How To Frame It

Modern Applied Logic from the Top Down,
or, Secrets of the Metamathematicians

Jon Doyle

Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

October 22, 1986

(PRELIMINARY DRAFT—FOR USE IN 15.780 ONLY)

Abstract: These notes survey modern logic's uses in formalizing subject matters, and its applications in artificial intelligence in particular. We assume familiarity with the elementary concepts of logic, and focus on logic's vocabulary for assessing, comparing, and iteratively improving tentative formulations of knowledge. We also consider logics of specific subjects, and present an annotated bibliography.

Contents

1 Introduction

This paper is about one theoretical and two practical applications of mathematical logic in artificial intelligence: the analysis of representational systems, the elucidation or formulation of knowledge, and the calculation of meaning. Our focus is on the second of these applications, the elucidation of knowledge, and we assume acquaintance with the elementary concepts of logic—logical languages, proofs, and models. (The appendix lists the elementary notions whose aquaintance is assumed here.) Almost any textbook of logic provides an adequate introduction to these concepts. See, for example, [Suppes 1957], [Margaris 1967] or the survey article [Prior 1967].

1.1 Elucidation of knowledge

The elucidation of knowledge is a central activity in many sorts of thinking. The basic process of elucidation is one of iteratively judging, diagnosing, and revising tentative formulations and formalizations of knowledge, and may be applied to all aspects of an agent's knowledge, including its beliefs, preferences, procedures, and plans. For example, the task of knowledge acquisition involves elucidating the knowledge possessed by an informant; the task of learning involves elucidating the organizing aspects or theoretical knowledge underlying a body of data; and the formulation of knowledge of how to best achieve one's goals is the essence of deliberation and planning. Logic supplies a vocabulary of ways of analyzing, judging, and criticizing tentative formulations of subject matters—specific expertise, data or experience, plans and possibilities—so as to guide the process of improving the formulations, and the bulk of this paper is devoted to explaining these concepts and their use in the process of formalization.

1.2 Analysis of representational systems

Logic's metatheoretical vocabulary is both fundamental and universal, and so supplies a means for theoretically assessing all proposed representational systems, whether or not these systems are based on logical languages and logical operations. Briefly put, logic may be used as a language of ascribed representations in analyzing the behavior of agents. Whether an agent is constructed using representations or using no explicit representations at all, by treating the agent's behavior as a subject matter we may ascribe representations to it if we wish—that is, present the theory of its behavior as if it operated representationally. These imputed representations may then be analyzed as if they were real to determine the properties and limits of the agent's knowledge. The logical properties and limits of the imputed representations are then true descriptions of the agent's powers, no matter how the agent is organized internally. By the same token, using logic to specify the desired behavior of an agent in no way commits one to using logical languages or logical operations in the implementation of the agent.

1.3 Calculation of meaning

Logic is a tool of mechanization as well as a tool of formalization. Much of artificial intelligence concerns computation or calculation, and logic also supplies ways of calculating meanings, implications, and conclusions. But these uses of logic are already well covered in the literature, and so we will only discuss them briefly. (See [Nilsson 1980], [Kowalski 1977], [Siekmann and Wrightson 1983].) The fundamental methods are to use proofs to determine entailments, and to use models to justify (or contradict) implications. Both methods have been widely used in artificial intelligence, though their usefulness is usually limited, both because of incompleteness and intractability of theories, and because of infinity and ungraspability of models. See [Rabin 1974] and [Harel 1987] for reflections on the computational limits of these techniques.

1.4 Logic and psychology

Logic has also been presented as a theory of thinking, with the rules of logic cast as laws of thought. But this characterization of logic is misguided. Logic is not a theory of reasoning or thinking, whether descriptively or normatively construed. It is not a descriptive theory for humans, for humans are not ideal. More importantly, it is also not a normative theory, for reasoning is an activity, and logic is not about an activity but about expression, meaning, and entailment. Unfortunately, the belief that logic is intimately related to reasoning is widespread, and this causes no end of confusion. Do not be fooled when you come across such claims. We cannot treat this subject here: see [Harman 1986], [McDermott 1986], and [Doyle 1986] for extended discussions.

2 The process of formalization

The process of formalizing a subject matter is easily described but of unpredictable difficulty. In this section, we sketch a recipe or procedure for formalizing knowledge that anyone can follow.

In essence, the process of formalization is iterative improvement and development of an initial formulation—a sequence of rational guesses and rational revisions of these guesses. In these notes we only sketch the process informally. For an extended treatment of these ideas in much greater detail, see [Polya 1945], [Polya 1962], and especially [Polya 1965]. Do not assume that Polya's focus on mathematical problem solving means that this method applies only to mathematics. On this point, see also [Truesdell 1984a]. For a formal treatment, see [Doyle 1986].

It is important to realize that, to paraphrase Hamming [1962], the purpose of formalization is insight, not axioms. Formalization aims at making intuitions exact, precisely so that they may be clearly understood. The first aim is not calculation, but criticism. Can one's intuitions be made coherent? When clearly stated, are they what one wants? Will everyone agree on them? Are one's initial intuitions adequate, or are there better (simpler, more powerful) ways of thinking? In answering such questions, computation serves only as a means for gaining experience with the implications of formalizations, not as an end in itself. While one may write and employ poorly understood programs, if we are ever to trust these programs enough to use them in consequential applications (e.g., medicine, business, engineering, manufacturing, administration, education, war), we must be able to understand them, and achieving such understanding is what formalization is all about. See [Doyle 1984] and [Doyle 1985] for further discussion.

2.1 Formulation and formalization

We may divide the process of codifying a subject matter into two parts, formulation and formalization. (Actually, this is not quite right, but we will correct the error as we go.) Formulation comes first, and is informal—a sketch of the codification—and formalization comes second—a detailed portrait over the basis of the sketch. The step of formulation looks much like philosophy (verbal proposals, arguments, and criticisms), while the step of formalization looks like mathematics (mathematical definitions and axioms, proofs, and counterexamples).

At the beginning of the process, if the subject is not well understood already, one must flounder around for a while, trying out various conceptualizations without worrying about the precise details, just to get a feel for what seems worth pursuing. This floundering constitutes the formulation step, and it may go on for a long while. One can only truly proceed from formulation to formalization if one can make the concepts of the favored informal formulation mathematically precise. If the mathematics of the day does not offer enough adequate formal concepts to do this, the formalization step must be postponed pending the search for new mathematical concepts to complete the task. In any event, it is always possible to regress from mathematical investigation to philosophical discussion if the precise formalization of a formulation shows up previously invisible flaws. If the mathematical formalization itself does not suggest fixes, it may be necessary to reopen philosophical debate in light of the new information to find a new formulation.

For some questions, it may be that we never get beyond the stage of philosophical discussion, but that is an open question. Philosophers like to view the central questions of philosophy as eternal, but even if that is so, the process of finding better and better formulations of these questions increases our knowledge and spins off many mathematical subjects of considerable interest. If you are interested in learning more about this sort of thing, read [Rorty 1979], [MacLane 1986], [Jaffe 1984], and [Truesdell 1958]. For fun, consider [Brams 1983].

2.2 Iterative formulation and formalization

The basic procedure is described by these steps.

1. Pick a vocabulary of concepts (a language of discussion), and axioms, rules of reasoning, and intended interpretations of the terms of this vocabulary.

2. See if all the obvious or intuitive elements of the subject matter are *expressible* in the language. If not, return to step 1 and fix the vocabulary.

3. See if the axioms and rules *correctly* capture *all* the obvious or intuitive elements of the subject matter. Do they lead to false or undesired con-

clusions? Are some elements missing? If so, return to step 1 and fix the axioms and rules.

4. See if the axioms and rules admit unwanted alternative interpretations (for instance, trivial ones), or if they in fact are incompatible with the intended interpretation, with some of the intended interpretations if there are more than one, or with each other. If they are *ambiguous*, *incompatible*, or *inconsistent*, return to step 1 and fix the axioms and rules, or widen or fix the set of intended interpretations.

5. Now that everything seems to be working right, we check to see if the formalization may be improved in various ways. First, see if concepts, axioms, and rules may be *simplified*. If so, back to step 1.

6. Next, see if the subject may be *separated* into disjoint sub-subject matters. If there are two concepts that appear roughly orthogonal, there may be a minor reformulation that fully separates them and so clarifies them. If they have some irreducible overlap, perhaps this intersection is an interesting concept in itself, and can be isolated in a way that makes separating the rest easy. If so, back to step 1.

7. See if the formulation can be generalized to cover more ground. If so, back to step 1.

8. See if the formulation can be efficiently mechanized, and if not, if some more mechanizable reformulation is possible. (These questions may refer to either a fixed means of mechanization, or to any possible mechanization.) If so, back to step 1.

Note that this procedure has no stopping point. You can always seek further improvements. How long you go depends on whether the subject remains interesting and important, and whether your current formalization is good enough for your current purposes.

Another point is that the same procedure works for either verbal formulations and mathematical formalizations. The two differ only in the standards for convincing definitions, arguments, and counterarguments. In philosophy, it is easy to seem to have defined or proved something, but people can argue with you forever about whether you have or not. In mathematics, it may be hard to find the formal definitions and proofs, but once you have, there is no further argument, at least about the proofs.

Similarly, the models employed may be formal or informal. Mathematical logic texts often convince students that models are always purely mathematical, abstract structures. In fact, any things may be used in interpreting theories. Especially in science, the intended interpretations are things in our world, not abstract structures.

2.3 Formalization and mechanization

When the procedure is used to develop a formalization, the question of mechanization need not be held until the end.

By mechanization we mean a procedure for assessing truth of sentences relative to the theory, or for assessing provability of sentences via the theory. The two notions of truth and proof yield two basic ways of mechanizing the process of answering questions within a formalization. One may use proofs to answer questions about models (using implication to answer questions about entailment), or use models to answer questions about proofs (discovering implications via entailments). If a theory is sound, then every implication reflects an entailment, so if one proves one thing from another, it reveals an entailment. If the theory is complete, every entailment reflects an implication, so if all models satisfying one thing also satisfy another, it reveals an implication (but not necessarily the proof of the implication in the theory).

In general, theories are unmechanizable, or, if mechanizable, are hopelessly difficult. But a mechanization or partial mechanization may be used in the early steps of the procedure to facilitate experiments with the formalization. Steps 3 and 4 consist of checking that the tentative formalization really captures the intended subject and not something else. To the extent that a mechanization relieves some of the chore of checking this, it speeds up the process of formalization. Of course, if one has to specially construct a mechanization for each subject and at each iteration of the procedure, the savings of effort may be minimal, so for these initial checks one may want to use a general purpose automated deduction system. Unfortunately, most automated deduction systems build in a single set of inference rules. The closest one comes to the theory-testing system that formalization requires are the Edinburgh ML system and the "production systems" employed in artificial intelligence—general interpreters of systems of rules and axioms. In fact, these are heavily employed in artificial intelligence primarily as means for experimenting with tentative formulations in expert system development (see [Doyle 1985]).

A word of caution: do not let questions of mechanizability (which often are very difficult to address) unduly influence the task of formulation too early. Yes, mechanizations can be convenient, but the most important thing is to first find the right theory, capturing the intended subject, and only then see how to modify or approximate it for mechanization. Without the discipline of a single subject, the procedure of formalization will wander aimlessly.

2.4 Formalization and intuition

In the initial stages of formalization, one ordinarily attempts to reflect the intuitive informal vocabulary and methods of the subject matter in the formal vocabulary and theory. In some cases, the intuitive conception may be formalized without substantial change. Beeson [1985, pp. 82-86], following Feferman, discusses the notions of adequacy and fidelity of formalizations for describing cases in which formalization mirrors intuition. A theory T is an *adequate* formalization of a body M of informal subject matter if every concept, argument, and result of M can be represented by a (basic or defined) concept, proof, and theorem, respectively, of T. (We might wish to place limits on the complexity and levels of definitions allowed to preserve a close connection between subject and theory.) Next, T is *faithful* to M if every basic concept of T corresponds to a basic concept of M, and every axiom and rule of T corresponds to or is implicit in the assumptions and reasoning followed in M. That is, T is faithful to M if T does not go beyond M conceptually or in principle. Beeson considers suitable formalizations to be ones both adequate and faithful to the subject. (See his exposition for more discussion.)

But in some cases, one's intutions involve several mutually inconsistent oversimplifications, and cannot be directly transformed into a formalization. Even when one's intuitions are consistent, further analysis of the formalization may reveal even better formulations whose concepts are quite removed from the superficial structures guiding the initial formulation, so that the task of formalization involves creating and comparing alternative views. In these cases, the best theory is not faithful to the intuitive theory. Relativity and quantum theory are examples of this in physics; the theory of logrolling is an example from politics and economics. Thus fidelity may not always be a desirable property of a theory. Similarly, in science and analytic thought generally, one attempts to separate distinct notions and study them separately. There need not be "natural" encodings of subjects; at best there will be natural encodings of particular views of the subject corresponding to individual intuitions. In these cases, adequacy need not be desirable either.

2.5 Metamathematics of formalization

Modern logic supplies mathematical tools not only for formalizing the axioms and rules of a subject (that is what is usually thought of as logic), but also the mathematical tools for judging or criticizing proposed formulations and formalizations according to expressibility, accuracy, ambiguity, and incompatibility, as well as mechanizability. In the following sections, we introduce the basic vocabulary of metamathematics for precise study of these questions. We will not try to be exact in every detail, or to spell out all uses and applications. The aim rather is to convince the reader that these questions of formulation may be posed precisely. To learn more about these metatheories, see [Barwise 1977] and [Barwise and Feferman 1985].

Section 3 presents the concepts concerning properties of individual theories or formalizations. These are of primary importance in steps 2-4 of the procedure, the steps of criticism of the tentative theory on internal grounds. Section 4 presents the concepts involved in comparing theories. These are of primary importance in steps 5-8 of the procedure. In section 5 we present the viewpoint of modern logic on finding the specialized logics of individual subject matters, and introduce several specialized logics that have been studied.

3 Assessing theories

Before proceding, let us clarify the terminology we use. Informally, the common meaning of "theory" is a description of a subject matter—for example, a list of axioms, rules, and intended interpretations. This is the meaning employed in this paper, and in Weyhrauch's [1980] formal system as well. In logic, however, the term theory has another meaning more convenient to logical studies. In logic, a theory is the set of all sentences true of a subject. (See [Keisler 1977] for precise definitions.) An axiomatization is any subset of a theory equivalent to the whole theory. Naturally, a theory may have many different axiomatizations. Indeed, one may revise the procedure of formalization to explicitly employ theories by, in step 1, writing down all things one can think of that are true of the subject, whether or not those facts seem fundamental or not. Then, in steps 5 and 6, more perspicuous axiomatizations may be sought. In this way, all the facts of the subject are initial data, and one seeks to classify the facts as consequences of better and better axiomatizations.

3.1 Consistency

The most important property to seek in constructing a theory is consistency. A theory is *consistent* just in case it has an interpretation in some model, so to demonstrate consistency, one need only exhibit a model. Any model, simple or not, intended or not, will do. An inconsistent theory has no interpretation in any model, and so certainly cannot be interpreted as a description of the subject matter. More exactly, an inconsistent theory describes the chosen subject matter just as well as it describes every other subject matter. But even though an inconsistent theory may be useless in practice, it is still valuable in the process of formalization, for it is usually easier to correct an inconsistent theory than to discover the theory in the first place. First drafts are almost always represent progress over blank notebooks.

For example, in some cases one constructs a theory listing all of one's intuitions about the subject matter, only to discover them inconsistent. This is not necessarily a tragedy, for it offers an opportunity to clarify one's thinking and to better understand the subject matter. Sometimes the theory may be fixed by modifying one or more of the axioms. In other cases, each of the axioms is really something desirable, and these desires conflict necessarily. (Anyone skeptical that one's intuitions may conflict should read any text on morality or functions of a real variable, or consider Milnor's theorem on decision rules [Luce and Raiffa 1957, pp. 297-298].) In these cases, one often must split the theory into two or more similar but incompatible theories, each one consistent by itself, but containing some axiom inconsistent with each of the sibling theories. In each particular application, one must then decide pragmatically which of these versions is most appropriate. Occasionally further experience and reflection will lead to improved insight and a better, single theory, but usually one must either pick one as a standard or live with a set of theories possessing a family resemblance. Be prepared to find that family resemblances occur very frequently in describing the ordinary objects of human life. These families of objects or properties are called *natural kinds*. Their commonness makes ordinary logical definitions a rarity in formalizing commonsense knowledge, and artificial intelligence has developed some specialized logics for capturing these concepts. See section 5.4.

In principle, proving a theory consistent is easy. One simply constructs an interpretation of the theory in a model; for example, an interpretation as one of the items covered by one's subject matter. This is usually a simple matter for theories whose intended interpretations include specific finite structures. If your theory is supposed to be one of finite structures and you are having trouble constructing a model, consider the possibility that the theory is inconsistent. In general, however, it is often difficult to tell if one's theory is consistent. Consistency of first-order theories is undecidable, and if the intended interpretation is infinite (for example, involving the natural numbers), one must judge how much effort, if any, to expend attempting to prove its consistency. If one cannot determine consistency or inconsistency in one's allotted time, one might simply assume the theory consistent and be on the lookout for indications of inconsistency that may crop up in future uses of the theory.

3.2 Categoricity

A theory is *categorical* if all of its models are isomorphic, so that it "really" describes a single structure. (For infinite subjects this notion must be restated, but it will do for this discussion.) If the intended subject matter is a single thing, it is important to check to see that the theory describes only that thing. If the theory is not categorical, it will possess other interpretations that are not what one intends. On the other hand, if one intends that the theory describe each member of a class of distinct structures, then one does not want a categorical theory. Categoricity is not like consistency, a uniformly desirable aim. Instead, it is a property that may or may not be desirable.

There are several metamathematical results (for example, the Löwenheim-Skolem theorems) in logic that say that it may be harder than one thinks to capture only the intended subject matter. One result is that finiteness and countability are inexpressible in first-order logic. Thus any set of first-order axioms that permit arbitrarily large finite models also have infinite models.

3.3 Completeness

A theory is *complete* if every statement in its language about the subject matter is either entailed by the theory or is inconsistent with the theory. That is, a complete theory entails all things that are true of the subject. Inconsistent theories are trivially complete, so if a complete theory is consistent, the set of its possible interpretations is exactly the subject matter (or things isomorphic to elements of the subject matter).

Clearly, a categorical theory is complete, but a complete theory need not be categorical. Like categoricity, completeness may or may not be desirable or possible. When it is achieved, however, it means that the theory says everything there is to say about the subject (at least in terms of the language of the theory), while one may always add new facts to an incomplete theory. Some subjects, for example number theory, possess no finite complete axiomatizations. One may always find some truths of arithmetic not entailed by a given finite axiomatization.

3.4 Decidability

A theory is *decidable* if the set of sentences it entails is recursive, that is, if there is an effective procedure that determines whether or not a sentence is a consequence of the axioms. Every finite complete set of axioms is decidable, but in general, decidable theories are rare, and the decision procedures for most decidable theories are either apparently or provably intractable, so ordinarily the issues of complexity discussed below are more important than decidability.

3.5 Complexity

While the preceding properties of theories are all absolute notions, either possessed or not by an individual theory, questions about complexity concern tradeoffs among equivalent axiomatizations.

The first questions concern the size of the axiomatization. Is there a finite axiomatization of the subject? Or is it not finitely axiomatizable? Are the axioms independent, or are some redundant? In general, it may be difficult to answer these questions. In simple cases, one demonstrates finite axiomatizability by exhibiting a finite axiomatization, and one demonstrates independence by exhibiting models of each subset of axioms falsifying the remaining axioms.

In artificial intelligence, we are especially interested in theories possessing finite (or perhaps recursive) sets of axioms, and if the subject matter does not admit such an axiomatization, we must live with that and choose some small partial axiomatization to work with. One can think of partial axiomatizations as approximations to the intended theory (though there may be other ways of approximating a theory too).

In foundational studies, logicians ordinarily are also concerned that the axioms be independent, as this reduces the number of cases that must be considered in proving properties of the theory. In applications, where the concern is more on using the theory than on proving properties of the theory, independence need not be as important. Indeed, when developing a formalization of a subject, one should not worry about independence at all, unless one is too slowed down by excessive lists of redundant axioms. The axiom set may always be simplified later.

The next questions about complexity concern the difficulty of proving consequences of the theory using the axioms and rules. Of course, if the theory is not decidable, some consequences are infinitely difficult to determine. But for those consequences that are provable, one may ask how costly are the proofs. As is known from the theory of computational complexity, this depends on both the size of the proofs and how frequently proofs occur among possible derivations. Both of these factors are influenced by the size and form of the axioms and rules, so that questions like independence of axioms, though not pressing in formalizing a subject, may be crucial in efficiently applying it.

4 Comparing theories

4.1 Relative strength

The first way to compare two theories is by looking at their sets of models. If two theories T_1 and T_2 have exactly the same set of models, they are said to be *equivalent* (or equipollent, or equal strength). If every model of T_1 is also a model of T_2, but not vice versa, we say that T_2 is *weaker* than T_1 (or T_1 is *stronger* than T_2). Weaker theories have more models; stronger theories fewer. If T_1 is not weaker, stronger, or equivalent to T_2, the two theories are *incomparable.*

The easiest way of weakening a theory is to remove axioms. The smaller set of axioms will have at least as many models as the larger. (If it has the same models, the removed axioms were redundant.) Conversely, a theory may be strengthened by adding nonredundant axioms. (If the new axioms make the theory inconsistent, it will have no models, and cannot be made any stronger.)

4.2 Expressiveness

We may also compare theories by comparing the expressiveness of their languages. If L_1 and L_2 are two languages, and there is a set of definitions D defining every term of L_1 as an expression of L_2, such that for every theory T_1 in L_1 and T_2 in L_2, T_1 is equivalent to $T_2 \cup D$, then L_1 is *reducible* to L_2. This means that the language L_1 is no more expressive than L_2.

4.3 Reducibility

Another way of looking at relative strength is as reducibility of one theory to another. Suppose that T_1 and T_2 are expressed in languages L_1 and L_2 respectively. Then if there is a set of definitions D defining every term of T_1 in L_2 such that T_1 is equivalent to $T_2 \cup D$, we say that T_1 is *reducible* to T_2. (Individual theories may be reducible without supposing their languages are.) Logical reducibility turns out to be related to relative strength: T_1 is stronger than T_2 if and only if T_1 is reducible to T_2, and T_2 is not reducible to T_1, and T_1 and T_2 are reducible to each other if and only if they are equivalent.

4.4 Relative assessments

Reducibility provides a way of comparing many properties of theories. If T_1 is reducible to T_2, and if T_1 is consistent if T_2 is, then we have demonstrated *relative consistency*, consistency of T_1 relative to T_2. Similarly, this reduction indicates the *relative completeness*, *relative categoricity*, or *relative decidability* of T_1 if one can prove T_1 complete, categorical, or decidable (respectively) under the assumption that T_2 is complete, categorical, or decidable (respectively).

Refined notions of reducibility also play a role in computational complexity. If T_1 is reducible to T_2, then the problem of answering questions in T_1 can be reduced to the problem of answering questions in T_2, and we may judge the *relative complexity* of T_1 by considering the complexity of T_2 and the costs of translation. For example, if queries in T_1 may be translated into queries in T_2 in time polynomial in the size of the original query, we say that T_1 is *polynomial-time reducible* to T_2.

4.5 Approximation

These ways of comparing theories provide ways—perhaps very weak ways—of saying that one theory approximates another. In cases in which the subject is not finitely axiomatizable, or is not feasibly decidable, we may have to employ more tractable theories that yield useful approximations. The least interesting notion of approximation is that of what can be answered in 10 steps or 10 seconds of reasoning in the ideal theory. Such approximation by amputation normally yields chaotic results, but is widely used in artificial intelligence as the principal means of approximate reasoning. Usually better means of approximation include using more tractable weaker or stronger theories (if any exist), using simplified theories that omit some of the complicating aspects of the subject, using probabilistic algorithms that compute correct answers with known chances of failures or error, and using more tractable theories whose answers are close to the exact ones (as in approximation algorithms for the traveling salesman problem). See [Harel 1987] for more on these ideas.

5 Specialized logics

One of the central ideas of modern logic is to pick a class of structures—the intended models—and to study the logic of these structures, that is, study the satisfaction relation $M \models \phi$ for models M ranging over this class. Such logics are called *model-theoretic logics*, logics embodying specific concepts. (See [Barwise 1985].) First-order logic, which builds in almost nothing, is interesting mathematically, but relatively limited in the concepts it can express. For many interesting subjects, one must depart from first-order logic. (This issue, like the relation of logic to psychology, is also controversial, as at least one prominent philosopher, W. V. O. Quine, has argued for decades that first-order logic is all anyone should need. There is growing evidence he is wrong on this point.)

From this point of view, a theory is just the set of all sentences true of some class of structures, and an axiomatization of the theory is some equivalent set of sentences. The question of completeness becomes the question of whether the valid sentences of the logic are recursively enumerable. Concepts whose logic is incomplete are not necessarily bad, merely very complex.

5.1 Logics of mathematical concepts

For example, some standard logics are best thought of as logics of mathematical concepts. Second-order logic, in which one can quantify over predicates and functions as well as individuals, is logic with set theory built in. Monadic and weak second-order logic are restricted variants of this. Logicians have also studied logics of probability theory that build in measure theory and have quantifiers like "has at least probability r" for each $r > 0$. The collection [Barwise and Feferman 1985] surveys these sorts of logics and their construction, from a logician's point of view.

5.2 Logics of logical concepts

Logics have been developed for several sorts of logical or computational reasoning. For instance, relevance logic [Anderson and Belnap 1975] builds in the concept of strict implication (in which the conclusion must have something to do with the antecedent). Very roughly speaking, intutionistic and constructive logics [Beeson 1985] build in notions of certainty, accumulation, and constructibility of conclusions, as do modal logics of provability [Boolos 1979]. And other modal logics of necessity and possibility [Chellas 1980], conditionals and hypotheticals [Lewis 1973] build in metaphysics, theories of what sorts of worlds are possible and how these different possible worlds are related. (See also [Turner 1984].)

5.3 Logics of computation and time

Some logics build in theories of non-mathematical, non-logical subjects. For example, temporal logics [Rescher and Urquhart 1971] build in theories of time, and have been discussed much in both artificial intelligence and computer science. In the latter they form one element of logics of programs [Harel 1979]—themselves theories of another subject matter, the classes of computational trajectories expressible in specific programming languages. (See also [Turner 1984].)

5.4 Logics of psychological concepts

But for artificial intelligence, the more interesting sorts of specialized logics are those building in either psychological notions or specific subject matters. While mathematicians have mainly studied logics of mathematical concepts, philosophical logicians have also studied logics of psychological (or pseudo-psychological) concepts. For example, there are numerous logics of belief, knowledge, desire, intent, obligation, and a large literature on each. In a logic of belief, for instance, the language typically includes a modal operator **Bel**, so that $\mathbf{Bel}(p)$ means that p is believed. The logic also includes rules and axioms for drawing conclusions about the agent's beliefs from such statements, such as the rule

$$\mathbf{Bel}(p \wedge q) \vdash \mathbf{Bel}(p) \wedge \mathbf{Bel}(q).$$

There have been a large variety of logics of belief ("doxastic" logics) proposed, mostly because no one agrees on what beliefs should behave like. Each doxastic logic builds in a different conception of beliefs. The same confusing situation holds true for logics of knowledge, desire, etc. as well. (See [Rescher 1968].)

Artificial intelligence has also developed logics for special sorts of psychological structures. For example, logics of defaults [Reiter 1980] and non-monotonic logics ([McDermott and Doyle 1980], [Moore 1983]) are logics of simple rules for making assumptions and of the conclusions these rules yield. (See also [Doyle 1983], [Doyle 1986].) Such assumptions also appear in some inheritance systems as well (see [Touretzky 1984]), which are logics of concepts like "part of," "subclass of," and "instance of." One might think of these inheritance and default logics as logics of prototypes or natural kinds. Natural kinds include things like lemons, lions, life, and money—things that everyone is familiar with, but which do not admit precise definitions because they admit many sorts of exceptions and variants—we just know them when we see them, and may argue about borderline cases. There are big literatures on natural kinds in both philosophy and psychology.

Acknowledgments

I thank Joseph Schatz for many discussions about this subject and related matters. This research was supported by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 4976, monitored by the Air Force

Avionics Laboratory under Contract F33615-84-K-1520. The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Government of the United States of America.

A Elementary logic

We will not give an exposition of elementary logic here. Instead, we simply give lists of fundamental notions, and leave to the reader the exercise of checking his familiarity with each. Definitions and discussions of these notions may be found in most textbooks, for example [Margaris 1967]. See also [Barwise 1977] and [Siekmann and Wrightson 1983].

A.1 Languages

symbols, constants, functions, predicates, propositions, variables, literals, connectives, quantifiers, operators, terms, wffs, sentences, schemata, bounded quantification and sorts, type theory, first and second order logic, infinitary logic

A.2 Models

interpretations, satisfaction, validity, tautology, contradiction, contingency, entailment, compactness, Löwenheim-Skolem theorems, finite axiomatizability, recursive axiomatizability, compactness, truth functionality, modality, Herbrand universes

A.3 Proofs

axioms, rules, natural deduction, soundness, completeness, normalization, constructibility, deduction theorem, provability, equality, Gödel incompleteness theorem, reflection principles

A.4 Mechanization

disjunctive normal form, conjunctive normal form, Horn clauses, matrix, Skolem functions, unification, subsumption, resolution, restricted resolution (locking, Horn), completeness, PROLOG, ML, ERGO

Bibliography

General logic texts and reference works

[Andrews 1986], [Barwise 1977], [Barwise 1985], [Keisler 1977], [Margaris 1967], [Prior 1967], [Scott 1982], [Suppes 1957], [Tarski 1946]

Philosophy of logic

[Barwise 1985], [Berger 1977], [Cherniak 1984], [Haack 1978], [Quine 1970]

Logic and psychology

[Cherniak 1984], [Doyle 1986], [Gazzaniga 1985], [Geertz 1983], [Harman 1973], [Harman 1986], [McDermott 1986], [Quine 1953], [Quine 1978]

Formulation and formalization

[Blanché 1962], [Doyle 1985], [Doyle 1986], [Hamming 1962], [Jaffe 1984], [Lakatos 1976], [Mac Lane 1986], [Montague and Montgomery 1984], [Polya 1945], [Polya 1962], [Polya 1965], [Rorty 1979], [Suppes 1957], [Tarski 1946], [Tarski 1959], [Truesdell 1984a], [Truesdell 1984b], [Wilder 1965]

Logic and artificial intelligence

[Charniak and McDermott 1985], [Doyle 1984], [Doyle 1985], [Fikes and Nilsson 1971], [Hayes 1970], [Kowalski 1979], [McCarthy 1958], [McCarthy 1977], [McCarthy 1980], [McCarthy 1986], [McCarthy and Hayes 1969], [McDermott 1978], [McDermott 1986], [Nilsson 1980], [Siekmann and Wrightson 1983], [Weyhrauch 1980]

Computability and effectiveness

[Harel 1987], [Rabin 1974], [Rabin 1976]

Logics of mathematical concepts

[Barwise 1985], [Barwise and Feferman 1985]

Logics of logical concepts

[Anderson and Belnap 1975], [Beeson 1985], [Boolos 1979], [Chellas 1980], [Lewis 1973], [Rescher 1968], [Turner 1984]

Logics of computation and time

[Harel 1979], [Rescher 1968], [Rescher and Urquhart 1971], [Scott 1982], [Turner 1984]

Logics of psychological concepts

[Barwise and Perry 1985], [Doyle 1983], [Doyle 1986], [Jeffrey 1983], [McDermott and Doyle 1980], [Moore 1983], [Rescher 1968], [Touretzky 1984], [Turner 1984]

Anderson, A. R., and Belnap, N. D., Jr., 1975. *Entailment*, Volume 1, Princeton: Princeton University Press.

Andrews, P. B., 1986. *An Introduction to Mathematical Logic and Type Theory: To truth through proof*, New York: Academic Press.

Barwise, J., 1977. *Handbook of Mathematical Logic*, Amsterdam: North-Holland.

Barwise, J., 1985. Model-theoretic logics: background and aims, *Model-Theoretic Logics* (J. Barwise and S. Feferman, eds.), New York: Springer-Verlag, 3-23.

Barwise, J., and Feferman, S., 1985. *Model-Theoretic Logics*, New York: Springer-Verlag.

Barwise, J., and Perry, J., 1983. *Situations and Attitudes*, Cambridge: MIT Press.

Beeson, M. J., 1985. *Foundations of Constructive Mathematics: Metamathematical Studies*, Berlin: Springer-Verlag.

Berger, F. R., 1977. *Studying Deductive Logic*, Englewood Cliffs: Prentice-Hall.

Blanché, R., 1962. *Axiomatics* (G. B. Keene, tr.), New York: Dover.

Boolos, G., 1979. *The Unprovability of Consistency: An essay in modal logic*, Cambridge: Cambridge University Press.

Brams, S. J., 1983. *Superior Beings*, New York: Springer-Verlag.

Charniak, E., and McDermott, D., 1985. *Introduction to Artificial Intelligence*, Reading: Addison-Wesley.

Chellas, B. F., 1980. *Modal Logic: An Introduction*, Cambridge: Cambridge University Press.

Cherniak, C., 1984. Computational complexity and the universal acceptance of logic, *J. Philosophy* Vol. LXXXI, No. 12, 739-758.

Doyle, J., 1983. Some theories of reasoned assumptions: an essay in rational psychology, Pittsburgh: Carnegie-Mellon University, Department of Computer Science, report 83-125.

Doyle, J., 1984. Expert systems without computers, or theory and trust in artificial intelligence, *AI Magazine*, V. 5, No. 2, 59-63. Errata, V. 5, No. 4, 78.

Doyle, J., 1985. Expert systems and the "myth" of symbolic reasoning, *IEEE Transactions on Software Engineering*, Vol. SE-11, No. 11, 1386-1390.

Doyle, J., 1986. Topics in rational self-government, in preparation.

Fikes, R. E., and Nilsson, N. J., 1971. STRIPS: a new approach to the application of theorem proving to problem solving, *Artificial Intelligence* **2**, 189-208.

Gazzaniga, M. S., 1985. *The Social Brain: Discovering the Networks of the Mind*, New York: Basic Books.

Geertz, C., 1983. Common sense as a cultural system, *Local Knowledge*, New York: Basic Books, 73-93.

Haack, S., 1978. *Philosophy of Logics*, Cambridge: Cambridge University Press.

Hamming, R. W., 1962. *Numerical Methods for Scientists and Engineers*, New York: McGraw-Hill, Chapter $N+1$, 394-401.

Harel, D., 1979. *First Order Dynamic Logic*, Berlin: Springer-Verlag.

Harel, D., 1987. *Algorithmics: The Spirit of Computing*, Reading: Addison-Wesley.

Harman, G., 1973. *Thought*, Princeton: Princeton University Press.

Harman, G., 1986. *Change of View: Principles of Reasoning*, Cambridge: MIT Press.

Hayes, P. J., 1970. Robotologic, *Machine Intelligence 5* (B. Meltzer and D. Michie, eds.), New York: American Elsevier, 533-554.

Jaffe, A., 1984. Ordering the universe: the role of mathematics, in *Renewing U.S. Mathematics* (E. E. David, Jr., et al.), Washington: National Academy Press, 117-162.

Jeffrey, R. C., 1983. *The Logic of Decision*, second edition, Chicago: University of Chicago Press.

Keisler, H. J., 1977. Fundamentals of model theory, *Handbook of Mathematical Logic* (J. Barwise, ed.), Amsterdam: North-Holland, 47-103.

Kowalski, R. A., 1979. *Logic for Problem Solving*, Amsterdam: Elsevier.

Lakatos, I., 1976. *Proofs and Refutations* (J. Worrall and E. Zahar, eds.), Cambridge: Cambridge University Press.

Lewis, D., 1973. *Counterfactuals*, Cambridge: Harvard University Press.

Luce, R. D., and Raiffa, H., 1957. *Games and Decisions*, New York: Wiley.

Mac Lane, S., 1986. *Mathematics: Form and Function*, New York: Springer-Verlag.

McCarthy, J., 1958. Programs with common sense, *Readings in Knowledge Representation* (R. J. Brachman and H. J. Levesque, eds.), Los Altos: Morgan Kaufmann (1985), 300-307.

McCarthy, J., 1977. Epistemological problems of artificial intelligence, *IJCAI-77*, 1038-1044. Reprinted in *Readings in Artificial Intelligence* (B. L. Webber and N. J. Nilsson, eds.), Los Altos: Morgan-Kaufmann (1981), 459-465.

McCarthy, J., 1980. Circumscription—a form of non-monotonic reasoning, *Artificial Intelligence* **13**, 27-39. Reprinted in *Readings in Artificial Intelligence* (B. L. Webber and N. J. Nilsson, eds.), Los Altos: Morgan-Kaufmann (1981), 466-472.

McCarthy, J., 1986. Applications of circumscription to formalizing common-sense knowledge, *Artificial Intelligence* **28**, 89-116.

McCarthy, J., and Hayes, P. J., 1969. Some philosophical problems from the standpoint of artificial intelligence, *Machine Intelligence 4* (B. Meltzer and D. Michie, eds.), New York: American Elsevier, 463-502. Reprinted in *Readings in Artificial Intelligence* (B. L. Webber and N. J. Nilsson, eds.), Los Altos: Morgan-Kaufmann (1981), 431-450.

McDermott, D., 1978. Planning and acting, *Cognitive Science* **2**, 71-109.

McDermott, D., 1986. Critique of pure reason, *Computational Intelligence*, to appear.

McDermott, D., and Doyle, J., 1980. Non-monotonic logic—I, *Artificial Intelligence* **13**, 41-72.

Margaris, A., 1967. *First Order Mathematical Logic*, Waltham: Xerox.

Montague, H. F., and Montgomery, M. D., 1984. How mathematicians develop a branch of pure mathematics, *Mathematics: People, Problems, Results* (D. M Campbell and J. C. Higgins, eds.), Belmont: Wadsworth, Vol. 1, 279-288.

Moore, R. C., 1983. Semantical considerations on nonmonotonic logic, *Eighth International Joint Conference on Artificial Intelligence*, 272-279.

Nilsson, N. J., 1980. *Principles of Artificial Intelligence*, Palo Alto: Tioga.

Polya, G., 1945. *How to Solve It: A new aspect of the mathematical method*, Princeton: Princeton University Press.

Polya, G., 1962. *Mathematical Discovery: On understanding, learning, and teaching problem solving*, Volume 1, New York: Wiley.

Polya, G., 1965. *Mathematical Discovery: On understanding, learning, and teaching problem solving*, Volume 2, New York: Wiley.

Prior, A. N., 1967. Logic, modern, *Encyclopedia of Philosophy* (P. Edwards, ed.), New York: Macmillan, Vol. 5, 12-34.

Quine, W. V., 1953. Two dogmas of empiricism, *From a Logical Point of View*, Cambridge: Harvard University Press.

Quine, W. V., 1970. *Philosophy of Logic*, Englewood Cliffs: Prentice-Hall.

Quine, W. V., and Ullian, J. S., 1978. *The Web of Belief*, second edition, New York: Random House.

Rabin, M. O., 1974. Theoretical impediments to artificial intelligence, *Information Processing 74*, Amsterdam: North-Holland, 615-619.

Rabin, M. O., 1976. Probabilistic algorithms, *Algorithms and Complexity: New directions and recent trends* (J. F. Traub, ed.), New York: Academic Press, 22-39.

Rescher, N., 1968. *Topics in Philosophical Logic*, Dordrecht: D. Reidel.

Rescher, N., and Urquhart, A., 1971. *Temporal Logic*, New York: Springer-Verlag.

Rorty, R., 1979. *Philosophy and the Mirror of Nature*, Princeton: Princeton University Press.

Scott, D. S., 1982. Domains for denotational semantics, *International Conference on Automata, Languages, and Programming.*

Siekmann, J., and Wrightson, G., 1983. *Automation of Reasoning* (2 vols.), Berlin: Springer-Verlag.

Suppes, P., 1957. *Introduction to Logic*, Princeton: Van Nostrand.

Tarski, A., 1946. *Introduction to Logic and the Methodology of the Deductive Sciences*, second edition, New York: Oxford University Press.

Tarski, A., 1959. What is elementary geometry? *The Axiomatic Method, with special reference to geometry and physics* (L. Henkin, P. Suppes, and A. Tarski, eds.), Amsterdam: North-Holland, 16-29.

Touretzky, D. S., 1984. *The Mathematics of Inheritance Systems*, London: Pitman.

Truesdell, C., 1958. Recent advances in rational mechanics, *Science* **127**, 729-739.

Truesdell, C., 1984a. Is there a philosophy of science? *An Idiot's Fugitive Essays on Science: Methods, Criticism, Training, Circumstances*, New York: Springer-Verlag, 471-502.

Truesdell, C., 1984b. The computer: ruin of science, threat to mankind *An Idiot's Fugitive Essays on Science: Methods, Criticism, Training, Circumstances*, New York: Springer-Verlag, 594-631.

Turner, R., 1984. *Logics for Artificial Intelligence*, Chichester: Ellis Horwood.

Weyhrauch, R., 1980. Prolegomena to a theory of mechanized formal reasoning, *Artificial Intelligence* **13**, 133-170.

Wilder, R. L., 1965. *Introduction to the Foundations of Mathematics*, second edition, New York: Wiley.