in a computer program involves making enough philosophical presuppositions to exclude most philosophy as irrelevant" [p. 24]. In McCarthy's view, a philosophical distinction becomes important only if and when it suggests alternative ways to proceed in the construction of the advice taker. Consider for instance McCarthy's attitude towards the alternative between realism and empiricism, as discussed in [6,7]. In his view, it is better to build realist computer programs, rather than empiricist computer programs. A realist program would build theories of the world independently of its sense data, while the latter would build such theories only on the basis of an interpretation of such data. The latter program would have more partial theories of the world, only those emerging from its experience; and would not use a priori knowledge in the process of data interpretation. According to McCarthy, the second approach seems less likely to succeed than the first. As a second example, consider the motivations for studying the epistemological part of intelligence independently of the heuristic part (quotation from

p. 77). They all have a pragmatical, design oriented flavour. Essentially, the point is that making such a distinction is a good strategy for constructing the advice taker.

2.2. The role of logic

Properties 2 and 3 require that the advice taker be able to increment its knowledge meaningfully and independently of its current state. Property 2 requires that this be done by reasoning. Property 3 requires that this be done by interaction with the outside world. As described in [2], an advantageous way to achieve this is by using a declarative representation language (this paper also lists the advantages and disadvantages of declarative and imperative languages). Once the need for a declarative language has been accepted, logic becomes the obvious choice (see [2,3,5,18]). It provides a declarative language which can be used to represent knowledge, a formal notion of deduction which can be used to formalize and mechanize reasoning, and a set of formal tools which allow us to study the properties of the formalisms developed, e.g. their expressiveness and their consistency.

In this perspective, logic is a tool, more than the object of interest; and, as such, it should be pursued to the extent needed for the development of the advice taker. This emerges very clearly from all the papers of the collection which use or talk about logic. In particular in [14] McCarthy writes:

> AI badly needs mathematical and logical theory, but the theory required involves conceptual innovations—not just mathematics. We won't reach human level intelligence by more algorithms reducing the complexity of a problem from n^2 to $n \log n$, and still less by proofs that yet another problem is unsolvable or NP-complete. Of course, these results are often very significant as mathematics or computer science [pp. 187-188]

The choice of a logic-based approach to AI has been very controversial and discussed at length in the literature (see again [24,25]). The main problem is that logic has been developed with goals quite different from AI, e.g. to prove the consistency of mathematical reasoning, or to provide semantics to (parts of) natural language (even if some parts of logic and the logic-based approach to AI share the goal of the formalization of reasoning). Though logic is a very good starting point which allows formalizing many forms of common sense, it is far from having the expressibility needed to represent common sense. As the collection makes clear, a lot of McCarthy's technical work can be characterized as developing some of the extensions of logic needed in order to achieve epistemological adequacy. McCarthy argues at length in favour of his choice of a logic-based approach to AI in [14,18]. However the following quotation articulates his position:

> There is an intuition that not all human reasoning can be translated into deduction in some formal system of mathematical logic, and therefore mathematical logic should be rejected as a formalism for expressing what a robot should know about the world. The intuition in itself doesn't carry a convincing idea of what is lacking and how it might be supplied.

> We can confirm part of the intuition by describing a previously unformalized mode of reasoning called *circumscription*, which we can show does not correspond to deduction in a mathematical system. The conclusions it yields are just conjectures and sometimes even introduce inconsistency. We will argue that humans often use circumscription and robots must, too. The second part of the intuition—the rejection of mathematical logic—is not confirmed, the new mode of reasoning is best understood and used within a mathematical logical framework and coordinates well with mathematical logical deduction. We think that *circumscription* accounts for some of the successes and some of the errors of human reasoning. [6, p. 83]

2.3. *A critique*

I am mostly in agreement with McCarthy's perspective on logic. My view of logic, which, to some extent, can be seen as a summary of McCarthy's position, is that logic (with respect to AI) should be seen as an applied science whose topic of study is common sense and its mechanization inside machines; and that logic should play for AI the same role that mathematics has played in the development of physics. The more common sense we formalize, the more we will need to develop logic ... exactly as has happened for physics and mathematics.

I am also mostly in agreement with McCarthy's perspective on philosophy. However I believe that, pragmatically, McCarthy concentrates too much on the differences between AI and philosophy, and on what AI has to teach philosophy; and too little on what they have in common, and on what philosophy has to teach AI. The result is, in my opinion, that there has been little cross-fertilization between the two disciplines; and that this cross-fertilization has not gone to the core of problems. Thus, on one side, a lot of philosophical debates have been about the feasibility of the AI project *in principle*, essentially understood as the project of building an intelligent machine. Most of this work is, from my point of view, largely uninteresting and useless to the goal of building an advice taker. It misses many important aspects of the advice taker project, most noticeably, the crucial role of implementation and all the practical and theoretical issues it rises (for more about this point see Section 6.1). Vice versa, on the AI side, McCarthy has never tried to exploit the many results and ideas developed in "classical" philosophy, for instance in order to get a better understanding of the foundations and/ or implications of certain assumptions (e.g. the subjective view of common sense, the split of intelligence into a heuristic and an epistemological part, the requirement of epistemological adequacy, the rejection of modal logics).[5]

[5] I am aware that this last observation is of very high level and abstract. However an in depth discussion of this issue would require a lot of space and it would take us out of the main goals of this paper. An example of the kind of work I have in mind can be found in [27]. This thesis shows that the use of contexts as described in [28] leads to a subjectivistic notion of meaning and makes a first attempt at a comparison with the context-based notion of meaning elaborated in [29]. Another thesis being currently developed tries to interpret the logic-based approach to AI in light of Husserl's phenomenology.

3. Problems

The work discussed in the previous sections sets the stage for the research which needs to be done. The next step, discussed by McCarthy in [6,15], is to isolate the set of problems which needs to be solved. Concerning common sense knowledge, we can distinguish the problems identified by McCarthy by their subject matter:

(1) Common sense knowledge about the world's extension in time. McCarthy is mainly interested in formalizing and reasoning about the capability of our actions (and other events) to influence the world's temporal evolution, that is, in providing a theory of action (and events). Using a previously mentioned example from [15], a program which reasons about bacterial infections, like MYCIN, should be able to realize that if a patient has cholera, while the antibiotic is killing the cholera bacteria, the damage to his intestines is causing loss of fluids, which may cause death.

(2) Common sense knowledge about the world's extension in space. Objects are immersed in space; they often change their position and sometimes their shape. While this kind of knowledge is not needed for a program like MYCIN, it becomes relevant for a program which drives a navigating robot.

(3) Common sense knowledge about internal states. The world is populated by objects, e.g. people, programs and thermostats, which have an internal state which influences their behavior. In order to successfully interact (e.g. cooperate, compete) with such objects it seems quite useful to represent and reason about their state. Taking an example from [15], while reasoning about internal states is not necessary for a program like MYCIN, it may become necessary if we want to upgrade MYCIN to interact with users in a way which is dependent on their understanding of bacterial infections.

McCarthy also discusses various problems in the formalization of common sense reasoning. Most of them can be characterized as being instances of:

(4) People's ability to jump to conclusions,

(as also suggested by the title of Section 3 in [6]). The most important instance is the "qualification problem", discussed in [6,9], i.e. the problem of avoiding considering all the possible qualifications needed in order to make any sentence absolutely true. McCarthy discusses the qualification problem mainly in terms of reasoning about action. In this domain the qualification problem manifests itself in the practical (in my opinion, also in the theoretical) impossibility of reasoning about all the possible preconditions which must be true for an action to be applicable. Another instance, again in reasoning about action, is the "frame problem", i.e. the problem of expressing information and reasoning about what remains unchanged over time. Finally, ascribing mental qualities to a person or a machine, as in [7,13], really amounts to jumping to the conclusion that that person or machine has certain beliefs.

Of course this is not to say that all of common sense reasoning amounts to jumping to conclusions. In other papers, e.g. [2,8,10], McCarthy describes and formalizes other forms of common sense reasoning. However all these formalizations are done in first order logic and are relatively unproblematic, once one has found the appropriate way to formalize the common sense knowledge involved. The point is that the notion of deduction in logic provides us a formalization of the notion of reasoning. This cannot be done in the formalization of the reasoning involved in jumping to conclusions. In fact, one of the main properties of logic is monotonicity, i.e. if $\Gamma \vdash A$ then $\Gamma, B \vdash A$, where A, B are formulas and Γ is a set of formulas, and this is obviously not the case in the reasoning involved in jumping to conclusions, which is said to be "nonmonotonic". For instance I may want to ascribe the belief *Itishot* to a thermostat until I realize that it has actually started the heating. This does not mean either that all common sense nonmonotonic reasoning amounts to jumping to conclusions. In [16], McCarthy lists some further uses of nonmonotonic reasoning, namely as a communication convention, as a database or information storage convention, as a representation of a policy, as a streamlined expression of probabilistic information, as a rule of auto-epistemic reasoning, as a rule of common sense physics and common sense psychology.

4. Solutions

McCarthy does not further elaborate on common sense spatial knowledge, nor does he propose any formalization of it (except the formalizations of "on" and "above" in [9,16]). Most of his technical work is instead devoted to the formalization of the other three forms of common sense: the knowledge about time and action, the knowledge about internal states and people's ability to jump to conclusions. The solutions proposed are briefly discussed below.

4.1. Formalizing the "knowledge about time and action"

McCarthy's proposed formalism for reasoning about time and action is the situation calculus, introduced in [3]. Further discussions and examples can be found in [9,16,17,18]. The interested reader can also find many publications on the situation calculus by other authors in this journal. The name of this formalism derives from the fact that it includes, as part of its basic ontology, a new kind of object, called situation, where "A situation s is the complete state of the universe at an instant of time" [3, p. 35].

Having situations as objects of the calculus allows us to use them as arguments to function and predicate symbols. McCarthy calls *fluents* those applicational symbols whose domain is the space *Sit* of situations. A propositional fluent returns a truth value, a situational fluent returns a situation. In [3] McCarthy introduces some of the fluents necessary for the formalization of the common sense informatic situation:

- $at(p, x, s)$ (which intuitively means that a person p is in the place x in the situation s) and $in(x, y, s)$ (which intuitively means that x is in the location y in the situation s). These two fluents allow us to express propositions about the world embedding in space.
- $time(s)$ (which intuitively is the time of the situation s). This fluent allows us to express propositions about the world embedding in time.
- $result(p, \sigma, s)$ (which intuitively means the situation resulting from the person p performing action σ in situation s). This fluent allows us to express propositions

about the world's evolution through situations as a consequence of some action being performed.

McCarthy introduces the situation calculus without much motivation. Most noticeably, he does not motivate the decision for a calculus of situations, as opposed to a calculus of time. Notice that he explicitly acknowledges in [3] that the calculi defined prior to [3] are temporal calculi. In [3] he also points out that, technically, for any temporal calculus a situation calculus can be constructed with the same expressibility and proof-theoretic strength (citing Prior's work described in [30]). At a first sight, time might seem, and seemed to me, the right horizon over which to project the world evolution. This is because, contrary to what is (partially) the case with the world state and spatial coordinates, time is *the* coordinate over which we do not have any control. In this view, time objects (e.g. instants, intervals) are to be taken as ontological primitives while situations can be constructed, e.g., by applying a function symbol *situation* to time instants. However, after a closer look, a calculus of situations seems epistemologically more adequate, at least in some applications. For instance in the case of temporal projection, it is possible to consider alternative developments of events, not only of the actual course of history, and impose with a very simple axiom like $time(result(p, \sigma_1, s)) = time(result(p, \sigma_2, s))$ that two actions have the same duration. The general argument is that a lot of the common sense reasoning about the world is about the effects of actions on its state. Time is considered only when needed, e.g. when it is a necessary precondition for some action to be performed, like catching a train.

4.2. *Formalizing the "knowledge about internal states"*

In [7,13] McCarthy proposes that we reason about internal states in terms of a certain set of high level abstract primitives, called *mental qualities*, such as belief and self-belief, knowledge and self-knowledge, consciousness and self-consciousness, free will, intention and so on. The argument is that mental qualities are needed in order to have epistemologically adequate theories of machines, especially when such machines are complex and not completely understood. In particular in [13] McCarthy writes that "the language [used to describe internal states] must be able to express the information our program can actually get about a person's or machine's 'state of mind'—not just what might be obtainable if the neurophysiology of the human or the design of the machine were more accessible" [p. 181].

Roughly speaking, McCarthy's mental qualities are a superset of what is usually called propositional attitudes. Historically, the most common approach to the formalization of propositional attitudes is based on modal logics [31]. McCarthy does not believe that modal logics are an appropriate tool for the formalization of mental qualities. In [8, p. 138] he argues that many problems can be solved in first order logics; that many of the existing problem solvers are based on first order logics; that some of the existing work in modal theorem proving, i.e. that described in [32], is based on a translation of modal logics into first order logics; and that modal logics change both the syntax and the semantics of first order logics. This last consideration is quite convincing as it points out that, using modal logics, the treatment of mental qualities, which might be a small part of a problem, affects the formalization of the other parts.

McCarthy introduces and develops the formalism for reasoning about mental qualities in [6,8]. His solution is based on two main recipes. The first is the reification of concepts and propositions as objects of the language. This allows him to obtain the same expressiveness as propositional modal logics. The reason for this choice is that "Admitting individual concepts as objects—with concept-valued constants, variables, functions, and expressions—allows ordinary first order theories of necessity, knowledge, belief, and wanting without modal operators or quotation marks and without the restrictions on substituting equals for equals that either device makes necessary. ... [This formalization] doesn't modify the logic and is more powerful, because it includes mappings from objects to concepts" [8, pp. 119–120]. The second recipe is the reification of possible worlds as objects of the language. This allows him to obtain the same expressiveness as quantified modal logics. McCarthy seems much less convinced of this second choice. In [8, p. 135] he explicitly admits that he uses possible worlds reluctantly.

McCarthy does not provide a comprehensive technical analysis, with, for instance, equivalence results with modal logics. However, he provides formalizations of many interesting examples. In particular in [10] he describes the formalization and mechanization inside the interactive theorem prover FOL [33] of two puzzles involving knowledge. These formalizations are quite interesting as they show how it is possible to exploit the reification of worlds in a syntax which expresses both knowing what and knowing that [8], joint knowledge, nonknowledge, and the evolution over time of knowledge. Most interesting, this last feature is obtained simply by making the predicate formalizing the accessibility relation between worlds take an extra time/situation argument (the idea being that what is known is a function of what holds in all the accessible worlds, and that the accessible worlds change in dependence with the new facts which are acquired over time).

4.3. Formalizing "Jumping to conclusions"

McCarthy's proposed formalism for nonmonotonic reasoning is "circumscription", i.e. a rule of conjecture which allows a person or program to jump to the conclusion that the objects which can be shown to have a certain property P by reasoning on a given set of facts are all the objects that satisfy P. Various axiomatic forms of circumscription, all variations of this basic idea, and many examples of use are given in [6,9,16]. Let us consider for instance the form of circumscription defined in [9]. Let P be a predicate symbol and A a first order sentence. Let $A(\Phi)$ be the result of replacing all occurrences of P in A with Φ. Then the circumscription of P in A is the sentence schema

$$A(\Phi) \wedge \forall \bar{x}.(\Phi(\bar{x}) \rightarrow P(\bar{x})) \rightarrow \forall \bar{x}.(P(\bar{x}) \rightarrow \Phi(\bar{x}))$$

where $\bar{x}$ stands for the tuple $x_1, \ldots, x_n$. Intuitively, this formula says that if $A(\Phi)$ holds, and if Φ has a smaller extension of P, then P and Φ have the same extension. In other words, the set of objects which satisfy P is made as small as that satisfying Φ. Thus for instance, taking an example from [9], let P be the predicate *isblock* (stating of its argument that it is a block), and A the following formula

$$\textit{isblock}\, B \wedge \textit{isblock}\, C \wedge \textit{isblock}\, D. \tag{1}$$

Then, computing the circumscription of *isblock* in (1) and performing some small amount of first order reasoning leads to the formula

$$\forall x.(isblock\ x \rightarrow (x = B \vee x = C \vee x = D)),$$

which states that B ,C, D are the only objects which are blocks. Circumscription, like the situation calculus is one of the most studied and well-known formalisms in AI. The interested reader can find a lot of material about circumscription in [34,35].

4.4. A critique

McCarthy's proposed formalisms are a step towards the solution of some of the problems described in Section 3. However, it is often not obvious to see how these formalisms can be effectively used in the implementation of the advice taker, or, more simply, in practical applications.

A first difficulty comes from the way solutions are discussed and proposed. McCarthy's approach is articulated in three steps. In the first he identifies a problem, in the second he constructs an example which isolates the problem in a simple and crisp way and, finally, in the third, he provides a solution to the example. (In [11] all three steps are performed within a single paper.) This approach allows for theoretically clean solutions. McCarthy's solutions are general, that is not example specific, and provide insights about how to solve the problem identified. Furthermore, he is always careful in pointing out the extent to which his proposals solve or do not solve the problem (see for instance the examples in [11,12,16], but also the discussion at the end of [4]). However, the examples considered are very idealized and detached from how the problem manifests itself in practice. As a consequence, it is often not obvious to see how the proposed solutions can be scaled up to more concrete applications. This is the case for instance for nonmonotonic reasoning. Here, many different forms of circumscription have been defined (e.g. domain circumscription, predicate circumscription, formula circumscription, prioritized circumscription, pointwise circumscription). The nonexpert is faced with a set of techniques which look like ad hoc solutions invented to solve specific examples, rather than facilities of general utility. He is left with many choices whose applicability is not clear, and with very little help. It is hard to see how to map the problems arising in practical applications into the idealized, simple problems used to introduce and motivate the various forms of circumscription.

A second difficulty arises because the solutions proposed introduce a certain level of complexity, which seems sometimes due to the formalism, more than to the intrinsic complexity of the phenomenon to be modeled. This is the case for instance for McCarthy's formalism for representing mental qualities. Here the reification of individual concepts, together with all the related machinery for handling them, allows us to reach the necessary expressibility. However this richer ontology and machinery introduces extra complexity which must always be dealt with, even when not necessary. This becomes cumbersome when dealing with "simple" situations where, at least in his mind, one is forced to make distinctions which would not naturally arise.

5. Difficulties

McCarthy's research program and the proposed solutions present many difficulties. McCarthy is aware of this and writes about it in various papers of the collection. In particular he identifies three major difficulties. As discussed in [11], a first difficulty comes from splitting the heuristic and the epistemological part of intelligence. The starting point of the discussion is Kowalski's equation `ALGORITHM = LOGIC + CONTROL` [36], an equation which resembles the distinction between the epistemological part and the heuristic part of intelligence. As stated at the very beginning of [11],

> The formula isn't precise, and it won't be until someone proposes a precise and generally accepted notion of how control is to be added to an expression of the logic of a program. [p. 167]

This paper discusses this problem through the study of various algorithms for coloring planar maps. It starts by discussing an algorithm written in Prolog by Pereira and Porto and proceeds by describing a new solution which implements two ideas adapted from Kempe's 1879 incorrect proof of the four color theorem. It poses the question of whether an algorithm using these ideas can be regarded as a form of control added to the logic program, or whether these ideas must necessarily be implemented as a new program.

A second difficulty is in the use of logic and in the declarative representation of knowledge. Progress in this direction is slower than one might have expected. In [18] McCarthy writes:

> Existing computer programs come more or less close to this goal [of expressing information in logical sentences], depending on the extent to which they use the formalisms of logic. I shall begin by describing four levels of their use.
>
> (1) A machine on the lowest level uses no logical sentences. All its "beliefs" are implicit in its state. ...
>
> (2) The next level of use of logic involves computer programs that use sentences in machine memory to represent their beliefs but use other rules than ordinary logical inference to reach conclusions. ... Moreover, the sentences that appear in memory are from a program-dependent subset of the logical language being used. Adding certain true sentences in the language may even spoil the functioning of the program. Logic is used at this second level in "expert systems". ...
>
> (3) The third level uses first order logic as well as logical deduction. Usually the sentences are represented as clauses, and the deduction methods are based on J. Alan Robinson's (1965) [37] method of resolution. ...
>
> Examples of such programs used commercially are "expert system shells" (ART, KEE, OPS-5)—computer programs that create generic expert systems. ...
>
> Although they express both facts and rules as logical sentences, third level systems are still rather specialized. ... For this reason, the facts of one program usually cannot be used in a database for other programs.

(4) The fourth level is still a goal. It involves representing general facts about the world as logical sentences. Once put in a database, the facts can be used by any program. The facts would have the neutrality of purpose characteristic of much human reasoning. The supplier of information would not have to understand the goals of the potential user or how his mind works. The present ways of "teaching" computer programs amount to education by brain surgery. [pp. 238–241]

(Notice how the properties requested at the fourth level are again the three basic properties of common sense described in Section 1.1.)

Finally, a third major problem is articulated in [17] where McCarthy writes:

> It was obvious in 1971 and even in 1958 that AI programs suffered from a lack of generality. It is still obvious and now there are many more details. The first gross symptom is that a small addition to the idea of a program often involves a complete rewrite beginning with the data structures ...
>
> Another symptom is that no one knows how to make a general database of common sense knowledge that could be used by any program that needed the knowledge. [p. 226]

5.1. A critique

The lack of generality is, in my opinion, a consequence of the fact that *all* our theories of the world are necessarily approximate, i.e. they do not (and cannot, not even in principle) describe the world completely. The point is that the approximations we perform in the formalization of a problem are dependent on the problem being solved. Any new problem will impose a new view of the world which in turn may hide some details and highlight others previously not considered.

McCarthy realizes that many theories of the world are approximate (even if I have not found in any of the papers of the collection a sentence stating that this is always the case). In [3] he points out that any formalization of common sense temporal knowledge is necessarily approximate, in that it will fail to capture all that is true in the world. Thus, for instance, $result(p, \sigma, s)$ should be regarded as being defined in a certain representation of the world, as opposed to the world itself; and such representation will be necessarily approximate. In [6] he makes clear that the frame problem and the qualification problems are specific instances of reasoning within approximate theories. In [7] he points out that the ascription of mental qualities, though very useful when reasoning about machines whose structure is not completely known, will lead to the development of theories which are necessarily approximate. This leads him in [7] to stating the need for definitions with respect to an approximate theory (where the constraint is that a definition cannot be more precise than the theory permits), and of second order structural definitions (as opposed to external "behavioral" definitions, where second orderness is used to assert that certain beliefs are "good" for a machine M living in a world W).

However, I have found nowhere in McCarthy's papers a statement which links the problem of generality and the problem of having approximate theories. In my opinion, the identification of this connection allows us to understand and approach both problems in a unified and more general perspective. First of all, the qualification problem and the frame problem become instances of the problem of generality (this is discussed in [38]). Second, as we know a priori that our theories of the world are necessarily approximate, we can exploit this to achieve locality, i.e. to reason with the least possible amount of information. The idea is to throw away as much information as possible and reason in a theory which is very approximate but still general enough to allow for a meaningful solution to the problem to be solved. (The advantages of achieving locality, its relevance to common sense reasoning, and a formalization using contexts, are discussed in [28].) Third, this perspective highlights the fact that the problem of generality cannot be solved a priori, e.g. by having a completely general theory of the world. This suggests a solution where the (approximate) theories we use are formulated and reformulated depending on the problem to be solved (see [38] for a more detailed discussion).

6. Achievements

Starting from the 1959 formulation of the advice taker project, McCarthy's research in AI has undergone three major conceptual steps:

(1) Identification of the need for common sense and of an epistemology of common sense. The main contributions in this phase are the identification of common sense as the key aspect of the advice taker, and of the need for epistemological adequacy (this has been discussed in Section 1).

(2) Identification of the need for a science of common sense. The main contribution here is the proposal of a logic-based approach to AI (this has been discussed in Section 2).

(3) Development of some major contributions. These consist of the identification of some epistemological problems to be solved, the development of some technical solutions, and the recognition of some outstanding difficulties (this has been discussed in Sections 3, 4, 5).

As discussed in various papers e.g. [11,14,17,18] and also in Section 5 of this review, the advice taker project is still very far from accomplishment. However since 1959, major progress has been made and the stage for further research has been set. We have identified the target of our research, i.e. common sense, we have identified how to study it, i.e. by developing an epistemology of common sense, we have identified the formal tool for formalizing it, i.e. logic, and we have identified problems and proposed (partial) solutions which we have tested against case studies. I think that all of this can be summarized by saying that McCarthy has been successful in identifying, proposing and pushing the development of a new science, which can be called an *epistemological science of common sense*. It is a science because of what we discussed in Section 2; it is an epistemological science because, following what we discussed in Section 1, it aims at developing epistemic level theories of common sense. This is by far McCarthy's main achievement, the one which has had and will have lasting influence in the scientific and technological development of AI and computer science.

6.1. A critique

The papers in the collection suggest that McCarthy has not paid much attention to some important problems which arise in the process of mechanizing the theories developed on paper. The development on paper of a theory is only the first step. A straightforward mechanization very rarely works. This is because the ideas and formalizations developed on paper are implicitly based on assumptions which often make their implementation infeasible in practice. The problems which usually arise are that the resulting programs are too big, or hard to write or to maintain, or simply too slow—all problems which make theories fail the requirement of "practicality" necessary in order to achieve epistemological adequacy.

Of course, this is not to say that McCarthy has not worried about mechanization. His contributions in the development of programming languages (and theory of computation) suitable for writing programs with common sense are (again) impressive. It is sufficient to recall his work on LISP [39,40] and the (still unpublished) work on Elephant 2000 [41]. LISP's sophisticated symbol manipulation facilities are a first important step towards building a program which solves problems by manipulating sentences in a formal language. Elephant's high level interaction facilities are very useful in order to build programs which can increment what they know by being told. (Along the same lines as Elephant but perhaps less ambitious, is the work on CBCL described in [12].) It does not mean either that he has not worried about the implementation of the theories he has proposed. For instance, the examples in [10] and some of those in [16] have been mechanized. Moreover, as discussed in Section 4.2, some of the motivations for rejecting modal logics are implementational.

These are all important contributions towards the mechanization of common sense, but not all the story. Consider for instance the implementation of propositional attitudes inside complex reasoning programs, e.g. natural language understanding systems or multiagent systems. Here, the most efficient problem solvers are first order theorem provers based on the reification of possible worlds in the syntax (see for instance [42]). This shows that some of the intuitions and motivations given by McCarthy in favour of a first order treatment of mental qualities are correct (see Section 4.2).[6] However these results have not had the impact that one might have expected. Most existing applications do not use any of the proposed formalisms, nor the associated theorem provers. A major problem is that it is not clear how to codify into these formalisms all the necessary information (e.g. the agents' knowledge; their reasoning capabilities, which are usually different for each agent; the usually very complex interactions among them) in a way to have efficient, easy to develop and to maintain, implementations. There is a practical impossibility to use these formalisms, which makes them fail the requirement of epistemological adequacy, once the task boils down to developing a concrete application.

As far I can judge, nowhere in the papers of this collection, does McCarthy discuss these issues. Indeed, there is a sense in which some of his work worsens this problem;

[6] For the sake of completeness, it must be said that these formalisms are quite different from McCarthy's. However they are arguably simpler to handle. This strengthens the argument given below.

see, for instance, Section 4.4. However, in my opinion these issues are crucial for the future development of AI. Failing to consider them may result in a situation where theory and applications proceed in parallel with very little cross-fertilization. Let us further consider the mechanization of propositional attitudes. In practical applications, the most common approach is to represent the beliefs of an agent as a set of facts plus some way to deduce consequences from them (see for instance [43–45]). Until recently, this and the logic-based work (e.g. modal logics or McCarthy's formalism) were developed independently, the first lacking theoretical foundations, the second being hardly applicable in real systems. This problem is fixed in [46], which, via an equivalence result, shows that the latter representation schema, under certain conditions, has the same or more expressibility than modal logics. This result sets the stage for further substantial development in the mechanization of propositional attitudes inside complex programs (and, ultimately, inside the advice taker). It provides in fact a foundation to the applied work; and a clear methodology for replicating and extending it in new implementations. It also provides motivation for further theoretical work aiming at incorporating into the existing formalisms some of the features needed in actual implementations.

7. Should we read the collection?

There are various books about common sense and the logic-based approach to AI, see for instance [47,48]. These books are more exhaustive, contain a lot of background material and also more recent developments. However, this collection provides a different (and maybe unique) perspective on AI, a perspective which comes after reading, thinking about, and comparing the ideas expressed in papers written by one of the founders of the field over a period of more than thirty years. Reading the collection has made me acquire a clearer and more general perspective of what I am doing, and of why I am doing it. Moreover many papers, even the earlier ones, have been a very good source of inspiration and have suggested, or in other cases reinforced, some ideas of possible lines of future research. For this reason I believe that any researcher in AI, even those not interested in a logic-based approach, should spend some time reading, or rereading, the papers of the collection.

Acknowledgments

Paolo Bouquet, Vladimir Lifschitz, Carlo Penco, Manuela Viezzer and Toby Walsh have read earlier versions of this review, and provided very valuable feedback; this has been an unforeseen but very welcome opportunity for discussing the "roots" of what we are doing.

References

[1] V. Lifschitz, Understanding common sense: John McCarthy's research in artificial intelligence, in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 1–8.

[2] J. McCarthy, Programs with common sense, in: *Proceedings Teddington Conference on the Mechanization of Thought Processes*, London (1959) 75-91; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 9-20.
[3] J. McCarthy and P. Hayes, some philosophical problems from the standpoint of artificial intelligence, in: B. Meltzer and D. Michie, eds., *Machine Intelligence* **4** (Edinburgh University Press, Edinburgh, 1969) 463-502; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 21-63.
[4] J. McCarthy, Review of "Artificial intelligence: a general survey", *Artif. Intell.* **5** (1973) 317-322; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 64-69.
[5] J. McCarthy, An example for natural language understanding and the AI problems it raises, in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 70-76.
[6] J. McCarthy, Epistemological problems of artificial intelligence, in: *Proceedings IJCAI-77*, Cambridge, MA (1977) 1038-1044; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 77-92.
[7] J. McCarthy, Ascribing mental qualities to machines, in: M. Ringle, ed., *Philosophical Perspectives in Artificial Intelligence* (Humanities Press, 1979) 161-195; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 93-118.
[8] J. McCarthy, First order theories of individual concepts and propositions, in: J.E. Hayes, D. Michie and L.I. Mikulich, eds., *Machine Intelligence* **9** (Ellis Horwood, Chichester, 1979) 129-148; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 119-141.
[9] J. McCarthy, Circumscription—a form of non-monotonic reasoning, *Artif. Intell.* **13** (1980) 27-39; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 142-157.
[10] J. McCarthy, Formalization of two puzzles involving knowledge, in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 158-166.
[11] J. McCarthy, Coloring maps and the Kowalski doctrine, in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 167-174.
[12] J. McCarthy, The common business communication language, in: A. Endres and J. Reetz, eds., *Textverarbeitung and Buerosysteme* (Oldenbourg, Munchen, 1982) 71-74; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 175-178.
[13] J. McCarthy, The little thoughts of thinking machines, *Psych. Today* **17** (12) (1983); also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 179-186.
[14] J. McCarthy, AI needs more emphasis on basic research, *AI Mag.* **4** (1983); also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 187-188.
[15] J. McCarthy, Some expert systems need common sense, in: H. Pagels, ed., *Computer Culture: The Scientific, Intellectual and Social Impact of the Computer*, Annals of the New York Academy of Sciences **426** (1984) 129-137; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 189-197.
[16] J. McCarthy, Applications of circumscription to formalizing common-sense knowledge, *Artif. Intell.* **28** (1986) 71-87; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 198-225.
[17] J. McCarthy, Generality in artificial intelligence, *Commun. ACM* **30** (1987) 1030-1035; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 226-236.
[18] J. McCarthy, Mathematical logic in artificial intelligence, *Daedalus* **117** (1988) 297-311; also in: V. Lifschitz, ed., *Formalizing Common Sense: Papers by John McCarthy* (Ablex Publishing Corporation, Norwood, NJ, 1990) 237-250.

[19] E.H. Shortliffe, *Computer-Based Medical Consultation: MYCIN* (Elsevier, New York, 1976).
[20] E. Shortliffe, R. Davis and B. Buchanan, Production rules as a representation for a knowledge-based consultation program, *Artif. Intell.* **8** (1977) 15–45.
[21] R.C. Schank, Conceptual dependency: A theory of natural language understanding, *Cogn. Psych.* **3** (1972) 552–631.
[22] S. Haack, *Philosophy of Logics* (Cambridge University Press, Cambridge, 1978).
[23] J. McCarthy, Artificial intelligence and logic, in: R.H. Thomason, ed., *Philosophical Logic and Artificial Intelligence* (Kluwer Academic Publishers, Dordrecht, 1989) 161–190.
[24] M. Minsky, *The Society of Mind* (Simon & Schuster, New York, 1985).
[25] R.A. Brooks, Intelligence without reason, in: *Proceedings IJCAI-91*, Sydney (1991) 569–595.
[26] A. Newell, The knowledge level, *Artif. Intell.* **18** (1982) 87–127.
[27] A. Ciceri, La nozione di significato in sistemi formali per la rappresentazione della conoscenza contestuale, Tech. Rept. Thesis 9404-01, IRST, Trento (1994).
[28] F. Giunchiglia, Contextual reasoning, *Epistemologia, special issue on I Linguaggi e le Macchine* **XVI** (1993) 345–364; short version in: *Proceedings IJCAI-93*, Chambery (1993) 39–49; also: IRST Tech. Rept. 9211-20, IRST, Trento (1993).
[29] D. Kaplan, On the logic of demonstratives, *J. Philos. Logic* **8** (1978) 81–98.
[30] A.N. Prior, *Past, Present and Future* (Clarendon Press, Oxford, 1968).
[31] J. Hintikka, *Knowledge and Belief* (Cornell University Press, Ithaca, NY, 1962).
[32] C.G. Morgan, Methods of automated theorem proving in nonclassical logics, *IEEE Trans. Comput.* **25** (1976) 852–862.
[33] R.W. Weyhrauch, Prolegomena to a theory of mechanized formal reasoning, *Artif. Intell.* **13** (1980) 133–176.
[34] M.L. Ginsberg, ed., *Readings in Nonmonotonic Reasoning* (Morgan Kaufmann, Los Altos, CA, 1987).
[35] V. Lifschitz, Circumscription, in: D.M. Gabbay, C.J. Hogger and J.A. Robinson, eds., *The Handbook of Logic in AI and Logic Programming*, Vol. 3 (Oxford University Press, Oxford, 1993) 298–352.
[36] R. Kowalski, Algorithm = logic + control, *Commun. ACM* **22** (1979) 424–436.
[37] J.A. Robinson, A machine oriented logic based on the resolution principle, *J. ACM* **12** (1965) 23–41.
[38] P. Bouquet and F. Giunchiglia, Reasoning about theory adequacy. a new solution to the qualification problem, *Fund. Inform.* (to appear, 1995); also: Tech. Rept. 9406-13, IRST, Trento (1994).
[39] J. McCarthy, Recursive functions of symbolic expressions and their computation by machine, Part I, *Commun. ACM* **3** (1960); also in: S. Rosen, ed., *Programming Systems and Languages* (McGraw-Hill, New York, 1967).
[40] J. McCarthy, A basis for a mathematical theory of computation, in: P. Brafford and D. Hirshberg, eds., *Computer Programming and Formal Systems* (North-Holland, Amsterdam, 1963) 33–70.
[41] J. McCarthy, Elephant 2000: a programming language based on speech acts, Incomplete preliminary draft (1990).
[42] H.J. Ohlbach, A resolution calculus for modal logics, in: *CADE9*, Vol. 310 (Springer, Berlin, 1988) 500–516.
[43] A.R. Haas, A syntactic theory of belief and action, *Artif. Intell.* **28** (1986) 245–292.
[44] Y. Wilks and A. Ballim, Multiple agents and the heuristic ascription of belief, in: *Proceedings IJCAI-87*, Milan (1987) 118–124.
[45] Y. Wilks and J. Bien, Beliefs, points of view and multiple environments, *Cogn. Sci.* **6** (1983) 120–146.
[46] F. Giunchiglia and L. Serafini, Multilanguage hierarchical logics (or: how we can do without modal logics), *Artif. Intell.* **65** (1994) 29–70.
[47] E. Davis, *Representations of Commonsense Knowledge* (Morgan Kaufmann, Los Altos, CA, 1990).
[48] M.R. Genesereth and N.J. Nilsson, *Logical Foundations of Artificial Intelligence* (Morgan Kaufmann, Los Altos, CA, 1987).