Commonsense Reasoning*

Ernest Davis
Courant Institute
New York, New York

December 6, 2012

1 Introduction

For an artificial system to act sensibly in the real world, it must know about that world, and it must be able to use its knowledge effectively. The common knowledge about the world that is possessed by every schoolchild and the methods for making obvious inferences from this knowledge are called common sense in both humans and computers. Almost every type of intelligent task — natural language processing, planning, learning, high-level vision, expert-level reasoning — requires some degree of commonsense reasoning to carry out. The encoding of commonsense knowledge has been recognized as one of the central issue of AI since the inception of the field (McCarthy, 1959).

Endowing a program with common sense, however, is a very difficult task. Common sense involves many subtle modes of reasoning and a vast body of knowledge with complex interactions. Consider the following quotation from *The Tale of Benjamin Bunny,* by Beatrix Potter:

> Peter did not eat anything; he said he should like to go home. Presently he dropped half the onions.

Except that Peter is a rabbit, there is nothing subtle or strange here, and the passage is easily understood by five year old children. Yet these three clauses involve, implicitly or explicitly, concepts

* *Encyclopedia of Artificial Intelligence.* Stuart Shapiro (ed.), Wiley, 1990

of quantity, space, time, physics, goals, plans, and speech acts. An intelligent system cannot, therefore, understand this passage unless it possesses a theory of each of these domains and the ability to connect this theory in a useful way to the story.

Many of the central issues in the automation of commonsense reasoning appear in all types of AI reasoning, particularly the development of domain-independent knowledge structures and inference techniques, and the analysis and implementation of plausible reasoning. Since these issues are common throughout AI, we will not study them in this article; see, instead, the articles on Knowledge Representation, Reasoning, Default Reasoning, Nonmonotonic Reasoning, and Plausible Reasoning. Here, we will focus on issues that arise in the study of specific commonsense domains.

2 General Issues and Methodology

The analysis of reasoning in a commonsense domain has three major parts:

1. Representation: The development of knowledge structures that can express facts in the domain.
2. Domain theory: The characterization of the fundamental properties of the domain and the rules that govern it.
3. Inference techniques: The construction of algorithms or heuristics that can be used to automate useful types of reasoning.

A popular methodology for carrying out these kinds of analysis runs along the following lines: (See (McCarthy, 1968), (McCarthy and Hayes, 1969), (Hayes, 1977), (Hayes, 1978), (McDermott, 1978), and (Davis, 1990) for more detailed expositions of this methodology; and (Minsky, 1975), (McDermott, 1987) for criticisms of it.) The researcher begins by defining a *microworld*, a small, coherent domain of study. Aspects of the real world that lie outside the microworld will either be ignored in her work, or will be represented in some very coarse, *ad hoc* manner. Next, she assembles a coherent collection of commonsensically obvious inferences in the microworld. She determines what problem-specific information and general domain knowledge is needed, explicitly or implicitly, to justify these inferences. She develops a language in which these facts can be expressed and these inferences can be validated; typically, this language is written in some known logic (qv). Having categorized the types of information and rules that she needs, she can work on developing data structures and procedures that allow her to solve some significant class of problems efficiently.

A knowledge representation for a commonsense domain must satisfy three requirements. First, the representation must be able to describe the relevant aspects of the domain involved. Second, it must be possible to use the representation to express the kinds of partial knowledge typically available; the design of the language must take account of likely kinds of ignorance. Thirdly, it must be possible to implement useful inferences as computations using the representation. These requirements are called *ontological adequacy, expressivity* or *epistemic adequacy,* and *effectiveness* or *heuristic adequacy* (McCarthy and Hayes, 1969). Each of these requirements is defined relative to a certain set of problems; a representation may be adequate for one kind of problem but not for another.

The greatest difference between representations for commonsense reasoning and representations used in other areas of computer science lies in the expressivity requirement. Most computer science representations assume either complete information, or information that is partial only along some very limited dimensions. By contrast, commonsense reasoning requires dealing with a wide range of possible types of partial information, and degrading gracefully as the quality and quantity of information declines. Reasoning from partial information is important for three reasons: (1) It may be expensive, time-consuming, or impossible to get complete information. (2) Computing with exact information may be too complex. (3) Reasoning with partial information allows the inference of general rules that apply across a wide class of cases. For instance, suppose you are driving a car at the top of a cliff, and you wish to determine whether it would be better to take the winding road to the bottom or to drive down the cliff. Given exact specifications of the car and the exact topography of the cliff, it may be possible to predict exactly what would happen if you drove off the cliff. But (1) you may not have this information, nor any way to get it; (2) even if you had the information, the computation would be horrendous; (3) the conclusion would apply only to this car and this cliff. The calculation would have to be redone for each new car and each new cliff.

3 Time

Temporal reasoning (qv) is probably the most central issue in commonsense reasoning. Almost every application involves reasoning about time and change; very few microworlds of interest are purely static.

The first task of a temporal representation is to express changes over time; for example, to represent such facts as "At one time, the light was off; later, it was on." Such a representation is often based on the concepts of *situations* and *fluents*. A situation is an instantaneous snapshot of the world at an instant. A fluent is a description that changes its value from one situation to another, such as "the light being on" or "the president of the U.S." A fluent like "the light being on" that has possible values "True" and "False" is called a *Boolean fluent* or *state*.

We can define a first-order language (see Predicate logic) for describing situations and fluents using the following non-logical symbols:

- true_in(S, A) — Predicate. State A is true in situation S.
- value_in(S, F) — Function. The value of fluent F in situation S.
- precedes$(S1, S12)$ — Predicate. Situation $S1$ precedes $S2$.

For example, we can express the sentence "At one time the light was off; later it was on," in the formula

$$\exists_{S1,S2}\ \text{precedes}(S1, S2) \wedge \neg\text{true\_in}(S1,\text{on}(\text{light1})) \wedge \text{true\_in}(S2,\text{on}(\text{light1}))$$

Most events do not occur instantaneously; they occur over finite stretches of time. To incorporate events into our representations, we introduce the concept of a time *interval*, a set of successive situations. We add the following symbols to our language:

- $S \in I$ — Predicate. Situation S is in interval I.
- $[S1, S2]$ — Function. The closed interval from $S1$ to $S2$.
- occurs(I, E) — Predicate. Event E occurs during interval I.

Using this language, we can describe a variety of dynamic microworlds. For instance, we can express the blocks world rule, "If X and Z are clear, then the result of putting X onto Z will be that X is on Z," as follows:

$$\forall_{S1,S2,X,Y}\ [\text{true\_in}(S1,\text{clear}(X)) \wedge \text{true\_in}(S2,\text{clear}(Y)) \wedge \text{occurs}([S1, S2],\text{puton}(X, Y))] \Rightarrow \text{true\_in}(S2,\text{on}(X, Y))$$

In developing such a theory of a microworld, where the occurrence of an event changes the state of the world, one encounters the following problem: The theory must specify, not only the fluents that change as a result of an event but also the fluents that remain the same. For example, in the blocks world, it is necessary to infer that, when the robot puts X on Y, the only "on" states affected are those involving X. The problem of expressing or deriving such rules efficiently is known as the "frame" problem (McCarthy and Hayes, 1969); it has been the focus of much recent research. (See, for example, (Hanks and McDermott, 1987), (Shoham, 1988), (Brown, 1987), (Pylyshyn, 1987).)

The language described above follows (McDermott, 1982). Many other types of temporal languages have been devised, including languages that use only intervals but no individual situations (Allen 1983) (Allen, 1984); languages that distinguish sections of space-time (Hayes, 1978); and modal temporal languages (Prior, 1967) (van Benthem, 1983).

4 Space

Commonsense spatial reasoning (qv) serves three major functions:

- High-level vision. The interpretation of visual information in terms of world knowledge and the integration of information gained through vision into a general knowledge base.
- Cognitive map maintenance. The formation, maintenance, and use of a knowledge base describing the spatial layout of the environment. In particular, the use of a cognitive map for navigation, planning a route to a destination.
- Physical reasoning. Spatial characteristics of physical systems are generally critical in understanding its behavior. The behavior of many physical systems consists largely of spatial motions. Spatial reasoning is therefore a vital component of physical reasoning.

Finding a language for spatial knowledge that is both expressive and computationally tractable is very difficult. Ideally, a spatial language would allow the description of any physically meaningful spatial layout and spatial behavior, including specifications of shapes, positions, and motions; it would allow the expression of all types of information that are relevant to commonsense reasoning; it would allow a wide range of partial specifications, corresponding to the types of information that may be obtained from perception, natural language text, or physical inference; and it would do all this in a way that supports efficient algorithms for commonsense reasoning. No such language has yet been found.

The following are some of the more extensively studied spatial representations: (For surveys, see (Requicha, 1980), (Ballard and Brown, 1982), (Hoffmann, 1989).)

- Occupancy array. The space is divided up into a rectangular grid, and each cell of the grid is associated with one element of an array. Each element of the array holds the name of the object(s) that intersect the corresponding rectangle in space. One disadvantage of this representation is that it is costly in terms of memory; this can be mitigated by the use of quad-trees or oct-trees, which merge adjacent array elements with identical labels.

- Constructive solid geometry. A shape is characterized as the union and difference of a small class of primitive shapes.
- Boundary representations. A shape is characterized in terms of its boundary. For example, the representation might approximate a two-dimensional shape as a polygon, which is defined by listing its edges, its vertices, and the coordinates of the vertices.
- Topological representations. A spatial layout is characterized by describing topological relations between objects. For instance, the TOUR program (Kuipers, 1978) describes a road map by stating the order in which places appear on a path, and the cyclic order in which paths meet at a place. Randell and Cohn (1989) characterize the spatial relations between objects in terms of such relations as abutment and overlapping.

5 Physical reasoning

Unlike the sciences, which aim at a simple description of the underlying structure of physical reality, the commonsense theory of physics must try to describe and systematize physical phenomena as they appear and as they may most effectively be thought about for everyday purposes. The problems addressed in physical reasoning include predicting the future history of a physical system, explaining the behavior of a physical system, planning physical actions to carry out a task, and designing tools to serve a given purpose.

Commonsense physical reasoning characteristically avoids the use of exact numerical values. Rather, it relies on qualitative characterization of the physical parameters involved, such as "If a kettle of water is placed on a flame, it will heat up; the higher the flame, the faster the water will heat." This rule does not specify the exact rate at which the temperature changes; it specifies that the change is positive, and that the rate is an increasing function of the height of the flame.

Accordingly, the mathematical structure of such constraints has been extensively studied. (See "Qualitative Physics," (de Kleer and Brown, 1985), (Kuipers, 1986).)

Complex physical systems, particularly artificial devices, can often be effectively analyzed by viewing them as a collections of *connected components*. In simple cases, the connections between the components remain constant over time; what varies are the values of various one-dimensional *parameters* of the system. Components are connected at *ports;* each parameter is associated with one port. The laws that govern these systems are *component characteristics,* which constrain the values of the parameters at the ports of the component, and *connection characteristics*, which constrain the values of parameters at ports that meet in a connection. For example, in electronics, the component characteristics are rules such as "The difference between the voltages at the two ends of a resistor is equal to the current through it times its resistance." The connection characteristics are the rules, "At a connection, the voltages of all the ports is equal, and sum of all the current flows into the ports is zero." The particular device is specified by describing the components it contains, and the connections between their ports. (de Kleer and Brown, 1985)

An alternative way to decompose physical systems (Forbus, 1985) focuses on the *processes* that occur. Consider, for example, a closed can of water above a flame. A process-based description of the behavior of this system would say that there is first a heating process, in which the temperature of the water rises to its boiling point; then a boiling process, in which the water turns from a liquid to a gas; then another heating process, in which the temperature and pressure of the gas rise steadily, and finally a bursting event, when the pressure of the gas exceeds the strength of the can. The central elements of such a representation are process types, such as heating and boiling, and parameters, such as temperature and pressure. The laws that govern the system describe how a process influences a parameter, such as "A boiling process tends to reduce the quantity of liquid and increase the quantity of gas;" they describe influences of one parameter on another, such as "The pressure of a gas tends to rise with its temperature;" and they describe the circumstances under which a process can take place, such a “Boiling occurs just if there is a heat-flow into a body of water that is at its boiling point.”

Reasoning about systems of solid objects involves techniques of a different kind. The central problem here is the geometric reasoning required, which can be very complicated, particularly with three-dimensional objects of complex shapes. However, the physical laws describing such systems are fairly straightforward. Many mechanisms, particularly man-made devices where the parts are tightly constrained can be used using only the physical laws that solid objects are rigid and cannot overlap. Such an analysis is known as a *kinematic* analysis (Faltings, 1987) (Joskowicz, 1987). For loosely constrained systems of solid objects, such as a bouncing ball, it is generally necessary to use *dynamic* analysis, invoking the concepts of Newtonian mechanics.

Liquids are still more complicated to represent and reason about, because they are not divided into discrete objects; they continually combine and separate. Hayes (1985) discusses a logical analysis of a commonsense theory of liquids.

One final issue in physical reasoning is causality. Ordinary discourse about physical events is often framed in terms of one event causing another; however, none of the theories mentioned above make any use of causality as a concept. Extracting a causal account from such theories has proven to be difficult, particularly as there is no consensus on exactly what purpose a causal account should serve. For discussions, see (de Kleer and Brown, 1985), (Iwasaki and Simon, 1986), (Shoham, 1988), (Pearl, 1988).

6 Knowledge and Belief

To reason about other agents, or even about oneself at other times, it is necessary to have a theory describing their mental life. AI studies of commonsense theories of cognitions have primarily focussed on agents’ knowledge and beliefs, discussed in this section, and their plans and goals, discussed in the next.

Representations of knowledge and belief must necessarily be quite different in structure from the representations we have considered above for temporal, spatial, and physical information. The relation, "A knows ϕ" takes as its argument a *proposition*, which may contain Boolean operators, quantifiers, or imbedded statements about knowledge. (Consider, for example, "John knows that Mary knows that all Libras were born either in September or in October.") Operators in first-order languages, by contrast, can take as arguments only terms that denotes entities. Moreover, first-order operators are *referentially transparent*; if two terms τ and ω denote the same entity then ω can be substituted for τ in any sentence without changing the truth of the sentence. By contrast, psychological relations, such as "A knows ϕ", are *referentially opaque*; substitution of equal terms may change the truth of a sentence. For example, given that Sacramento is the capital of California, it follows that "John is in Sacramento" is true just if "John is in the capital of California" is true. However, it is possible for "John knows that he is in Sacramento" to be true but "John knows that he is in the capital of California" to be false, if John believes that Los Angeles is the capital of California.

Three general types of representations have been developed for these kinds of relations:

- "Know" and "believe" can be represented as operators in a *modal logic* (qv), a logic that extends first-order logic by allowing additional operators on sentences. In such a theory, the sentence mentioned above could be represented

  know(john, know(mary, $\forall_X$ libra(X) $\Rightarrow$ born_in(X, sept) $\vee$ born_in(X, oct)))

- "Know" and "believe" can be represented as first-order predicates that take as argument a string of characters that spell out the sentence known or believed. The above sentence would be represented

  know(john, $\prec$know(mary, $\prec\forall_X$ $\bar{\text{l}}$ibra(X) $\Rightarrow$ born_in(X, sept) $\vee$ born_in(X, oct)$\succ$)$\succ$)

  where $\prec$ and $\succ$ are string delimiters. (In this example, the difference between the two theories appears trivial; in fact, there are deep differences in their logical characteristics.)

- Facts about knowledge and belief can be expressed in terms of accessibility relations among *possible worlds*. A possible world is one conceivable way that the world could be; a fact may be true in one possible world and false in another. A world W1 is accessible from world W0 relative to the knowledge of agent A if nothing in W1 contradicts something that A knows in W0. We can then express the fact that A knows ϕ in world W by saying that ϕ is true in every world accessible from W. Thus the statement, "John knows in world w0 that he is in Sacramento" can be represented

  $$\forall_{W1} \text{ know\_acc(john, w0, } W1) \Rightarrow \text{true\_in}(W1,\text{in(john, sacramento))}$$

  where the predicate "know_acc$(A, W0, W1)$" means that world $W1$ is accessible from world $W0$ relative to the knowledge of A, and "true_in" means the same as in the section on temporal reasoning. Our other sample sentence can be represented

  $$\begin{aligned} &\forall_{W1,W2} \text{ [know\_acc(john, w0, } W1) \wedge \text{know\_acc(mary, } W1, W2)] \Rightarrow \\ &\quad \forall_X \text{ true\_in}(W1,\text{libra}(X)) \Rightarrow \\ &\qquad [\text{ true\_in}(W2,\text{born\_in}(X,\text{sept})) \vee \text{true\_in}(W2,\text{born\_in}(X,\text{oct}))] \end{aligned}$$

For extensive discussions of these representations and their relative merits, see (Moore, 1980), (Halpern and Moses, 1985), and (Morgenstern, 1988).

The next problem is to characterize what agents know and believe in a way that supports reasonable commonsense inferences. Most theories to date have been modeled *implicit* knowledge and belief. An agent implicitly knows ϕ if he, in principle, has enough information to determine ϕ; that is, if ϕ is a logical consequence of facts that he knows. However, in many situations, such as teaching, implicit knowledge is not a reasonable theory; a teacher who assumes that the students can immediately perceive all the consequences of what he says will be disappointed. It has been difficult to find more psychologically plausible theories of knowledge and belief, that accommodate the fact that reasoners are limited in the speed and power of their inferential abilities. (See (Konolige, 1986), (Levesque, 1984).)

Other issues that have been studied in AI theories of knowledge include the gaining of knowledge through perception (Davis, 1988) and the *auto-epistemic* inference, which allows an agent to infer that ϕ is false from the fact that he does not know ϕ. (Moore, 1985)

7 Plans and Goals

The second major focus of AI commonsense psychological theories has been in representing and reasoning about plans and goals. Plans and goals have been studied primarily in connection with two high-level tasks: *plan construction,* the problem of finding actions that an agent can perform to accomplish his goal; and *motivation analysis,* the problem of explaining an agent's actions in terms of his plans and goals. The chief problems in analyzing plans and goals are the following:

- Constructing a language to describe plans and goals. Defining what it means to carry out a plan or to accomplish a goal described in the language.
- Characterizing the feasibility of a plan; the validity of a plan for achieving a given goal; and the cost of a plan.
- Characterizing the typical high-level goals of human actors.
- Giving criteria for evaluating alternative explanations of actions in terms of plans and goals.
- The problem of searching for the best plan in plan construction or the best explanation in motivation analysis.

Most the planning literature in AI has assumed a particularly simple model of plans and goals. A goal is taken to be a desired state of the world, such as "Block C is on block B" or "John is home." A plan is taken to be a sequence of primitive actions, actions that can be directly carried out by a low-level robotic controller. For example, in the blocks world, the action, "Put X on Y" could be taken to be primitive. Plans would then be sequences of "put on" instructions such as "First put A on the table; then put C on B." Furthermore, it is assumed that the planner is omniscient; he knows everything that can possibly be relevant. Under these assumptions, the definitions of feasibility and correctness are straightforward: A plan is feasible if the preconditions of each successive action hold at the time that the action is scheduled to be performed; a plan accomplishes a goal if the goal holds after all the actions have been performed. The main problem is then one of search; finding a correct plan, given a starting situation and a goal.

More sophisticated theories generalize this basic notion of plans and goals in a number of different ways:

1. A plan may only partially specify the actions to be taken, leaving details to be completed at execution time. Consider, for example, a plan to mail a letter consisting of five steps: (1) Insert the letter in an envelope; (2) Address the envelope; (3) Attach a stamp to the envelope; (4) Seal the envelope; (5) Put the envelope in a mail box. When forming the plan, it is probably not necessary to identify exactly which envelope, stamp, and mail box should be used; when the plan is executed, the most convenient objects of these types may be chosen. Moreover, the steps need not be totally ordered at planning time. As long as step 4 follows step 1, and step 5 is the last operation, other ordering relations among the steps may be chosen at execution time (Sacerdoti, 1975) (Chapman, 1987).

2. If it will be necessary for an agent to achieve repeatedly goals of a similar form, it may be worthwhile constructing a *generic* plan, that will accomplish these goals in all circumstances, rather than planning each case individually. For example, in the blocks world, one might want to construct the generic plan "Clear block X; clear block Y; put X on Y" for achieving the goal "X on Y." (Sussman, 1975) (Manna and Waldinger, 1987).

3. If we consider planners that are not omniscient, but have only partial knowledge of the environment, then the analysis of plans become more complicated in several respects. First, in this context, plans and goals become referentially opaque operators, like knowledge and belief. John may plan or wish to go to the capital of California, and yet not plan or wish to go to Sacramento,

if he does not know that they are the same place. It is therefore necessary to use one of the techniques described in the previous section — modal logic, syntactic operators, or possible worlds — to represent the plans and goals of agents who are not omniscient.

Planners with partial knowledge must also deal with circumstances in which the planner must gain information in order to achieve his goal. For example, if John wants to call Mary, but does not know her phone number, he may construct the plan, "First look up Mary's number in the phone book; then dial that number." The analysis of this kind of plan is known as the *knowledge preconditions* problem (Moore, 1980) (Morgenstern, 1988).

4. For either generic plans (2) or planning with partial knowledge (3), it may be useful to augment the planning language so that a plan can specify actions that depend on the state of the world. For this purpose, it may be useful to introduce operators similar to those of programming languages, such as conditionals, loops, variable binding, interrupts, and so on.

In order to carry out motivation analysis — the explanation of an agent's actions in terms of his goals and plans — it is necessary to have a theory that describes characteristic goals. (Otherwise, it would be possible to explain any action as done for the fun of it.) Schank and Abelson (1977) suggest five general categories of top-level goals:

- Satisfaction goals — Basic physical needs, such as hunger, thirst, and fatigue, that arise periodically.
- Preservation goals — The desire to preserve certain key personal states, such as preservation of life, health, and possessions.
- Achievement goals — Large-scale ambitions accomplished over a long terms, such as raising a family or success in a career.
- Entertainment goals — The short-term enjoyment of some activity, such as seeing a movie.
- Delta goals — The acquisition of certain goods, particularly wealth and knowledge.

8 Other issues

Other commonsense domains that have been studied in the AI literature, including emotions (Dyer, 1983), (Sanders, 1989) interactions among agents (Wilensky, 1983), (Bond and Gasser, 1988), communication (Perrault and Allen, 1980), and thematic relations between people (Schank and Abelson, 1977).

9 Bibliography

J. Allen, "Maintaining knowledge about temporal intervals," *Comm. ACM,* **26,** 832-843 (1983).

J. Allen, "Towards a general theory of action and time," *Artif. Intell.* **23,** 123-154 (1984).

D. Ballard and C. Brown, *Computer Vision,* Prentice Hall, Englewood Cliffs, NJ, 1982

A. Bond and L. Gasser, *Readings in Distributed Artificial Intelligence,* Morgan Kaufmann, San Mateo, CA, 1988.

F. Brown (ed.) *The Frame Problem in Artificial Intelligence: Proceedings of the 1987 Workshop,* Morgan Kaufmann, San Mateo, CA, 1987.

D. Chapman, "Planning for conjunctive goals," *Artificial Intelligence,* **32,** 333-378, (1987).

J. de Kleer and J.S. Brown, A Qualitative Physics Based on Confluences, in D. Bobrow (ed.), *Qualitative Reasoning about Physical Systems,* MIT Press, Cambridge, MA., pp. 7-84, 1985.

E. Davis, Inferring Ignorance from the Locality of Visual Perception, *Proceedings of the Sixth National Conference on Artificial Intelligence,*, Minneapolis, pp. 786-790, 1988

M. Dyer, *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension,* MIT Press, Cambridge, MA, 1983

B. Faltings, Qualitative Kinematics in Mechanisms, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence,* Milan, pp. 1331-1336, 1987.

K. Forbus, Qualitative Process Theory, in D. Bobrow (ed.), *Qualitative Reasoning about Physical Systems,* MIT Press, Cambridge, MA., pp. 85-168, 1985

J. Halpern and Y. Moses, A Guide to the Modal Logics of Knowledge and Belief, *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* Los Angeles, pp. 480-490, 1985.

S. Hanks and D. McDermott, "Nonmonotonic logic and temporal projection," *Artif. Intell.* **33,** 379-412 (1987).

P. Hayes, In Defense of Logic, Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, MA, pp. 559-565, 1977.

P. Hayes, The Naive Physics Manifesto, in D. Michie (ed.), *Expert Systems in the Micro-electronic Age,* Edinburgh University Press, Edinburgh, Scotland, 1978.

P. Hayes, Naive Physics 1: Ontology for Liquids, in J. Hobbs and R. Moore (eds.), *Formal Theories of the Commonsense World,* Ablex, Norwood, New Jersey, pp. 71-108, 1985.

C. Hoffmann, *Geometric and Solid Modeling: An Introduction,* Morgan Kaufmann, San Mateo, CA, 1989.

Y. Iwasaki and H. Simon, "Causality in device behavior," *Artif. Intell.* **29** pp. 3-32 (1986).

L. Joskowicz, Shape and function in mechanical devices, *Proceedings of the Sixth National Conference on Artificial Intelligence,*, Seattle, WA, pp. 611-618, 1987.

K. Konolige, "Belief and Incompleteness," in J. Hobbs and R. Moore (eds.), *Formal Theories of the Commonsense World,* Ablex, Norwood, New Jersey, pp. 71-108, 1985.

B. Kuipers, "Modeling spatial knowledge," *Cog. Sci.* **2**(2) 129-154 (1978).

B. Kuipers, "Qualitative simulation," *Artif. Intell.* **29**, pp. 289-338 (1986).

H. Levesque, A Logic of Explicit and Implicit Belief, *Proceedings of the Fourth National Conference on Artificial Intelligence,* Austin, TX. 1984.

Z. Manna and R. Waldinger, A Theory of Plans, in M. Georgeff and A. Lansky (eds.), *Reasoning about Actions and Plans,* Morgan Kaufmann, San Mateo, CA, 1987.

J. McCarthy, Programs with Common Sense, in *Proceedings of the Symposium on Mechanisation of Thought Processes 1*, London, 1959.

J. McCarthy., Programs with Common Sense, in M. Minsky (ed.) *Semantic Information Processing,* MIT Press, Cambridge, MA, pp. 403-418, 1968.

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, pp. 463-502, 1969.

D. McDermott, "Tarskian semantics, or no notation without denotation!" *Cog. Sci.*, **2**,3 277-282 (1978).

D. McDermott, "A temporal logic for reasoning about processes and plans," *Cog. Sci.*, **6,** 101-155 (1982).

D. McDermott, "A Critique of Pure Reason," *Computational Intelligence*, **3** 151-160 (1987).

M. Minsky, A Framework for Representing Knowledge, in P. Winston (ed.) *The Psychology of Computer Vision,* McGraw-Hill, 1975.

R. Moore, A Formal Theory of Knowledge and Action, in J. Hobbs and R. Moore (eds.), *Formal Theories of the Commonsense World,* Ablex, Norwood, NJ, pp. 319-358, 1985.

R. Moore, "Semantical Considerations on Nonmonotonic Logic," *Artif. Intell.,* **25** 75-94, 1985.

L. Morgenstern, *Foundations of a Logic of Knowledge, Action, and Communication,* NYU Ph.D. Thesis, 1988.

J. Pearl, *Probabilistic Reasoning in Intelligent Systems; Networks of Plausible Inference,* Morgan Kaufmann, San Mateo, CA, 1988.

C. Perrault and J. Allen, "A plan-based analysis of indirect speech acts," *American Journal of Computational Linguistics,* **6**, 167-182 (1980).

A.N. Prior, *Past, Present, and Future,* Clarendon Press, Oxford, 1967.

Z. Pylyshyn, *The Frame Problem and Other Problems of Holism in Artificial Intelligence,* Ablex, Norwood, NJ, 1987.

D.A. Randell and A.G. Cohn, Modeling Topological and Metrical Properties in Physical Process, in *Proceedings of the First International Conference on Principles of Knowledge Representations and Reasoning,* Toronto, 1989.

A.A.G. Requicha, "Representations for rigid solids: Theory, methods, and systems," *ACM Comput. Surv.,* **12**(4) 437-464 (1980).

E. Sacerdoti, *A Structure for Plans and Behaviors,* American Elsevier, New York, 1975.

K. Sanders, A Logic for Emotion, *Proc. Conf. for Cognitive Science,* Ann Arbor, MI, pp. 357-363, 1989.

R. Schank and R. Abelson, *Scripts, Plans, Goals, and Understanding,* Erlbaum, Hillsdale, NJ, 1977

Y. Shoham, *Reasoning About Change,* MIT Press, Cambridge, MA, 1988.

G. Sussman, *A Computer Model of Skill Acquisition,* American Elsevier, New York, 1975.

J. van Bethem, *The Logic of Time,* Reidel, Dordrecht, 1983.

R. Wilensky, *Planning and Understanding,* Addison-Wesley, 1983.

9.1 General References

R. Brachman and H. Levesque (eds.), *Readings in Knowledge Representation,* Morgan Kaufman, Menlo Park, CA, 1985. Reprints of many classic articles. Extensive bibliography.

E. Charniak and D. McDermott, *Introduction to Artificial Intelligence,* Addison-Wesley, Reading, MA, 1985. Chapters 1, 6, and 7 give an excellent introduction to knowledge representation and commonsense reasoning.

E. Davis, *Representations of Commonsense Knowlege,* Morgan Kaufmann, San Mateo, 1990. Comprehensive textbook.

M. Genesereth and N. Nilsson, *Logical Foundations of Artificial Intelligence,* Morgan Kaufmann, San Mateo, 1988. Extensive discussion of logic, nonmonotonic inference, and knowledge representation.

J. Hobbs and R. Moore, *Formal Theories of the Commonsense World,* ABLEX Publishing, Hilldale, NJ, 1985. Collection of research papers.

D. Weld and J. de Kleer, *Qualitative Reasoning about Physical Systems,* Morgan Kaufmann, San Mateo, CA, 1989. Extensive collection of research papers on physical reasoning.