



---

## Chapter 20

# Logic

*G. Aldo Antonelli*

### Origins of the Modern Conception of Logic

Logic is an ancient discipline that, ever since its inception some 2500 years ago, has been concerned with the analysis of patterns of valid reasoning. Aristotle first developed the theory of the *syllogism* (a valid argument form involving predicates and quantifiers), and later the Stoics singled out patterns of *propositional* argumentation (involving sentential connectives). The study of logic flourished in ancient times and during the middle ages, when logic was regarded, together with grammar and rhetoric (the other two disciplines of the *trivium*), as the foundation of humanistic education.

Throughout its history, logic has always had a *prescriptive* as well as a *descriptive* component. As a descriptive discipline, logic aims to capture the arguments accepted as valid in everyday linguistic practice. But this aspect, although present throughout the history of the field, has since the inception of the modern conception of logic, some 100 or 150 years ago, taken up a position more in the background, and in fact some have argued that it is no longer part of logic proper, but belongs to other disciplines (linguistics, psychology, or what have you). Nowadays logic is, first and foremost, a prescriptive discipline, concerned with the identification and justification of valid inference forms.

The articulation of logic as a prescriptive discipline is, ideally, a two-fold task. The first task requires the identification of a class of valid arguments. The class thus identified must have certain features: not just any class of arguments will do. For instance, it is reasonable to require that the class of valid argument be closed under the relation “having the same logical form as,” in that if an argument is classified as valid, then so is any other argument of the same logical form. It is clear, then, that such an identification presupposes, and rests on, a notion of *logical form*.

The question of what constitutes a good theory of logical form exceeds the boundary of the present contribution, and hence we will not be concerned with



it. We shall limit ourselves to the observation that one can achieve the desired closure conditions by requiring that the class of valid arguments be generated in some uniform way from some restricted set of principles. For instance, Aristotle's theory of the syllogism accomplishes this in a characteristically elegant fashion: subject–predicate propositions are classified on the basis of their forms into a small number of classes, and syllogisms are then generated by allowing the two premises and the conclusion to take all possible forms.

The second task, however, is much harder. Once a class of arguments is identified, one naturally wants to know what it is that makes these arguments *valid*. In other words, in order to accomplish this second task, one needs a general theory of *logical consequence*, and such a theory was not only unavailable to the ancients, it would not be available until the appearance of modern symbolic logic in the late 1800s, when an effort was undertaken to formalize and represent mathematical reasoning, and it would not be completely developed until the middle of the twentieth century.

It is only with the development of the first general accounts of the notion of logical consequence that modern symbolic logic was born. Modern symbolic logic is only a little over 100 years old, dating back to the end of the nineteenth century, and in particular to Gottlob Frege's *Begriffsschrift* (1879) and Richard Dedekind's *Was sind und was sollen die Zahlen?* (1888). With these two works we have the beginning of a rigorous account of logical consequence, an account that will be perfected by Alfred Tarski in the early 1930s.

This chapter focuses on the development of modern symbolic logic from the point of view of the notion of *logical consequence*. After presenting a streamlined account of what is regarded as the crowning achievement of modern symbolic logic, i.e., the systematization of *first-order logic*, we consider consequence relations from an abstract point of view. In the next section we look at consequence relations that are of particular conceptual interest, in that they aim to capture patterns of *defeasible* reasoning in which conclusions are drawn tentatively, subject to being retracted in the light of additional evidence. Finally, we look at nonmonotonic logics devised to capture such defeasible inference.

## First-order Logic

First-order logic (henceforth: FOL) was originally developed (through the work of Frege, Dedekind, Russell & Whitehead, Hilbert, Gödel, and Tarski) for the representation of mathematical reasoning. As such, FOL turned out to be nothing but a stunning success. Its mathematical properties provide a crucial benchmark for the assessment of alternative logical frameworks. We are not going, in this chapter, to provide an introduction to the nuts and bolts of FOL: the interested reader can consult any one of the many excellent introductory texts that are available, such as, for example, Enderton 1972.



FOL provides an implementation of the so-called “no-counterexample” consequence relation: a sentence  $\phi$  is a consequence of a set  $\Gamma$  of sentences if and only one cannot reinterpret the language in which  $\Gamma$  and  $\phi$  are formulated in such a way as to make all sentences in  $\Gamma$  true and  $\phi$  false. An inference from premises  $\psi_1, \dots, \psi_k$  to a conclusion  $\phi$  is *valid* if  $\phi$  is a consequence of  $\{\psi_1, \dots, \psi_k\}$ , i.e., if the inference has no counterexample.

For this to be a rigorous account of logical consequence, the underlying notion of interpretation needs to be made precise. This was accomplished by Alfred Tarski in 1935, who defined the notion of truth on an interpretation (see Tarski 1956 for a collection of his technical papers). In so doing, Tarski overcame both a technical and a philosophical problem. The technical problem has to do with the fact that in FOL quantified sentences are obtained from components that are not, in turn, sentences, so that a direct recursive definition of truth for sentences breaks down at the quantifier case. In order to overcome this problem Tarski introduced the auxiliary notion of *satisfaction*. The philosophical obstacle had to do with the fact that the notion of *truth* was considered suspiciously metaphysical among logicians trained in the environment of the Vienna Circle. This was a factor, for instance, in Gödel’s reluctance to formulate his famous undecidability results in terms of truth.

Tarski’s analysis yielded a mathematically precise definition for the “no-counterexample” consequence relation  $\vDash$  of FOL: we say that  $\phi$  is a consequence of a set  $\Gamma$  of sentences, written  $\Gamma \vDash \phi$ , if and only if  $\phi$  is true on every interpretation on which every sentence in  $\Gamma$  is true. At first glance, there would appear to be something intrinsically infinitary about  $\vDash$ . Regardless of whether  $\Gamma$  is finite or infinite, to check whether  $\Gamma \vDash \phi$  one has to “survey” infinitely many possible interpretations, and check whether any one of them is a counterexample to the entailment claim, i.e., whether any one of them is such that all sentences in  $\Gamma$  are true on it while  $\phi$  is false.

However, surprisingly, in FOL the infinitary nature of  $\vDash$  is only apparent. As Gödel showed in 1929, the relation  $\vDash$ , although defined by universally quantifying over all possible interpretations, can be analyzed in terms of the existence of finite objects of a certain kind, viz., formal proofs. A *formal proof* is a finite sequence of sentences, each one of which is either an *axiom*, or an *assumption*, or is obtained from previous ones by means of one of a finite number of inference rules, such as *modus ponens*. Many different axiomatizations of FOL exist, and a particularly simple and elegant one can be found in Enderton 1972. If a sentence  $\phi$  occurs as the last line of a proof, then we say that the proof is a *proof of  $\phi$* ; and we say that  $\phi$  is *provable from  $\Gamma$* , written  $\Gamma \vdash \phi$ , if and only if there is a proof of  $\phi$  all of whose assumptions are drawn from  $\Gamma$ .

Gödel’s famous completeness theorem of 1929 states that the two relations  $\vDash$  and  $\vdash$  are extensionally equivalent: for any  $\phi$  and  $\Gamma$ ,  $\Gamma \vDash \phi$  if and only if  $\Gamma \vdash \phi$ . This is a remarkable feature of FOL, which has a number of consequences. One of the deepest consequences follows from the fact that proofs are finite objects, and hence that  $\Gamma \vdash \phi$  if and only if there is a *finite* subset  $\Gamma_0$  of  $\Gamma$  such that  $\Gamma_0 \vdash \phi$ .



This, together with the completeness theorem, gives us the *compactness theorem*:  $\Gamma \vdash \phi$  if and only if there is a finite subset  $\Gamma_0$  of  $\Gamma$  such that  $\Gamma_0 \vdash \phi$ . There are many interesting equivalent formulations of the theorem, but the following is perhaps the most often cited. Say that a set of sentences is *consistent* if they can all be made simultaneously true on some interpretation; then the compactness theorem says that a set  $\Gamma$  is consistent if and only if each of its finite subsets is by itself consistent.

Another important consequence of Gödel's completeness theorem is the following form of the Löwenheim–Skolem theorem: if all the sentences in  $\Gamma$  can be made simultaneously true in some interpretation, then they can also be made simultaneously true in some (other) interpretation whose universe is no larger than the set  $\mathbb{N}$  of the natural numbers.

Together, the compactness and the Löwenheim–Skolem theorem are the beginning of one of the most successful branches of modern symbolic logic: model theory. Compactness and the Löwenheim–Skolem characterize FOL; as shown by Per Lindström in 1969, any logical system (meeting certain “regularity” conditions) for which both compactness and Löwenheim–Skolem hold, is no more expressive than FOL (see Ebbinghaus et al. 1994: ch. 13 for an accessible treatment).

Another important consequence of Gödel's completeness theorem has to do with the question of whether and to what extent one can devise an effective procedure to determine if a sentence  $\phi$  is valid, or, more generally, whether  $\Gamma \vDash \phi$  for given  $\Gamma$  and  $\phi$ . First, some terminology. We say that a set  $\Gamma$  of sentences is *decidable* if there is an effective procedure, i.e., a mechanically executable set of instructions, that determines, for each sentence  $\phi$ , whether  $\phi$  belongs to  $\Gamma$  or not. Notice that such a procedure gives both a positive and a negative test for membership of a sentence  $\phi$  in  $\Gamma$ . A set of sentences is *semidecidable* if there is an effective procedure that determines if a sentence  $\phi$  is a member of  $\Gamma$ , but might not provide an answer if  $\phi$  is not a member of  $\Gamma$ . In other words,  $\Gamma$  is semidecidable if there is a positive, but not necessarily a negative test for membership in  $\Gamma$ . Equivalently,  $\Gamma$  is semidecidable if it can be given an effective listing, i.e., if it can be mechanically generated. These notions can be generalized to relations among sentences of any number of arguments. For instance, it is an important feature of the axiomatizations of FOL, such as that of Enderton 1972, that both the set of axioms and the relation that holds among  $\phi_1; \dots; \phi_k$  and  $\psi$  when  $\psi$  can be inferred from  $\phi_1; \dots; \phi_k$  by one of the rules, are decidable. As a result, the relation that holds among  $\phi_1; \dots; \phi_k$  and  $\phi$  whenever  $\phi_1; \dots; \phi_k$  is a proof of  $\phi$  is also decidable. (See Chapter 2, COMPUTATION, for further details on these notions.)

The import of Gödel's completeness theorem is that if the set  $\Gamma$  is decidable (or even only semidecidable), then the set of all sentences  $\phi$  such that  $\Gamma \vDash \phi$  is semidecidable. Indeed, one can obtain an effective listing for such a set by systematically generating all proofs from  $\Gamma$ . The question arises of whether, beside this positive test, there might not also be a negative test for a sentence  $\phi$  being a consequence of  $\Gamma$ . This “decision problem” (*Entscheidungsproblem*) was originally proposed by David Hilbert in 1900, and it was solved in 1936 independently by





Alonzo Church and Alan Turing. The Church–Turing theorem states that, in general, it is not decidable whether  $\Gamma \vDash \phi$ , or even if  $\phi$  is valid. (It is important to know that for many, even quite expressive, fragments of first-order logic the decision problem is solvable; see Börger et al. 1997 for details). We should also notice the following fact that will be relevant later in our discussion: say that a sentence  $\phi$  is *consistent* if  $\{\phi\}$  is consistent, i.e., if its negation  $\neg\phi$  is not valid. Then the set of all sentences  $\phi$  such that  $\phi$  is consistent is not even semidecidable, for a positive test for such a set would yield a negative test for the set of all valid sentences, which would thus be decidable, against the Church–Turing theorem.

### Consequence relations

In the previous section, we defined a consequence relation  $\vDash$  by saying that  $\Gamma \vDash \phi$  if and only if  $\phi$  is true on every interpretation on which every sentence in  $\Gamma$  is true. In general, it is possible to consider what abstract properties a relation of consequence between sets of sentences and single sentences could have. Let  $\vdash$  be any such relation. Consider the following properties, all of which are satisfied by the consequence relation  $\vDash$  of FOL:

*Supraclassicality*: if  $\Gamma \vDash \phi$  then  $\Gamma \vdash \phi$ .

*Reflexivity*: if  $\phi \in \Gamma$  then  $\Gamma \vDash \phi$ ;

*Cut*: If  $\Gamma \vdash \phi$  and  $\Gamma, \phi \vdash \psi$  then  $\Gamma \vdash \psi$ ;

*Monotony*: If  $\Gamma \vdash \phi$  and  $\Gamma \subseteq \Delta$  then  $\Delta \vdash \phi$ .

The first property is supraclassicality, which states that if  $\phi$  follows from  $\Gamma$  in FOL, then it also follows according to  $\vdash$ ; i.e.,  $\vdash$  extends  $\vDash$  (the relation  $\vDash$  is trivially supraclassical). Of the remaining conditions, the most straightforward is reflexivity: it says that if  $\phi$  belongs to the set  $\Gamma$ , then  $\phi$  is a consequence of  $\Gamma$ . This is a very minimal requirement on a relation of logical consequence. We certainly would like all sentences in  $\Gamma$  to be inferable from  $\Gamma$ . It's not clear in what sense a relation that fails to satisfy this requirement can be called a *consequence* relation.

Cut, a form of transitivity, is another crucial feature of consequence relations. Cut is as a conservativity principle: if  $\phi$  is a consequence of  $\Gamma$ , then  $\psi$  is a consequence of  $\Gamma$  together with  $\phi$  only if it is already a consequence of  $\Gamma$  alone. In other words, by adjoining to  $\Gamma$  something which is already a consequence of  $\Gamma$  does not lead to any *increase* in inferential power. Cut is best regarded as the statement that the “length” of a proof does not affect the degree to which the assumptions support the conclusion. Where  $\phi$  is already a consequence of  $\Gamma$ , if  $\psi$  can be inferred from  $\Gamma$  together with  $\phi$ , then  $\psi$  can also be obtained via a longer “proof” that proceeds indirectly by first inferring  $\phi$ . It is immediate to check that FOL satisfies Cut.



It is worth noting that many forms of probabilistic reasoning fail to satisfy Cut, precisely because the degree to which the premises support the conclusion is inversely correlated to the length of the proof. To see this, we adapt a well-known example. Let  $Ax$  abbreviate “ $x$  was born in Pennsylvania Dutch country,”  $Bx$  abbreviate “ $x$  is a native speaker of German,” and  $Cx$  abbreviate “ $x$  was born in Germany.” Further, let  $\Gamma$  comprise the statements “Most  $A$ ’s are  $B$ ’s,” “Most  $B$ ’s are  $C$ ’s,” and  $Ax$ . Then  $\Gamma$  supports  $Bx$ , and  $\Gamma$  together with  $Bx$  supports  $Cx$ , but  $\Gamma$  by itself does not support  $Cx$ . Statements of the form “Most  $A$ ’s are  $B$ ’s” are interpreted probabilistically, as saying that the conditional probability of  $B$  given  $A$  is, say, greater than 50 percent; likewise, we say that  $\Gamma$  supports a statement  $\phi$  if  $\Gamma$  assigns  $\phi$  a probability  $p > 50$  percent.

Since  $\Gamma$  contains “Most  $A$ ’s are  $B$ ’s” and  $Ax$ , it supports the  $Bx$  (in the sense that the probability of  $Bx$  is greater than 50 percent); similarly,  $\Gamma$  together with  $Bx$  supports  $Cx$ ; but  $\Gamma$  by itself cannot support  $Cx$ . Indeed, the probability of someone who was born in Pennsylvania Dutch country being born in Germany is arbitrarily close to zero. Examples of inductive reasoning such as the one just given cast some doubt on the possibility of coming up with a well-behaved relation of probabilistic consequence (see Chapter 21, PROBABILITY IN ARTIFICIAL INTELLIGENCE).

Special considerations apply to monotony. Monotony states that if  $\phi$  is a consequence of  $\Gamma$  then it is also a consequence of any set containing  $\Gamma$  (as a subset). The import of monotony is that one cannot preempt conclusions by adding new premises to the inference. It is clear why FOL satisfies monotony: semantically, if  $\phi$  is true on every interpretation on which all sentences of  $\Gamma$  are true, then  $\phi$  is also true on every interpretation on which all sentences in a larger set  $\Delta$  are true (similarly, proof-theoretically, if there is a proof of  $\phi$  all of whose assumptions are drawn from  $\Gamma$ , then there is also a proof of  $\phi$  – indeed, the same proof – all of whose assumptions are drawn from  $\Delta$ ).

Many consider this feature of FOL as inadequate to capture a whole class of inferences typical of everyday (as opposed to mathematical or formal) reasoning, and therefore question the descriptive adequacy of FOL, when it comes to representing commonsense inferences. In everyday life, we quite often reach conclusions tentatively, only to retract them in the light of further information. For instance, when told that Stellanuna is a mammal, we infer that she does not fly, because mammals, by and large, don’t fly. But the conclusion that Stellanuna doesn’t fly can be retracted when we learn that Stellanuna is a bat, because bats are a specific kind of mammals, and they do fly. So we infer that Stellanuna does fly after all. This process can be further iterated. We can learn, for instance, that Stellanuna is a baby bat, and that therefore she does not know how to fly yet. Such complex patterns of *defeasible* reasoning are beyond the reach of FOL, which is, by its very nature, monotonic.

For these and similar reasons, people have striven, over the last 20 years or so, to devise nonmonotonic formalisms capable of representing defeasible inference. We will take a closer look at these formalisms below, but for now we want to consider the issue from a more abstract point of view.



When one gives up monotony in favor of descriptive adequacy, the question arises of what formal properties of the consequence relation to put in its place. Two such properties have been considered in the literature, for an arbitrary consequence relation  $\vdash$ :

*Cautious Monotony*: If  $\Gamma \vdash \phi$  and  $\Gamma \vdash \psi$ , then  $\Gamma, \phi \vdash \psi$ .

*Rational Monotony*: If  $\Gamma \vdash \neg \phi$  and  $\Gamma \vdash \psi$ , then  $\Gamma, \phi \vdash \psi$ .

Both Cautious Monotony and the stronger principle of Rational Monotony are special cases of Monotony, and are therefore not in the foreground as long as we restrict ourselves to the classical consequence relation  $\vDash$  of FOL.

Although superficially similar, these principles are quite different. Cautious Monotony is the converse of Cut: it states that adding a consequence  $\phi$  back into the premise-set  $\Gamma$  does not lead to any *decrease* in inferential power. Cautious Monotony tells us that inference is a cumulative enterprise: we can keep drawing consequences that can in turn be used as additional premises, without affecting the set of conclusion. Together with Cut, Cautious Monotony says that if  $\phi$  is a consequence of  $\Gamma$  then for any proposition  $\psi$ ,  $\psi$  is a consequence of  $\Gamma$  if and only if it is a consequence of  $\Gamma$  together with  $\phi$ . It has been often pointed out by Dov Gabbay that Reflexivity, Cut and Cautious Monotony are critical properties for any well-behaved nonmonotonic consequence relation (see Gabbay et al. 1994, Stalnaker 1994).

The status of Rational Monotony is much more problematic. As we observed, Rational Monotony can be regarded as a strengthening of Cautious Monotony, and like the latter it is a special case of Monotony. However, there are reason to think that Rational Monotony might not be a correct feature of a nonmonotonic consequence relation. A counterexample due to Stalnaker (1994: 19) involves three composers: Verdi, Bizet, and Satie. Suppose that we initially accept (correctly but defeasibly) that Verdi is Italian, while Bizet and Satie are French. Suppose now that we are told by a reliable source of information that Verdi and Bizet are compatriots. This lead us no longer to endorse the propositions that Verdi is Italian (because he could be French), and that Bizet is French (because he could be Italian); but we would still draw the defeasible consequence that Satie is French, since nothing that we have learned conflicts with it. By letting  $I(v)$ ,  $F(b)$ , and  $F(s)$  represent our initial beliefs about the nationality of the three composers, and  $C(v; b)$  represent that Verdi and Bizet are compatriots, the situation could be represented as follows:

$$C(v; b) \vdash F(s).$$

Now consider the proposition  $C(v; s)$  that Verdi and Satie are compatriots. Before learning that  $C(v; b)$  we would be inclined to reject the proposition  $C(v; s)$  because we endorse and  $I(v)$  and  $F(s)$ , but after learning that Verdi and Bizet are compatriots, we can no longer endorse  $I(v)$ , and therefore no longer reject  $C(v; s)$ . The situation then is as follows:



$$C(v; b) \vdash \neg C(v; s).$$

However, if we added  $C(v; s)$  to our stock of beliefs, we would lose the inference to  $F(s)$ : in the context of  $C(v; b)$ , the proposition  $C(v; s)$  is equivalent to the statement that all three composers have the same nationality, and this leads us to suspend our assent to the proposition  $F(s)$ . In other words, and contrary to Rational Monotony:

$$C(v; b); C(v; s) \vdash F(s).$$

The previous discussion gives a rather clear picture of the desirable features a nonmonotonic consequence relation. Such a relation should satisfy Supraclassicality, Reflexivity, Cut, and Cautious Monotony.

### Varieties of Defeasible Reasoning

A separate issue from the formal properties of a nonmonotonic consequence relation, although one that is strictly intertwined with it, is the issue of how *conflicts* between potential defeasible conclusions are to be handled.

There are two different kinds of conflicts that can arise within a given nonmonotonic framework: (i) conflicts between defeasible conclusions and “hard facts”; and (ii) conflicts between one potential defeasible conclusion and another (many formalisms, for instance, provide some form of defeasible inference rules, and such rules might have conflicting conclusions). When a conflict (of either kind) arises, steps have to be taken to preserve or restore consistency.

All defeasible formalisms handle conflicts of the first kind in the same way: indeed, it is the very essence of defeasible reasoning that conclusions can be retracted when new facts are learned. But conflicts of the second kind can be handled in two different ways: one can draw inferences either in a “cautious” or “bold” fashion (also known as “skeptical” or, respectively, “credulous”). These two options correspond to widely different ways to construe a given body of defeasible knowledge, and yield different results as to what defeasible conclusions are warranted on the basis of such a knowledge base.

The difference between these basic attitudes comes to this. In the presence of potentially conflicting defeasible inferences (and in the absence of further considerations such as specificity – see below), the credulous reasoner always commits to as many defeasible conclusions as possible, subject to a consistency requirement, whereas the skeptical reasoner withholds assent from potentially conflicted defeasible conclusions.

A famous example from the literature, the so-called “Nixon diamond,” will help make the distinction clear. Suppose our knowledge base contains (defeasible) information to the effect that a given individual, Nixon, is both a Quaker and a

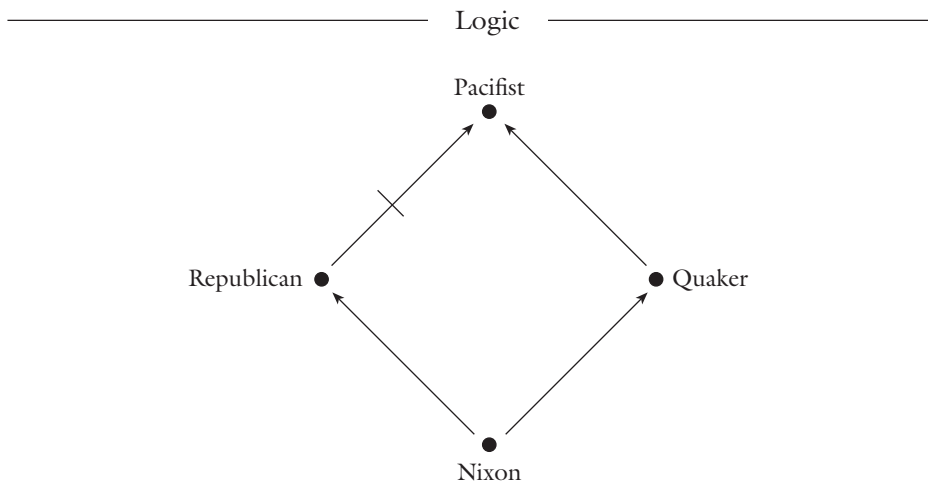


Figure 20.1: The Nixon diamond

Republican. Quakers, by and large, are pacifists, whereas Republicans by and large are not. The question is, what defeasible conclusions are warranted on the basis of this body of knowledge, and in particular whether we should infer that Nixon is a pacifist or that he is not pacifist. Figure 20.1 provides a schematic representation of this state of affairs in the form of a (defeasible) network.

The credulous reasoner has no reason to prefer either conclusion (“Nixon is a pacifist”; “Nixon is not a pacifist”) to the other one, but will definitely commit to one or the other. The skeptical reasoner recognizes that this is a conflict not between hard facts and defeasible inferences, but between two different defeasible inferences. Since the two possible inferences in some sense “cancel out,” the skeptical reasoner will refrain from drawing either one.

Whereas many of the early formulations of defeasible reasoning have been credulous, skepticism has gradually emerged as a viable alternative, which can, at times, be better behaved. Arguments have been given in favor of both skeptical and credulous inference. Some have argued that credulity seems to better capture a certain class of intuitions, while others have objected that although a certain degree of “jumping to conclusions” is by definition built into any nonmonotonic formalism, such jumping to conclusions needs to be regimented, and that skepticism provides precisely the required regimentation. (A further issue in the skeptical/credulous debate is the question of whether so-called “floating conclusions” should be allowed; see Horty [to appear] for a review of the literature and a substantial argument that they should not.)

### Nonmonotonic Logics

As we have mentioned, over the last twenty years or so, a number of so-called “nonmonotonic” logical frameworks have emerged, expressly devised for the





purpose of representing defeasible reasoning. The development of such frameworks represents one of the most significant developments both in logic and artificial intelligence, and has wide-ranging consequences for our philosophical understanding of argumentation and inference.

Pioneering work in the field of nonmonotonic logics was carried out beginning in the late 1970s by (among others) J. McCarthy, D. McDermott, & J. Doyle, and R. Reiter (see Ginsberg 1987 for a collection of early papers in the field). With these efforts, the realization (which was hardly new) that ordinary first-order logic was inadequate to represent defeasible reasoning was for the first time accompanied by several proposals of formal frameworks within which one could at least begin to talk about defeasible inferences in a precise way, with the long-term goal of providing for defeasible reasoning an account that could at least approximate the degree of success achieved by FOL in the formalization of mathematical reasoning. The publication of a monographic issue of the *Artificial Intelligence Journal* in 1980 can be regarded as the “coming of age” of defeasible formalisms.

The development of nonmonotonic (or defeasible) logics has been guided all along by a rich supply of examples. One of the early sources of motivation for the development of nonmonotonic logic comes from database theory. Consider for instance the *closed world assumption*: suppose that you need to travel from Oshkosh to Minsk, so you consult your travel agent, who, not surprisingly, tells you that there are no direct flights. How does the travel agent know? In a sense, he doesn't: his database does not list any direct flights between Oshkosh and Minsk, and he assumes that the database is *complete*. In other words, what we have in this example is an attempt to *minimize* the extension of a given predicate (“flight-between” in this case). Moreover, such a minimization needs to take place not with respect to what the database explicitly contains but with respect to what it implies.

The idea of minimization is at the basis of one of the earliest nonmonotonic formalisms, McCarty's *circumscription*. Circumscription makes explicit the intuition that, all other things being equal, extensions of predicates should be *minimal*. Again, consider principles such as “all normal birds fly.” Here we are trying to minimize the extension of the abnormality predicate, and assume that a given bird is normal unless we have positive information to the contrary. Formally, this can be represented using second-order logic. In second-order logic, in contrast to FOL, one is allowed to explicitly quantify over predicates, forming sentences such as  $\exists P \forall x Px$  (“there is a universal predicate”) or  $\forall P (Pa \leftrightarrow Pb)$  (“ $a$  and  $b$  are indiscernible”). In circumscription, given predicates  $P$  and  $Q$ , we abbreviate  $\forall x (Px \rightarrow Qx)$  as  $P \leq Q$ , and  $P \leq Q \wedge Q \not\leq P$  as  $P < Q$ . If  $A(P)$  is a formula containing occurrences of a predicate  $P$ , then the circumscription of  $P$  in  $A$  is the second-order sentence  $A^*(P)$ :

$$A(P) \wedge \neg \exists Q [A(Q) \wedge Q < P].$$

$A^*(P)$  says that  $P$  satisfies  $A$ , and that no smaller predicate does. Let  $Px$  be the predicate “ $x$  is abnormal,” and let  $A(P)$  be the sentence “All normal birds fly.”



Then the sentence “Tweety is a bird,” together with  $A^*(P)$  implies “Tweety flies,” for the circumscription axiom forces the extension of  $P$  to be empty, so that “Tweety is normal” is automatically true. In terms of consequence relations, circumscription allows us to define, for each predicate  $P$ , a nonmonotonic relation  $A(P) \vdash \phi$  that holds precisely when  $A^*(P) \vDash \phi$ . (This basic form of circumscription has been generalized, for in practice one needs to minimize the extension of a predicate, while allowing the extension of certain other predicates to vary.) From the point of view of applications, however, circumscription has a major shortcoming, namely the absence of a complete inference procedure, due to the fact that, in general, second-order logic lacks such a procedure. The price one pays for the greater expressive power of second-order logic is that there are no complete axiomatizations, as we have for FOL.

Another nonmonotonic formalism inspired by the intuition of minimization of abnormalities is *nonmonotonic inheritance*. Whenever we have a taxonomically organized body of knowledge, we presuppose that subclasses inherit properties from their superclasses: dogs have lungs because they are mammals, and mammals have lungs. However, there can be exceptions, which can interact in complex ways. To use an example already introduced, mammals, by and large, don’t fly; since bats are mammals, in the absence of any information to the contrary, we are justified in inferring that bats do not fly. But then we learn that bats are exceptional mammals, in that they do fly: the conclusion that they don’t fly is retracted, and the conclusion that they fly is drawn instead. Things can be more complicated still, for in turn, as we have seen, baby bats are exceptional bats, in that they do not fly (does that make them unexceptional mammals?). Here we have potentially conflicting inferences. When we infer that Stellaluna, being a baby bat, does not fly, we are resolving all these potential conflicts based on a *specificity* principle: more specific information overrides more generic information. Nonmonotonic inheritance networks were developed for the purpose of capturing taxonomic examples such as the above. Such networks are collections of nodes and directed (“is a”) links representing taxonomic information. When exceptions are allowed, the network is interpreted *defeasibly*. Figure 20.2 gives a network representing this state of affairs. In such a network, if there is a link of the form  $A \rightarrow B$ , then information about  $A$ ’s is more specific than information about  $B$ ’s, and hence should override it. Research on nonmonotonic inheritance focuses on the different ways in which one can make this idea precise.

The main issue in defeasible inheritance is to characterize the set of assertions that are supported by a given network. It is of course not enough to devise a representational formalism, one also needs to specify how the formalism is to be interpreted, and this is precisely the focus of much work in nonmonotonic inheritance. Such a characterization is accomplished through the notion of *extension* of a given network. There are two competing characterizations of extension for this kind of networks, one that follows the credulous strategy and one that follows the skeptical one. Both proceed by first defining the *degree* of a path through the network as the length of the longest sequence of links connecting its endpoints,

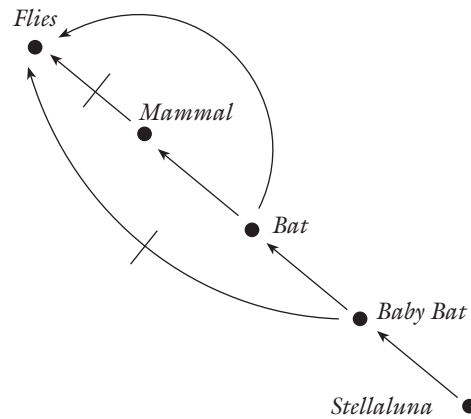


Figure 20.2: An Inheritance network; links of the form  $A \rightarrow B$  represent the fact that typical  $A$ 's are  $B$ 's, and links  $A \nrightarrow B$  represent the fact that typical  $A$ 's are not  $B$ 's

and then building extensions by considering paths in ascending order of their degrees. We are not going to review the details, since many of the same issues arise in connection with default logic (which is treated to greater length below), but Horty 1994 provides an extensive survey. It is worth mentioning that since the notion of degree makes sense only in the case of acyclic networks, special issues arise when networks contain cycles (see Antonelli 1997 for a treatment of inheritance on cyclic networks).

Although the language of nonmonotonic networks is expressively limited by design (in that only links of the form “is a” can be represented in a natural fashion), such networks represent an extremely useful setting in which to test and hone one’s intuitions and methods for handling defeasible information, which are then extended to more expressive formalisms. Among the latter is Reiter’s “Default Logic,” which is perhaps the most flexible among nonmonotonic frameworks. In Default Logic, the main representational tool is that of a *default rule*, or simply a *default*. A default is a *defeasible inference rule* of the form

$$\frac{\eta : \theta}{\xi},$$

(where  $\eta$ ,  $\theta$ ,  $\xi$  are sentences in a given language, respectively called the prerequisite, the justification, and the conclusion of the default). The interpretation of the default is that if  $\eta$  is known, and there is no evidence that  $\theta$  might be false, then the rule allows the inference of  $\xi$ . As is clear, application of the rule requires that a consistency condition be satisfied, and rules can interact in complex ways. In particular it is possible that application of a rule might cause the consistency condition to fail (as when  $\theta$  is  $\neg\xi$ ). Reiter’s default logic uses the





notion of an extension to make precise the idea that the consistency condition has to be met both before and after the rule is applied. Given a set  $\Gamma$  of defaults, an extension for  $\Gamma$  is, roughly, a set of defaults whose consistency condition is met both before and after their being triggered; an extension therefore represents a set of inferences that can be reasonably and consistently drawn using defaults from  $\Gamma$ . More in particular (and in typical circular fashion), an extension for  $\Gamma$  is a maximal subset  $\Delta$  of  $\Gamma$  the conclusions of whose defaults both imply all the prerequisites of defaults in  $\Delta$  and are consistent with all the justifications of defaults in  $\Delta$ .

This definition can be made precise as follows. By a *default theory* we mean a pair  $(W; \Delta)$ , where  $\Delta$  is a (finite) set of defaults, and  $W$  is a set of sentences (a world description). The idea is that  $W$  represents the strict or background information, whereas  $\Delta$  specifies the defeasible information. Given a pair  $(T_1; T_2)$  of sets of sentences, a default such as the equation above is *triggered* by  $(T_1; T_2)$  if and only if  $T_1 \vDash \eta$  and  $T_2 \not\vDash \neg\theta$  (i.e.,  $\theta$  is consistent with  $T_2$ ). Notice how this definition is built “on top” of  $\vDash$ : we could, conceivably, employ a different relation here. Finally we say that a set of sentences  $E$  is an *extension* for a default theory  $(W; \Delta)$  if and only if

$$E = E_0 \cup E_1 \cup \dots \cup E_n \cup \dots,$$

where:  $E_0 = W$ , and

$$E_{n+1} = E_n \cup \left\{ \xi : \frac{\eta : \theta}{\xi} \in \Delta \text{ is triggered by } (E_n, E) \right\}$$

(notice the occurrence of the limit  $E$  in the definition of  $E_{n+1}$ ). There is an alternative characterization of extensions: given a default theory, let  $\mathfrak{S}$  be an operator defined on sets of sentences such that for any set  $S$  of sentences,  $\mathfrak{S}(S)$  is the smallest set containing  $W$ , deductively closed (i.e., such that if  $\mathfrak{S}(S) \vDash \phi$  then  $\phi \in \mathfrak{S}(S)$ ), and such that if a default with consequent  $\xi$  is triggered by  $(S; S)$  then  $\xi \in \mathfrak{S}(S)$ . Then one can show that  $E$  is an extension for  $(W; \Delta)$  if and only if  $E$  is a fixed point of  $\mathfrak{S}$ , i.e., if  $\mathfrak{S}(E) = E$ .

For any given default theory, extensions need not exist, and even when they exist, they need not be unique. Let us consider a couple of examples of these phenomena. Our first example is a default theory that has no extension: let  $W$  contain the sentence  $\eta$ , and let  $\Delta$  comprise the single default

$$\frac{\eta : \theta}{\neg\theta}.$$

If  $E$  were an extension, then the default above would have to be either triggered or not triggered by it, and either case is impossible.



Let us now consider an example of a default theory with multiple extensions. Like before, let  $W$  contain the sentence  $\eta$ , and suppose  $\Delta$  comprises the two defaults

$$\frac{\eta : \theta}{\neg \xi}, \quad \text{and} \quad \frac{\eta : \xi}{\neg \theta}.$$

This theory has exactly two extensions, one in which the first default is triggered and one in which the second one is. It is easy to see that at least a default has to be triggered in any extension, and that both defaults cannot be triggered by the same extension.

These examples are enough to bring out a number of features. First, it should be noted that neither one of the two characterizations of default logic given above gives us a way to “construct” extension by means of anything resembling an iterative process. Essentially, one has to “guess” a set of sentences  $E$ , and then verify that it satisfies the definition of an extension.

Further, the fact that default theories can have zero, one, or more extensions raises the issue of what inferences one is warranted in drawing from a given default theory. The problem can be presented as follows: given a default theory  $(W; \Delta)$ , what sentences  $\phi$  can be regarded as *defeasible consequences* of the theory? On the face of it, there are several options available.

One option is to take the union of the extensions of the theory, and consider  $\phi$  a consequence of a default theory  $(W; \Delta)$  if and only if  $\phi \in E$ , for some extension  $E$ . But this option is immediately ruled out, in that it leads to endorsing contradictory conclusion, as in the second example above. It is widely believed that any viable notion of defeasible consequence for default logic must have the property that the set  $\{\phi : (W; \Delta) \vdash \phi\}$  must be consistent whenever  $W$  is. Once this option is ruled out, only two alternatives are left.

The first alternative, known as the “credulous” or “bold” strategy, is to pick an extension  $E$  for the theory, and say that  $\phi$  is a defeasible consequence if and only if  $\phi \in E$ . The second alternative, known as the “skeptical” or “cautious” strategy, is to endorse a conclusion  $\phi$  if and only if  $\phi$  is contained in *every* extension of the theory.

Both the credulous and the skeptical strategy have problems. The problem with the credulous strategy is that the choice of  $E$  is arbitrary: with the notion of extension introduced by Reiter, extensions are *orthogonal*: of any two distinct extensions, neither one contains the other. Hence, there seems to be no principled way to pick an extension over any other one. This has led a number of researcher to endorse the skeptical strategy as a viable approach to the problem of defeasible consequence. But as shown by Makinson, skeptical consequence, as based on Reiter’s notion of extension, fails to be cautiously monotonic. To see this, consider the default theory  $(W; \Delta)$ , where  $W$  is empty, and  $\Delta$  comprises the two defaults:



$$\frac{}{\vdash \theta}, \text{ and } \frac{\theta \vee \eta : \neg\theta}{\neg\theta}.$$

This theory has only one extension, coinciding with the deductive closure of  $\{\theta\}$ . hence, if we put  $(W; \Delta) \vdash \phi$  if and only if  $\phi$  belongs to every extension of  $(W; \Delta)$ , we have  $(W; \Delta) \vdash \theta$ , as well as  $(W; \Delta) \vdash \theta \vee \eta$  (by the deductive closure of extensions). Now consider the theory with  $\Delta$  as before, but with  $W$  containing the sentence  $\theta \vee \eta$ . This theory has two extensions: one the same as before, but also another one coinciding with the deductive closure of  $\{\neg\theta\}$ , and hence not containing  $\theta$ . It follows that the intersection of the extensions no longer contains  $\theta$ , so that  $(\{\theta \vee \eta\}, \Delta) \not\vdash \theta$ , against cautious monotony. (Notice that the same example establishes a counterexample for Cut for the credulous strategy, when we pick the extension of  $(\{\theta \vee \eta\}, \Delta)$  that contains  $\neg\theta$ .)

It is clear that the issue of how to define a nonmonotonic consequence relation for default logic is intertwined with the way that *conflicts* are handled. The problem of course is that in this case neither the skeptical nor the credulous strategy yields an adequate relation of defeasible consequence. In Antonelli 1999 a notion of *general extension* for default logic is introduced, showing that this notion yields a well-behaved relation of defeasible consequence that satisfies all four requirements of Supraclassicality, Reflexivity, Cut, and Cautious Monotony.

A different set of issues arises in connection with the behavior of default logic from the point of view of computation. As we have seen for a given semidecidable set  $\Gamma$  of sentences, the set of all  $\Gamma$  that are a consequence of  $\Gamma$  in FOL is itself semidecidable. In the case of default logic, to formulate the corresponding problem one extends (in the obvious way) the notion of (semi)decidability given above to sets of defaults. The problem, then, is to decide, given a default theory  $(W, \Delta)$  and a sentence  $\phi$  whether  $(W, \Delta) \vdash \phi$ , where  $\vdash$  is defined, say, skeptically (it doesn't really make a difference computationally whether  $\vdash$  is defined skeptically or credulously). Such a problem is not even semidecidable, the essential reason being that in general, in order to determine whether a default is triggered by a pair of sets of sentences, one has to perform a consistency check. But the consistency checks are not the only source of complexity in default logic. For instance, we could restrict our language to conjunctions of atomic sentences and their negations (making consistency checks feasible). Even so, the problem of determining whether a given default theory has an extension would still be highly intractable (NP-complete, to be precise, as shown by Kautz & Selman 1991), seemingly because the problem requires checking all possible sequences of firings of defaults (see Chapter 2, COMPLEXITY, for these and related notions).

Default logic is intimately connected with certain *modal* approaches to nonmonotonic reasoning, which belong to the family of *autoepistemic logics*. Modal logics in general have proved to be one of the most flexible tools for modeling all sorts of dynamic processes and their complex interactions. Beside the applications in knowledge representation, which we are going to treat below,



there are modal frameworks, known as *dynamic logics*, that play a crucial role, for instance, in the modeling of serial or parallel computation. The basic idea of modal logic is that the language is interpreted with respect to a give set of *states*, and that sentences are evaluated relative to one of these states. What these states are taken to represent depends on the particular application under consideration (they could be epistemic states, or states in the evolution of a dynamical system, etc.), but the important thing is that there are *transitions* (of one or more different kinds) between states. In the case of one transition that is both *transitive* (i.e., such that if  $a \rightarrow b$  and  $b \rightarrow c$  then  $a \rightarrow c$ ) and *euclidean* (if  $a \rightarrow b$  and  $a \rightarrow c$  then  $b \rightarrow c$ ), the resulting modal system is referred to as K45. Associated with each kind of state transition there is a corresponding modality in the language, usually represented as a box  $\Box$ . A sentence of the form  $\Box A$  is true at a state  $s$  if and only if  $A$  is true at every state  $s'$  reachable from  $s$  by the kind of transition associated with  $\Box$  (see Chellas 1980 for a comprehensive introduction to modal logic).

In autoepistemic logic, the states involved are epistemic states of the agent (or agents). The intuition underlying autoepistemic logic is that we can sometimes draw inferences concerning the state of the world using information concerning our own knowledge or ignorance. For instance, I can conclude that I do not have a sister given that if I did I would probably know about it, and nothing to that effect is present in my “knowledge base.” But such a conclusion is defeasible, since there is always the possibility of learning new facts.

In order to make these intuitions precise, consider a modal language in which the necessity operator  $\Box$  is interpreted as “it is known that/” As in default logic or defeasible inheritance, the central notion in autoepistemic logic is that of an *extension* of a theory  $S$ , i.e., a consistent and self-supporting sets of beliefs that can reasonably be entertained on the basis of  $S$ . Given a set  $S$  of sentences, let  $S_0$  be the subset of  $S$  composed of those sentences containing no occurrences of  $\Box$ ; further, let the *introspective closure*  $S_0^i$  of  $S_0$  be the set

$$\{\Box\phi : \phi \in S_0\},$$

and the *negative introspective closure*  $S_0^n$  of  $S_0$  the set

$$\{\neg\Box\phi : \phi \notin S_0\}.$$

The set  $S_0^i$  is called the introspective closure because it explicitly contains positive information about the agent’s epistemic status:  $S_0^i$  expresses what is known (similarly,  $S_0^n$  contains negative information about the agent’s epistemic status, stating explicitly what is not known). With these notions in place, we define an extension for  $S$  to be a set  $T$  of sentences such that:

$$T = \{\phi : \phi \text{ follows from } S \cup T_0^i \cup T_0^n \text{ in K45}\}.$$



Autoepistemic logic provides a rich language, with interesting mathematical properties and connections to other nonmonotonic formalisms. It is faithfully intertranslatable with Reiter's version of default logic, and provides a defeasible framework with well-understood modal properties.

## Conclusion

There are three major issues connected with the development of logical frameworks that can adequately represent defeasible reasoning: (i) material adequacy; (ii) formal properties; and (iii) complexity. Material adequacy concerns the question of how broad a range of examples is captured by the framework, and the extent to which the framework can do justice to our intuitions on the subject (at least the most entrenched ones). The question of formal properties has to do with the degree to which the framework allows for a relation of logical consequence that satisfies the above-mentioned conditions of Supraclassicality, Reflexivity, Cut, and Cautious Monotony. The third set of issues has to do with computational complexity of the most basic questions concerning the framework.

There is a potential tension between (i) and (ii): the desire to capture a broad range of intuitions can lead to *ad hoc* solutions that can sometimes undermine the desirable formal properties of the framework. In general, the development of nonmonotonic logics and related formalisms has been driven, since its inception, by consideration (i) and has relied on a rich and well-chosen array of examples. Of course, there is some question as to whether any single framework can aspire to be universal in this respect.

More recently, researchers have started paying attention to consideration (ii), looking at the extent to which nonmonotonic logics have generated well-behaved relations of logical consequence. As Makinson (1994) points out, practitioners of the field have encountered mixed success. In particular, one abstract property, Cautious Monotony, appears at the same time to be crucial and elusive for many of the frameworks to be found in the literature. This is a fact that is perhaps to be traced back, at least in part, to the above-mentioned tension between the requirement of material adequacy and the need to generate a well-behaved consequence relation.

The complexity issue appears to be the most difficult among the ones that have been singled out. Nonmonotonic logics appear to be stubbornly intractable with respect to the corresponding problem for classical logic. This is clear in the case of default logic, given the ubiquitous consistency checks. But beside consistency checks, there are other, often overlooked, sources of complexity that are purely combinatorial. Other forms of nonmonotonic reasoning, beside default logic, are far from immune from these combinatorial roots of intractability. Although some important work has been done trying to make various nonmonotonic formalisms more tractable, this is perhaps the problem on which progress has been slowest in coming.



## References

- Antonelli, G. Aldo. 1997. "Defeasible inheritance over cyclic networks." *Artificial Intelligence* 92(1): 1–23. [A treatment of nonmonotonic inheritance networks that include cycles, inspired by Kripke's 3-valued approach to truth theories.]
- . 1999. "A directly cautious theory of defeasible consequence for default logic via the notion of general extension." *Artificial Intelligence* 109 (1–2): 71–109. [Introduces a well-behaved consequence relation for default logic, based on a generalization of Reiter's original notion of extension.]
- Börger, Egon, Grädel, Erich, and Gurevich, Yuri. 1997. *The Classical Decision Problem*. Berlin and New York: Springer-Verlag. [A standard reference on the complexity of the decision problem for FOL and many of its fragments.]
- Chellas, Brian F. 1980. *Modal Logic: An Introduction*. Cambridge: Cambridge University Press. [The standard introduction to modal logic. Cover standard possible-world semantics as well as other variants.]
- Ebbinghaus, H.-D., Flum, J., and Thomas, W. 1994. *Mathematical Logic*, 2nd ed. New York and Berlin: Springer-Verlag. [A graduate-level introduction to symbolic logic. Notable for its treatment of Lindström's results.]
- Enderton, Herbert. 1972. *A Mathematical Introduction to Logic*, 2nd ed. New York: Academic Press. New York: Harcourt/Academic Press. [An introduction to logic aimed at advanced undergraduates. Enderton's axiomatization of FOL is widely used.]
- Fitting, Melvin. 1990. *First-order Logic and Automated Theorem Proving*. New York and Berlin: Springer-Verlag. [An introduction to formal logic emphasizing the mechanization of theorem-proving. Notable for its treatment of both resolution and tableaux methods.]
- Gabbay, D. M., Hogger, C. J., and Robinson, J. A., eds. 1994. *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3. Oxford: Oxford University Press. [Contains excellent surveys of all the major nonmonotonic formalisms as well as article addressing the foundations of nonmonotonic logic.]
- Gallier, Jean H. 1986. *Logic for Computer Science: Foundations of Automated Theorem Proving*. New York: Harper & Row. [A technically sophisticated treatment of first-order logic using mainly proof-theoretic methods. Covers Cut-elimination for Gentzen systems, resolution, and many-sorted first-order logic.]
- Galton, Antony. 1990. *Logic for Information Technology*. Chichester and New York: John Wiley & Sons. [A thorough introduction to proof systems for first-order and their metatheory.]
- Ginsberg, M. L., ed. 1987. *Readings in Nonmonotonic Reasoning*. Los Altos, CA: Morgan Kaufman. [A collection of early papers in nonmonotonic logic. Somewhat hard to find, but invaluable.]
- Gödel, Kurt. 1930. "Die Vollständigkeit der Axiome des logischen Funktionenkalküls." *Monatshefte für Mathematik und Physik* 37. [A classic.]
- Horty, John F. 1994. "Some direct theories of nonmonotonic inheritance." In Gabbay et al. 1994: 111–87. [A very clear presentation of all the major issues in defeasible inheritance.]
- . 2002. "Skepticism and floating conclusions." *Artificial Intelligence Journal* 135(1–2): 55–72. [This article addresses some issues in the foundations of defeasible reasoning. Makes the case that, contrary to what many have argued, floating conclusions are not always warranted.]





---

Logic

---

- Kautz, H. and Selman, B. 1991. "Hard problems for simple default logic." *Artificial Intelligence Journal* 49: 243–79. [A seminal work addressing complexity issues in nonmonotonic inference.]
- Lindström, Per. 1969. "On extensions of elementary logic." *Theoria* 35. [The original reference for Lindström's theorems.]
- Makinson, David. 1994. "General patterns in nonmonotonic reasoning." In Gabbay et al. 1994: 35–110. [Best general treatment of the issues concerning abstract consequence relations for various defeasible formalisms.]
- Stalnaker, Robert. 1994. "Nonmonotonic consequence relations." *Fundamenta Informatica* 21: 7–21. [A deep paper assessing several abstract properties of nonmonotonic consequence relations.]
- Tarski, Alfred. 1935. "Der Wahrheitsbegriff in den formalisierten Sprachen." *Studia Logica*, pp. 261–405; English tr. in Tarski 1956: 152–278. [The classic, original treatment of semantics for FOL.]
- . 1956. *Logic, Semantics, and Metamathematics*. Oxford: Oxford University Press, ed. and tr. J. H. Woodger. [A collection of Tarski's mostly technical papers.]