COMPUTER SCIENCES DEPARTMENT
University of Wisconsin-Madison

AN ANALYSIS OF DEFAULT REASONING SYSTEMS IN TERMS OF CONVENTIONAL INFERENCE

by

Kenneth Robert Whitebread

Computer Sciences Technical Report #487

December 1982

AN ANALYSIS OF DEFAULT REASONING SYSTEMS IN TERMS OF CONVENTIONAL INFERENCE

by

Kenneth Robert Whitebread

A thesis submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN - MADISON
1982

TABLE OF CONTENTS

AN ANALYSIS OF DEFAULT REASONING SYSTEMS IN TERMS OF CONVENTIONAL INFERENCE

Kenneth Robert Whitebread

Under the supervision of Professor Larry E. Travis

An understanding of default reasoning is important both for the understanding of human intelligence and for artificial intelligence applications such as robotics and expert systems. Various systems have been built which use heuristic rules to do default reasoning in some limited domain. However, questions about default reasoning remain. To help answer these, a formal model of the process is desirable.

The notions of default theory, due to Reiter, and nonmonotonic theory, due to McDermott and Doyle, have both been proposed as models of default reasoning. These are the only fully developed models of default reasoning to appear in the literature thus far.

It is our contention that these models are based on hypotheses about the nature of default reasoning that have not been sufficiently justified. Therefore, neither of these models can yet be accepted as adequate formalizations of default reasoning.

An example of the type of hypothesis we consider is the supposition, made for both default and nonmonotonic theories, that the assumptions introduced by default reasoning are logical consequences of the reasoner´s initial assumptions according to some nonstandard logic. It is this assumption which leads to the well known nonmonotonicity property that has been attributed to default reasoning.

To support our contention we present a new model called a two-level system. The hypotheses about the nature of default reasoning that form the basis of our definition of a two-level system are quite different from those made in the development of the notions of default and nonmonotonic theory. We argue that a two-level system is at least as viable as a formalization of default reasoning as either a default theory or nonmonotonic theories. Thus, the hypotheses made about default reasoning in developing the latter models are not necessarily the only or the best choices. Also, various properties attributed to default reasoning, such as nonmonotonicity, are seen to be peculiar to default and nonmonotonic theories, not to the phenomenon itself. Finally, we argue that the notion of a two-level system has a number of advantages over the other two approaches.

Acknowledgments

I would like to thank Professor Larry Travis, my advisor, for his encouragement and suggestions. Special thanks are also due to Professor Jon Barwise. Without his advice on questions about mathematical logic and his careful criticism this thesis could not have been completed. Finally, I would like to thank my wife Christine for her constant support during my career as a graduate student at Wisconsin.

1. Introduction

1.1 Background

In ordinary or everyday reasoning humans often introduce new assumptions which they have justified in some way from the information they already know. Typically, one factor in "justifying" the new assumption is the fact that nothing contrary to the assumption is known or believed. For example, my car was in good working order yesterday and I have no reason to believe anything has happened to it in the meantime that would cause it not to start so I assume that it will start today. I may then draw further conclusions from this assumption. I might, for instance, conclude that I can use the car to reach some destination. In artificial intelligence this sort of reasoning is often called _default_ _reasoning_.

The ability to perform default reasoning appears to be important for computer reasoning systems intended to do common-sense reasoning. Various experimental systems have been constructed that employ heuristic default reasoning principles. Winograd [W] has surveyed such systems. However, the relationship of the heuristic principles used to the intuitive notion of default reasoning is unclear. Is the reasoning done by systems using these principles acceptable in the sense that an experienced human reasoner would be likely to produce the same results? Are the principles that have been tried the only possible ones? If there are others, are they better in some way than those already tried? Do these principles represent what a human reasoner actually does when performing default reasoning? To answer such questions a theory of default reasoning is desirable.

McDermott and Doyle have attempted such a theory, as has Reiter. In [MD] and [M] McDermott and Doyle present a mathematical model called a nonmonotonic theory. In [R] Reiter develops an alternative mathematical model called a default theory. These two models of default reasoning are the only fully developed ones that have appeared in the literature so far. Both models turn out to have some odd properties, for example, nonmonotonicity (to be explained below).

1.2 Discussion of Thesis

Once a proposed model has been accepted as a faithful characterization of a phenomenon, properties discovered to be true of the model are uncritically attributed to the phenomenon however odd those properties may be. The step of accepting a proposed model is therefore a crucial one. It is our contention that neither Reiter´s notion of a default theory nor McDermott´s and Doyle´s concept of a nonmonotonic theory can yet be accepted as a model of default reasoning.

To develop a formal model of a phenomenon it is necessary to formulate certain hypotheses about the nature of the phenomenon. It is from these hypotheses that we develop the precise definitions of the model. Although the notions of default theory and nonmonotonic theory appear to be quite different from each other, they share a number of basic hypotheses about the nature of default reasoning. The assumptions made about default reasoning in the developments of these models are often implicit or, at best, presented without any argument for the elimination of alternatives. In the few cases where such arguments were made we will argue that they are weak. The effects of possible alternatives to the assumptions employed, whether explicit or implicit have thus not been thought through and understood.

We will argue that alternative assumptions lead to a preferable model, and we will show that certain properties considered intrinsic to default reasoning on the basis of previous models are instead the result of unjustified assumptions. Therefore, neither Reiter´s model nor that of McDermott and Doyle should be accepted at this point as a satisfactory characterization of default reasoning.

To support our contention we present a new mathematical model called a two-level system. The definition of a two-level system depends on quite different hypotheses about the nature of default reasoning. However, we argue that a two-level system is at least as viable as a model of default reasoning as either a default or a nonmonotonic theory. Thus, the hypotheses made about default reasoning in developing the latter two models are not necessarily the only or the best alternatives. Also, we shall see that various properties attributed to default reasoning are peculiar to default and nonmonotonic theories, not to the phenomenon itself. Finally, we will argue that the notion of a two-level system has certain advantages over the other two approaches.

1.3 Organization

We begin Chapter 2 with some additional examples of default reasoning and a further explanation of the intended scope of the concept. We then present the definitions of default and nonmonotonic theories. In the process we point out the various assumptions on which these definitions are based as well as certain peculiar features of the resulting models.

In Chapter 3 we consider the assumptions pointed out in Chapter 2, comparing them to alternatives and discussing their intuitive basis.

Although it is possible to determine certain isolated assumptions about the nature of default reasoning that underlie the definitions of default and nonmonotonic theories, the presentations of these two models do not contain any kind of overall informal theory of default reasoning. Because such an informal theory is necessary to motivate a formal definition, we present our own informal theory of default reasoning in chapter 4. The formal definition of a two-level system is also presented in this chapter.

In Chapter 5 we compare the notions of two-level system and default theory. Various results concerning the relation of the two models are given. It is argued that the notion of a two-level system characterizes default reasoning at least as well as the notion of a default theory. In addition the properties of a two-level system are compared to those of a default theory.

Chapter 6 contains a comparison of the notions of nonmonotonic theory and two-level system similar to the comparison done in Chapter 5.

Several types of heuristic rule for default reasoning are considered in Chapter 7. We show that such rules can be modelled by a two-level system.

In Chapter 8 we discuss and summarize our results. We also consider the possibility of using two-level systems as models of computer reasoning systems that do default reasoning.

2. Previous Models of Default Reasoning

In this chapter we introduce the notions of default and nonmonotonic theories. We precede the description of these models with a discussion of the idea of default reasoning.

2.1 The Notion of Default Reasoning

The phrase "default reasoning" refers to an aspect of informal or common-sense reasoning. In particular, human reasoners seem frequently to introduce new assumptions during the reasoning process. A statement, determined by some process to be plausible and acceptable, is assumed true and treated as such until and unless discovered to be false. The newly accepted statement normally does not follow by deductive inference from prior assumptions, but once accepted it may be used both as the basis for making additional acceptable assumptions and as the basis for making logical inferences. The statements generated by such means are not necessarily true and sometimes must be discarded later in the reasoning process when additional facts are learned.

For example, if our car ran properly when we used it yesterday, we accept that it will run properly today. Our information about the car does not allow us to infer this statement. Nevertheless, we do expect the car to run today. The fact that the car ran yesterday is only part of the reason for the expectation. It is also based on the fact that *we don´t know* anything which contradicts it. There are many reasons why the car might not run today even though it did yesterday, but we don´t know anything implying that any of these are true. Thus we are saying, in effect, that since the car ran yesterday and since we do not know any information to the contrary, it is reasonable to assume that the car will run today. The information on which we base the adoption of this assumption may be incomplete. Some condition which makes the car inoperable may indeed be true but unknown to us.

Another example of default reasoning concerns the characteristics of birds. We know that most birds fly but that there are exceptions such as penguins. If we are told that a tern is a kind of bird but know nothing else about it, we would probably conclude that it is reasonable to suppose that a tern can fly. This assumption is justified similarly to the previous one. Since most birds fly and since *we don´t know* anything to the contrary about terns, it is reasonable to assume that a tern can fly.

In both of these examples an assumption is introduced on the strength of information which supports the assumption without entailing it and which is such that it does not contradict the assumption. Thus, although the newly introduced assertion is indeed an assumption in the sense that it does not follow logically from the assertions held prior to its introduction, the new assertion is in some way justified by the assertions already believed. The assumption that our car will run is justified both by the existence of certain statements among those we currently believe and by the absence of others. Note that if, for example, we knew that our car had been damaged in an accident while parked on the street overnight, we would not accept so uncritically the assertion that it will run properly today.

The fact that the adoption of assumptions such as those described in our examples depends on the information known to the reasoner suggests that some sort of reasoning process is involved. It has therefore been hypothesized that there is a reasoning process used to justify assumptions of the sort discussed above. This hypothetical process is called default reasoning because it depends on the absence of certain information. The assumptions thus introduced are called default assumptions.

Humans appear to freely intermix default reasoning and conventional reasoning. Given a set of statements representing what one currently knows, one might generate default assumptions directly on the basis of these statements, or one might generate them on the basis of statements not among those currently known but which can be inferred logically. One might also logically infer statements from those currently known or from default assumptions.

2.2 Default Theories

In this section we present the definition of a default theory due to Reiter [R]. We begin by considering the discussion in [R] that prefaces the formal definition. This allows us to identify certain assumptions on which the formal definition is based.

Reiter considers a default reasoning argument about a bird similar to the example given above. His initial rendition of this argument is:

> If x is a bird, then, in the absence of any information to the contrary, infer that x can fly.

Reiter contends that given this rendition the problem is to interpret the phrase "in the absence of any information to the contrary". His choice of interpretation (presented without argument) is "it is consistent to assume that x can fly". He then restates the above default reasoning argument as:

> If x is a bird and it is consistent to assume that x can fly, then infer that x can fly.

In fact, Reiter has already made two important assumptions about the nature of default reasoning at this point, one explicit, the other implicit.

The explicit assumption is that the sense in which the introduction of a default assumption depends on the absence of certain information from the reasoner´s knowledge can be rendered in terms of logical consistency. The implicit assumption concerns the use of the word "infer" in the above arguments. By stating that under certain (still ill-defined) conditions one can infer that x can fly, Reiter is assuming that the default reasoning process is some sort of logical inference process. Although the assumption of an inference process which allows the derivation of default assumptions is basic to Reiter´s approach, no argument in its favor is given and in fact it is apparently not even recognized as open to challenge.

The question arises as to exactly how arguments of the sort illustrated by the above example can be interpreted as logical inferences. One possibility would be to define a suitable form for inference rules, and Reiter appears to begin with this approach in mind. He defines syntactic objects called <u>defaults</u> which are expressions of the form:

$$\alpha: M\alpha_1, \ldots, M\alpha_k/\beta$$

where α, $\alpha_1, \ldots, \alpha_k$, and β are wffs of first order language L. The symbol M is not part of L and is meant to be interpreted as asserting that the wff to which it is attached can consistently be assumed. Defaults are intended to abstract the features common to examples of informal default reasoning arguments thus providing a form of "default inference" rule. Reiter would like the meaning of $\alpha:M\alpha_1, \ldots, M\alpha_k/\beta$ to be something like " if α can be inferred and $\alpha_1, \ldots, \alpha_k$ can consistently be assumed then infer β". Thus, we would have a rule for inferring the default assumption β. To say that a wff can consistently be assumed presumably means that the negation of the wff is not provable from the reasoner´s knowledge (whatever that may be). A default would therefore be a proof rule that refers to unprovability within the very system for which it is a rule. Because of this a logic based on the notion of a default could not simply be a conventional formal theory.

Note that the definition of a default involves another assumption about the nature of default reasoning. Informal default reasoning arguments involve the criterion that the default assumption not be contrary to what the reasoner knows. The simplest way to view this requirement is that the default assumption itself satisfies some relation with "what the reasoner knows". For instance, in the above example from [R], where the relation in question is consistency, it is required that the default assumption "x can fly" be consistent with some unspecified set of assertions. The desired interpretation of a default in effect changes this requirement into the condition that some finite set of assertions, which may not even include the default assumption, not be contrary to the reasoner´s knowledge. Thus, Reiter is postulating that the introduction of a default assumption may depend, not on concluding that the assumption is not contrary to what is known, but on concluding that each member of some set of assertions is not contrary to what is known. Reiter does not provide an argument for this assumption in [R] but does in [RC]. We will discuss this argument in the next chapter.

We have not yet given a precise meaning to a default, and in fact this does not appear to be possible.

The shortcoming of the rough interpretation given above is, of course, that it does not specify what $\alpha_1,..,\alpha_k$ are to be consistent with. In our informal examples the requirement has been that the default assumption not be contrary to the reasoner´s knowledge. If "not contrary" is to mean consistent, we presumably want the set of assertions to be consistent with the reasoner´s knowledge. A precise definition of what constitutes the reasoner´s knowledge would therefore seem to solve our problem.

Perhaps the most obvious choice would be to say that whenever a default is to be applied, the reasoner´s knowledge at that point is just the set of assertions that the reasoner accepts as true at that time. Thus, we could imagine the reasoner beginning with some set of initial assumptions that represent his knowledge prior to any reasoning. If for a given default, say $\alpha:M\alpha_1,...,M\alpha_k/\beta$, it is the case that α is provable from these initial assumptions and each of $\alpha_1,..., \alpha_k$ are consistent with them, then β could be added to the set. The reasoner could then attempt to apply some default to this new set and so on. Unfortunately, this approach fails.

Suppose that the reasoner´s initial assumptions are represented by some set of first order wffs W. Suppose also that the default $\alpha:M\alpha_1,...,M\alpha_k/\beta$ applies to W in the manner just described. Then if we apply this default, we have a new set of formulas representing what is known, namely $W \cup \{\beta\}$. Now, however, suppose we have another default, say $\alpha':M\alpha'_1,\ldots,M\alpha'_n/\beta'$ such that α' is provable from $W \cup \{\beta\}$, $\alpha'_1,\ldots,\alpha'_n$ are consistent with $W \cup \{\beta\}$, and β' is, say, $\sim\alpha_1$. Thus, we could add $\sim\alpha_1$ to $W \cup \{\beta\}$ but the resulting set contains β when it should not because the justification for the earlier inclusion of β has now been undermined.

For example, let $W = \{P(a)\}$ and let the set of defaults be D where $D = \{P(a):MQ(b)/R(a), P(a):MR(a)/\sim Q(b)\}$. If the first default is applied to W in the manner described above, we get $W \cup \{R(a)\}$ as the new set representing what is known. If the second default is then applied to $W \cup \{R(a)\}$, we get $W \cup \{R(a), \sim Q(b)\}$ representing what is known. But the condition for applying the first default was that Q(b) be consistent with what is known. Q(b) is not consistent with the last set derived so we must ask whether we are justified in including R(a) in that set. In fact the sort of interpretation desired for defaults is such that the justification for any default assumption appears to depend on all other default assumptions that the reasoner might accept starting from the given initial assumptions.

Thus, on the one hand defaults as Reiter has defined them apparently cannot be given a well defined meaning. On the other hand Reiter still wishes to treat default assumptions as some sort of logical consequences of the reasoner's initial assumptions. Since the most important function of the proposed model would be, given any set of initial assumptions, to define the set of consequences entailed by those assumptions, Reiter defines such sets of consequences without using any formal definition of inference rule.

The set of theorems of a conventional formal theory can be thought of as the fixed point defined by the operation of deduction using the inference rules of the theory. Similarly, Reiter defines a fixed point for a default theory but without defining explicit inference rules. This fixed point is the set Reiter calls an extension. We now present the formal definitions of default theory and extension.

A <u>default theory</u> is a pair of sets (D,W). W is a set of closed wffs (i.e., sentences) in some first order language L. D is a set of defaults of the form:

$$\alpha: M\alpha_1,\ldots,M\alpha_k/\beta$$

where α, $\alpha_1,\ldots,\alpha_k$, and β are wffs of L. Both D and W are allowed to be infinite but are countable.

If each wff occurring in some member of D is a sentence, the default theory is said to be closed. Since Reiter deals mainly with closed default theories and treats nonclosed theories by relating each one to a certain closed theory, we consider only those which are closed.

Given the definition of default theory we define the extensions of a default theory. Let (D,W) be a closed default theory. For any set of sentences S let $\Gamma(S)$ be the smallest set X satisfying:

1. $W \subseteq X$;
2. X contains the usual axioms for and is closed under the usual inference rules of predicate calculus;
3. If $(\alpha: M\alpha_1,\ldots,\alpha_k/\beta) \in D$ and $\alpha \in X$ and $\sim\alpha_1,\ldots,\sim\alpha_k \notin S$ then $\beta \in X$. (Here, $\sim\alpha$ is the negation of the wff α.)

A set E of sentences is defined to be an extension of (D,W) if $E = \Gamma(E)$. We will use Th(X) to stand for the closure of X as defined in condition 2.

Note that Γ is not a monotonic operator (see [A], for example). Since logicians generally agree that the concept of a monotonic operator constitutes the most general setting in which one can speak of inference rules, the fact that Γ is not nonmonotonic should not surprise us, given our observation of the problems in assigning a formal meaning to defaults.

An extension is any fixed point of the operator Γ. We see that an extension E is deductively closed and includes W, the initial assumptions. In addition E contains various default assumptions. If α:Mα_1,...,Mα_k/β $\in$ D, $\alpha \in$ E, and $\alpha_1,\ldots,\alpha_k$ are consistent with E, then $\beta \in$ E also. Thus, an extension is defined to be any set having those properties that Reiter would expect to be true of the set resulting from applying defaults as inference rules if they could be so applied.

A default theory can have more than one extension because the reasoner might have a choice between inconsistent default assumptions. A default theory can also have no extensions. This appears to be intended to handle situations such as would occur, for example, with a default theory whose only rule is :M(A v ~A)/(A & ~A). The wff A v ~A is a tautology and consistent with any consistent set of wffs while A & ~A is of course inconsistent. We do not wish to have inconsistent extensions of consistent sets of assumptions. In fact, if W is consistent, then the default theory having only this default has no extensions. If W is inconsistent, then the default theory having only this default has the single extension consisting of the whole language L.

The definition of an extension in effect supplies a definition of a logical consequence relation. Although this relation is not actually based on a notion of default inference rule, it still retains consistency as a criterion for satisfying the relation. The consequence relation defined by the notion of an extension is therefore still a nonstandard one.

In examining the notion of a default theory we have discerned three underlying hypotheses about the nature of default reasoning. First, we have the hypothesis that default assumptions are consequences of the reasoner´s initial assumptions according to some nonstandard logical consequence relation. Second, there is the hypothesis that to say that there is no information contrary to a default assumption means that the assumption is consistent with some set of assertions. Finally, it is hypothesized that there may be default assumptions which are justified, not because they themselves are not contrary to the reasoner´s knowledge, but because the members of some finite set of assertions are not contrary to the reasoner´s knowledge.

The notion of a default theory displays two important properties. The first of these is nonmonotonicity.

Suppose we have a conventional formal theory with axiom set A. Suppose also that the assertion p is provable from A. Then any set of axioms containing A as a subset will also have p as one of its theorems. Thus, there is an obvious monotonic relation that holds among pairs of formal theories. Adding to the set of axioms cannot reduce the set of theorems.

If we had succeeded in translating the above ill-defined rule for inferring that x can fly into a well defined one, we would have produced a rule which violates monotonicity. A reasoner might well use such a rule to conclude that x could fly only to find out later that x could not fly. Adding this discovery as a new axiom would make the inference that x can fly impossible. Thus, such a system would be nonmonotonic. In fact, default theories can be nonmonotonic in the following sense. There are default theories, say (D,W) and (D´,W´) such that D is a subset of D´, W is a subset of W´, and yet there exists a formula p such that p is a member of some extension of (D,W) and not a member of any extension of (D´,W´).

Nonmonotonicity is an unusual property for a reasoning process to have. No human reasoning activity observed so far has displayed it. It is therefore important to ask whether the nonmonotonicity displayed by default theories reflects a property that is inherently part of default reasoning. If not, we must question the suitability of default theories as a model of default reasoning.

The second property of interest is the lack of a notion of inference rule in a default theory. As we have already pointed out, defaults, the only candidates for rules in a default theory, do not have a well-defined meaning and therefore cannot be taken as inference rules. Instead, the sets of assertions that are to be considered the consequences of a given set of initial assumptions are defined as certain fixed points. As can be seen from the examples that we have given, informal default reasoning arguments have a deductive flavor about them. It is therefore natural to attempt to define some notion of a rule for introducing a default assumption. Although even ordinary deductive reasoning need not be thought of in terms of inference rules, it is important to discover whether the lack of a notion of rule in default theories reflects an intrinsic feature of default reasoning.

2.3 Nonmonotonic Theories

Examination of the development of the notion of a nonmonotonic theory in [MD] shows that this model is based on the same three hypotheses about default reasoning that we found underlying the definition of a default theory. Furthermore, nonmonotonic theories can indeed be nonmonotonic and do not possess any notion of default inference rule just as is the case for default theories. The chief difference between the two models lies in the attempt by McDermott and Doyle to include within the nonmonotonic theory itself assertions that can be interpreted as "α is consistent with what is known".

Reiter uses expressions of the form Mα in defining defaults and gives them the informal meaning of "it is consistent to assume α". However, the language in which the axioms and consequences of a default theory are expressed does not itself contain such assertions about consistency. If the assertion "it is consistent to assume α" were to be made at all in a default theory, it would have to be made outside of the language in which the reasoner´s initial assumptions, default assumptions, and inferred assertions are expressed. Actually, no such assertion can be made at all in a default theory.

It is assumed by McDermott and Doyle that assertions of the form "α is consistent with what is known" or "it is consistent to assume α" should be expressible in a model of default reasoning. (The reasons for this assumption will be considered in Chapter 6.) Their approach to realizing this involves the introduction of a special symbol into the language in which the assertions of a nonmonotonic theory are to be expressed.

We begin with a formal language, L_M, which is based on an ordinary first order language. L_M contains wffs built up in the usual way from quantifiers, connectives, and atomic formulas consisting of predicate symbols applied to suitable arguments. However, a special symbol, M is included in the alphabet for L_M. If α is any wff of L_M, then $M\alpha$ is also a wff of L_M. The goal of McDermott and Doyle is to develop a system in which a formula such as $M\alpha$ can be interpreted to mean that it is consistent with "what is known" to believe α.

Along with L_M we assume a set of logical axioms and inference rules exactly analogous to those for the predicate calculus. From these axioms and rules, provability is given the usual syntactic definition. Thus, up to this stage, the symbol M is transparent to the definitions made.

The next step is to define the set of formulas which are "nonmonotonically provable" from a set, A, of wffs of L_M, thus defining the notion of nonmonotonic provability. This requires some intermediate definitions. For A and S sets of wffs, let

$$\mathrm{As}_A(S) = \{M\beta \mid {\sim}\beta \notin S\} - \mathrm{Th}(A);$$

$$NM_A(S) = Th(A \bigsqcup As_A(S)).$$

Here, we assume that A contains the usual axioms for the predicate calculus and define Th as before. We then define the class of *fixed points of A*, FP(A), for any A by

$$FP(A) = \{S \mid S \text{ a set of wffs } \& \ NM_A(S) = S\}.$$

Finally, we define the set of wffs TH(A) to be the intersection of all sets in the class FP(A) if FP(A) is not empty and L_M if FP(A) is empty. TH(A) represents the set of wffs nonmonotonically provable from A and is called the *nonmonotonic* theory of A. $As_A(S)$ is called the set of *assumptions from S*.

The idea behind the concept of a nonmonotonic theory is that the ability to prove $M\alpha$, where $M\alpha$ is interpreted as "α is consistent with what is known", will allow us to express rules for default reasoning similar to those which Reiter hoped to capture through the definition of a default. Thus, if we have a theory in which it is possible to prove expressions of the form $M\alpha$, we can add to such a theory axioms of the form:

B(x) and MF(x) implies F(x).

If B(x) is interpreted as "x is a bird" and F(x) as "x can fly", this axiom would express the sort of rule intended by Reiter in the example given above. The difficulty is that expressions of the form $M\alpha$ are not provable since the definition of a nonmonotonic theory depends on fixed points not on explicit inference rules just as the definition of a default theory does.

Note that the definition of the class FP(A) is similar to the definition of the class of extensions of a default theory. However, the extensions of a default theory are treated as alternative sets of beliefs. The set of formulas accepted as true by the reasoner, given the sets W and D, may be any one of the extensions of (D,W). Here, instead of treating each member of FP(A) as one possible set of beliefs for the reasoner, the set of formulas defined to be accepted as true given that the reasoner accepts the formulas of A is TH(A), which is just the set of formulas common to every member of FP(A). We will discuss this difference between default theories and nonmonotonic theories in Chapter 6.

Although nonmonotonic theories look quite different from default theories, these two models share certain basic hypotheses about the nature of default reasoning and also have certain unusual properties in common. In later chapters we will consider the relationships among such properties and hypotheses.

Because the formulas of a default theory are expressed in an ordinary first order language, there is no difficulty in seeing how one could attach meanings to them. The only difficulty connected with interpretation arises from the definition of a default. Since defaults are not part of the language in which the assertions of the model are expressed, Reiter´s method of getting around the problem by defining an extension as a fixed point leaves us with a model for which we know how to define an interpretation.

In the case of nonmonotonic theories, however, the language includes a symbol, M, for which there is no standard interpretation. In [MD] an attempt is made to provide a semantics for nonmonotonic theories. As Davis points out in [D], this attempt does not succeed. In [M] a modified notion of nonmonotonic theory is presented and for this version a semantics is given. The difference between the definition of a nonmonotonic theory in [M] and the one we have presented here is that the modified definition depends on axioms and inference rules for a modal logic instead of axioms and rules for the predicate calculus. The syntactic definition of a nonmonotonic theory remains the same in both cases. Furthermore, the modified definition is based on the same hypotheses and displays the same properties as the original version. When we discuss nonmonotonic theories in Chapter 6 we will be concerned with the informal meaning intended for this model. As our analysis and conclusions will apply equally to the original definition and the modified version, we will consider only the original definition as just presented.

3. Alternative Hypotheses About the Nature of Default Reasoning

In this chapter we consider the three hypotheses that we have found to underlie the definitions of default and nonmonotonic theories. Their plausibility will be considered, and alternatives will be suggested. Further, we argue that the suggested alternatives are more intuitive.

3.1 The Relation of Default Assumptions to Initial Assumptions

It is natural to conceive of default reasoning in terms of some sort of inference. Examples of the process consist of "premises" and a "conclusion". However, the nature of the conclusion is open to argument. For instance, the exemplary rule concerning a bird given in the previous chapter was stated according to Reiter´s (and McDermott´s and Doyle´s) interpretation. It had the form:

> If x is a bird and it is consistent to assume that x can fly, then infer x can fly.

To apply such a rule, one would need two premises, one asserting that x is a bird and the other that x's ability to fly could consistently be assumed. According to this interpretation, one would then conclude the default assumption "x can fly". But does a human reasoner actually conclude a default assumption as the result of a default reasoning process?

There is another possible version of the above rule. It is as follows:

> If x is a bird and it is consistent to assume that x can fly, then it is reasonable to assume x can fly.

In this case, the reasoner would be concluding from his argument, not that x can fly, but that it is reasonable to assume that x can fly.

The first form of the rule manifests the view that a default assumption is itself some sort of logical consequence of the reasoner's prior assumptions. This view seems a distortion of the intuitive notion of a logical consequence relation. The idea behind the notion of logical consequence is that a given assertion α is a consequence of a set of assertions just if it must hold whenever all members of the set hold. Yet this is clearly not the case with a default assumption.

If I believe that x is a bird and that it is consistent to assume x can fly, I am not as a result convinced that x can fly. I know that it might turn out that x cannot fly. What I do believe is that I can reasonably treat "x can fly" as a working assumption. The second form of the rule manifests this point of view. The default assumption is treated as a statement that is not *derived* but *assumed because of the successful derivation of a statement about it*. The relation between default assumptions and the reasoner´s knowledge appears according to this interpretation as an indirect one, depending on drawing a conclusion from that knowledge about the assumption.

Thus, while informal default reasoning arguments do indeed appear to be inferential in nature, we have at least two ways of interpreting the apparent inference. On the one hand, we can assume that the reasoner is actually inferring the default assumption. On the other, we can suppose that what is being inferred is an assertion about the reasonableness or plausibility of the default assumption. The second of these hypotheses appears the more intuitive.

3.2 Consistency as a Criterion for the Introduction of Default Assumptions

We would not wish to assume that x can fly if it were possible for us, without receiving any new information, to conclude that x cannot fly. That is, if the adoption of a default assumption is to be correct, it is necessary that the assumption be consistent with the reasoner´s knowledge (both at the time of adoption and during any subsequent reasoning). This observation appears to be the basis for the hypothesis (accepted by Reiter and by McDermott and Doyle) that the part of a default reasoning argument asserting that something is not contrary to what the reasoner knows can be interpreted as an assertion about consistency.

As we saw in the previous chapter, the general form assumed for default reasoning by these authors stipulates that default assumptions are to be justified because some finite set of assertions is not contrary to the reasoner´s knowledge. We consider this generalization in the next section. Here we wish to examine the suitability of consistency just in those informal default reasoning arguments where it is the default assumption itself which must not conflict with what the reasoner knows.

The following example is essentially given in [RC] although the conclusions drawn from it are our own. Consider the following arguments:

> If x is a high-school dropout and it is not contrary to what is known to assume that x is an adult, then it is reasonable to assume x is an adult.
>
> If x is an adult and it is not contrary to what is known to assume that x is employed, then it is reasonable to assume x is employed.

Both these arguments are intuitively correct.

Suppose we are told that Arnold is a dropout, but we know nothing about his employment status so we can deduce neither that Arnold is or is not employed. Suppose also that we do not know Arnold´s age. We might conclude that it is reasonable to assume that Arnold is an adult, but we would not be willing to go on from there and conclude that we can also reasonably assume that he is employed. The reason seems clearly to be that we do indeed know something "contrary" to the assumption that Arnold is employed. We know that he is a dropout. The assumption that he is employed is <u>logically</u> <u>consistent</u> with what we know. Nevertheless, his status as a dropout constitutes sufficient evidence against the assumption to make us feel that it should not be accepted.

Thus, although it is necessary that a default assumption be consistent with what we know in order to be correct, logical consistency alone does not seem to be an adequate interpretation of the phrase "not contrary to". In fact, the examples of default reasoning considered in the literature so far do not seem to point to any obvious interpretation. An alternative approach is to define a model in such a way as to allow a wide range of possible definitions for the notion of "not contrary to".

The above discussion does suggest that the interpretation of the assertion that a default assumption is not contrary to what the reasoner knows might be based on some notion of "evidence against" a statement. The idea is that if an assertion is inconsistent with the reasoner´s knowledge, then this fact surely constitutes evidence against the assertion. In addition, facts known by the reasoner that do not contradict the assertion but make its likelihood questionable also constitute such evidence. It remains to be seen whether any well-defined form of this concept of "evidence against" a statement can be developed, but it might, for example, be dealt with in terms of the notion of inductive reasoning.

3.3 The Scope of "Not Contrary to"

In this section we consider whether or not it is desirable to postulate the existence of default assumptions that are justified because the members of some finite set of assertions are not contrary to the reasoner´s knowledge.

The examples of informal default reasoning arguments that we have presented all are based on the requirement that the default assumption itself not be contrary to what is known. This seems to be natural and, as Reiter points out in [R], it is difficult to think of natural examples of default reasoning that do not fit this pattern.

In [RC], however, an argument is presented for the existence of cases in which justification depends on a finite set of assertions. The argument is based on examples similar to the one given above concerning the likelihood of a dropout being employed, but the analysis given relies on the hypothesis that "not contrary to" means consistent with. Thus, the two rules for default reasoning stated above are rendered as:

> If x is a high-school dropout and it is consistent with what is known to assume that x is an adult, then infer that x is an adult.

> If x is an adult and it is consistent with what is known to assume that x is employed, then infer that x is employed.

Suppose as before that we know that Arnold is a dropout and do not know anything about either his employment status or his age. If human default reasoning follows the pattern realized in the above two rules, one should be able to use the rules to conclude that Arnold is an adult and, from his adulthood, that he is employed.

It is stated in [RC] that the conclusion that Arnold is employed (or, in our view, that his employment can reasonably be assumed) should not occur. However, the problem is claimed to lie in the fact that in our "default inference rules" we are only requiring that the default assumption itself satisfy the condition of not being contrary to the reasoner´s knowledge. Accordingly, the solution offered is to replace the second of the above two rules with a rule that says something like:

> If x is an adult and it is consistent with what is known to assume both that x is not a dropout and x is employed, then infer that x is employed.

As was pointed out in the previous section, however, this example can also be interpreted as indicating that the criterion of consistency may be inadequate.

Suppose the two rules are stated in the following fashion:

> If x is a high-school dropout and there is no evidence against the assumption that x is an adult, then it is reasonable to assume x is an adult.
>
> If x is an adult and there is no evidence against the assumption that x is employed, then it is reasonable to assume x is employed.

Then, since the fact that Arnold is a dropout constitutes evidence against the assumption that he is employed, we do not find ourselves making an unwarranted assumption.

We conclude that there is no intuitive reason for the hypothesis that a default assumption may need to be justified on the basis of assertions other than itself being deemed not contrary to the reasoner´s knowledge. At the same time, natural examples of default reasoning certainly do appear to require that the default assumption not be contrary to the reasoner´s knowledge. It therefore seems reasonable to postulate that in default reasoning the object which must not be contrary to the reasoner´s knowledge is just the default assumption itself.

4. Definition of a Two-Level System

Our next task is to define the notion of a two-level system. To motivate our definition and to provide evidence for its appropriateness, we first present what we believe to be the most intuitive informal theory of default reasoning.

One must have some informal theory in mind in order to define any formal model. Without the intermediate step of making one´s informal theory explicit, it is difficult to understand whether any problems encountered in the formal model are a result of an incorrect formalization of an intuitively correct informal theory or a result of the incorrectness of the informal theory itself. Also, it is the underlying informal view that makes the model convincing, and it is therefore important to clearly understand what that view is in order to judge the reasonableness of the formal model. In the development of both default and nonmonotonic theories, the presentation of an explicit informal theory is omitted. We will argue in later chapters that this obscures the causes of a number of problems in the two models.

4.1 An Informal Theory of Default Reasoning

The general idea behind default reasoning is the concept of a reasoning process which depends on an assertion being found not contrary to the reasoner´s knowledge. We begin the development of our informal theory by making this notion somewhat narrower and more precise.

Consider the example concerning an automobile that was presented earlier. We can state that example as:

> Since my car ran properly yesterday and since I know nothing contrary to the assumption that it will start today, it is reasonable to assume that my car will start.

We will treat this example as typical of the reasoning process referred to as default reasoning, and develop our theory by abstracting and generalizing it.

We first note that the above argument can be considered to consist of two premises and a conclusion:

> My car ran properly yesterday. I know nothing contrary to the assumption that it will start today. Therefore, it is reasonable to assume that my car will start.

The import of the first premise is that it provides evidence or support for the assumption that the car will start. The second premise tells us that there is no known reason to suppose that the car will not start. From these two facts it is concluded that one can reasonably assume that the car will start. (We state the conclusion in this way on the basis of the arguments given in the previous chapter. Further arguments for this form of conclusion will be discussed at a later point in our development.)

We have already argued that the most reasonable course is to suppose that the assertion which is taken not to be contrary to the reasoner´s knowledge is just the default assumption itself. Furthermore, it seems clear that, as in the above example, one does not accept an assumption as plausible unless there is some sort of definite, positive evidence indicating that the assumption is (or is likely to be) true. It is therefore natural to view default reasoning as the process of concluding that there is reason to suppose an assumption is true and no known reason to suppose it is not true. We take this view and postulate that default reasoning is the process of generating arguments of the form:

There is evidence to support α. Nothing contrary to

> α is known. Therefore, it is reasonable to assume α.

We will call premises of the form:

> Nothing contrary to α is known
>
> or
>
> α is not contrary to what is known

default premises.

Several questions occur about this abstract form for a default reasoning argument. First, is there some general definition of "evidence to support" a default assumption? Second, what does it mean to say that nothing contrary to a default assumption is known? Finally, how would such an argument be carried out?

The concept of evidence supporting an assertion is basic to inductive logic. It is therefore natural to look to that field for a definition of "evidence to support" a default assumption. However, the exact form the definition would take is not clear. We will therefore leave our notion of evidence supporting the default assumption undefined. The notion of a two-level system as defined in the next section will simply provide a framework in which a wide range of definitions of this concept would be possible.

The problem of interpreting the notion of not knowing anything contrary to a default assumption has two parts. First, what is meant by "what is known" ? Second, what is meant by "not contrary to"?

The most natural definition of what is known by the reasoner is the set of assertions accepted by him as true. However, this set obviously changes during the reasoning process. Suppose we think of the reasoner as beginning the reasoning process with some initial set of assumptions. We can think of them as representing what he knows at that point. From the point of view of default reasoning this set can be changed in two ways. First, the reasoner may derive a new assertion from his initial assumptions by conventional deductive inference. Second, the reasoner might adopt a default assumption. Thus, the set of assertions accepted by the reasoner would change during the course of the reasoning process. Therefore, we will interpret "what is known" to mean _what the reasoner knows at the current point_ in the reasoning process. This interpretation will be refined below.

As already discussed in the previous chapter, the problem of defining "not contrary to" , like that of defining "evidence to support" does not have a ready solution. It seems clear that consistency is necessary but not sufficient as a correctness criterion. The notion of inductive reasoning may also be applicable. Various heuristic criteria are also possible, as will be discussed in a later chapter. We will therefore leave "not contrary to" unspecified and attempt to define our model in such a way as to allow a wide range of definitions.

We next consider the question of how a default reasoning argument might be carried out.

Viewing default reasoning as a process of inferring an assertion about a default assumption rather than as one of inferring the default assumption was discussed in the previous chapter. Similarly, a default premise is naturally viewed as an assertion about the default assumption. We will adopt these views here. However, a question arises concerning the nature of any reasoning process which could yield the conclusion of a default reasoning argument.

The reasoner, in carrying out a default reasoning argument, must be reasoning from some of his knowledge. Yet, the default premise refers to the reasoner´s knowledge, thus apparently leading to problems of self-reference. Similarly, we wish to interpret the conclusion of a default reasoning argument as asserting that a given default assumption is reasonable and so may be assumed. If the reasoning that leads to such an assertion is itself based on the reasoner´s knowledge (taken to be the reasoner´s assumptions), how can it lead to the conclusion that another assumption can be made? Our solution to these difficulties is to employ the following observation.

Consider again the example of default reasoning concerning a car. The default premise of this argument asserts that the reasoner does not know anything contrary to the assertion that his car will start. In determining the correctness of this default premise must the reasoner consider <u>everything</u> he knows? It seems clear that he need not. For example, the reasoner´s knowledge concerning his own reasoning processes has nothing to do with starting his car. This observation leads us to conclude that "what the reasoner knows" need refer only to knowledge that could affect the argument.

We hypothesize that the reasoner´s knowledge is divided into components. The reasoning involved in adding a default assumption to one component will be done, not within that component, but in a second component in which the reasoner can reason <u>about</u> the first one. In fact, we will assume that there are only two components.

Examples of default reasoning generally involve introducing a default assumption about the world exterior to the reasoner (or at least exterior to his reasoning process). We will therefore view the reasoner´s knowledge as consisting of two components or levels. The knowledge at the first or <u>object</u> level concerns everything exterior to the reasoning process. It is to this level that the reasoner may introduce default assumptions. The knowledge at the second or <u>meta</u> level represents the reasoner´s rules for introducing default assumptions. It is at this level that the reasoning leading to the adoption of a default assumption is done. Here we use the terms "metalevel" and "object level", not in the formal sense associated with the notion of a metalanguage in which it is possible to talk about sentences of an object language as objects, but in the informal sense conveyed by the observation that, for example, a default premise can·be seen as an assertion about an assertion. We consider in a later chapter the possibility of the rules for default reasoning being themselves the subject of default reasoning.

Given the above notion of default reasoning as a process involving two levels, we next develop our notion of "what the reasoner knows" and fit it into our overall theory. We have already chosen to interpret "what the reasoner knows" to mean what he currently knows. We now make the interpretation more precise, taking into account our assumption of two levels.

The notion of what is known, as it applies to a default reasoning argument, is to be taken as what is known at the object level. We therefore need to give some specification of what is to constitute the reasoner's current object-level knowledge during the reasoning process.

There are three types of assertion that can occur at the object level. First, there are the reasoner's initial object-level assumptions. Next, there are default assumptions, and finally there are assertions derived by conventional deductive inference. The reasoner's current object-level knowledge can therefore reasonably be taken to consist of no more than the initial assumptions, any default assumptions introduced prior to that point, and any conventionally inferred assertions derived prior to that point. We tentatively eliminate conventionally inferred assertions.

Our reason for counting only initial and default assumptions as current knowledge stems from the difference between the roles played by assumptions (whether initial or default) and inferred assertions in our view of default reasoning. The reasoning done to introduce default assumptions is hypothesized to occur at the metalevel so that the only reasoning occurring at the object level must be conventional inference. From the point of view of conventional deductive inference default and initial assumptions appear as the same type of assertion. Neither can be inferred conventionally from other assertions occurring at this level and therefore represent axioms. It is therefore natural to think of just the reasoner´s current object-level assumptions as his current object-level knowledge. (In a later chapter we will see an illustration of how one might incorporate the more general definition of assumptions plus inferred assertions into our model, as well as a possible reason for doing so.)

So far, then, we are viewing default reasoning as a process divided into two levels. Default assumptions are introduced at the object level by reasoning done at the metalevel. This metalevel reasoning determines that the given default assumption is plausible, given the reasoner´s current object level assumptions, and given that these assumptions are not contrary to the default assumption. Object-level reasoning, on the other hand, consists of conventional deductive inference from the reasoner´s current object-level assumptions.

A final consideration for our informal theory involves the nature of the metalevel reasoning. It would clearly be desirable to be able to postulate that the reasoning done at the metalevel is also conventional deductive inference, and in fact, such a hypothesis seems reasonable. The reasoning being carried out is about the set of current object level assumptions and its relation to a default assumption which is also an object level assertion. It is not about any assertions from the metalevel. We therefore do not encounter the difficulties that occur if we try to view this reasoning as being done from the same assumptions to which it is intended to add. Thus, we add to our informal theory the hypothesis that the metalevel reasoning which results in the adoption of a default assumption is itself conventional inference.

Before turning to the formal definition of our model, we discuss one additional argument for the intuitive appeal of the two-level approach to default reasoning.

A default assumption appears to be treated in two different ways by a human reasoner. When deciding whether the assumption is justified or not, the reasoner recognizes that he is indeed dealing with an assumption. But once accepted, the default assumption is treated in the same way as any other assertion that the reasoner believes. For example, if I ask myself whether or not my car will start, I realize that the assertion that it will start is an assumption and cannot be guaranteed by me.

Yet, having concluded that I can assume the car will start, in subsequent reasoning the assumption is treated as a fact.

This behavior fits our two-level approach. By postulating that reasoning takes place on two levels we account for the two ways in which a human views a default assumption. At the metalevel a potential default assumption is recognized as indeed an assumption. When reasoning at the object level, however, the reasoner no longer questions the assumption but accepts it on a par with all other object-level assumptions and uses it as a basis for further reasoning.

4.2 The Definition of a Two-Level System

We now need to define a formal model. The approach that we take is guided by the purpose that the model is intended to serve. Our intention is to develop a model that can serve as a basis for justifying the claims we have made concerning the hypotheses on which default and nonmonotonic theories are based. We therefore wish to define a model based on formal concepts that are as simple and well understood as possible even if the internal mechanisms of the model may not be exactly analogous to human behavior. At the same time, the model should in some reasonable sense be a formalization of the informal theory presented above. As we will see below, the model resulting from our approach also displays potential as a way of representing computer reasoning systems that do default reasoning.

The basis of our informal theory is the postulate that default reasoning is a process of reasoning about assumptions. This view led to the conception of the reasoner´s knowledge being divided into levels in such a way that the assumptions of one level could be reasoned about at the other level. The simplest way to formally represent reasoning about assumptions is as a process of reasoning about representations of assumptions. In particular, we will employ the common notions of metalanguage and object language. A default assumption will be represented by an appropriate sentence in an object language, and reasoning about the default assumption will be represented by reasoning in the metalanguage about the object language sentence corresponding to the default assumption.

We have already hypothesized that the reasoning done at the object level is conventional inference and noted that the reasoner´s current object level assumptions simply appear as a set of axioms in the context of reasoning at this level. A well known way of representing conventional inference from a set of axioms is by way of a formal theory defined as a set of inference rules and a set of axioms in a formal language. It is reasonable to suppose that the assertions dealt with by the reasoner can be expressed in a formal language and that the inferences he performs at the object level can be defined by some set of inference rules. Further, although the reasoner can add new object-level assumptions, it does not seem plausible that his rules of inference change. We therefore conceive each set of assertions that might become the reasoner´s current object-level assumptions at some point during the reasoning process as being represented by a formal theory. Also, the object-level reasoning that can be done from any such set of current assumptions is represented by deduction within the corresponding formal theory.

The reasoner reaches a given set of current-object level assumptions by default reasoning. If no default assumptions are introduced, the current object-level assumptions are just the initial assumptions. Thus, we can think of default reasoning as a process of changing from one set of assumptions to another. This observation suggests the idea of representing the effects of default reasoning in terms of the sets of current object level assumptions it might produce. We can represent each such set of assumptions by a formal theory. Default reasoning could then be represented as a process of changing from one formal theory to another. Therefore, one component of our model will be a collection of formal theories, defined in some formal language, each of which represents one of the sets of current object level assumptions that might be arrived at during the reasoning process. One such formal theory will contain only the initial object level axioms and so will represent the starting point of the process.

Object level reasoning is to be represented by reasoning in a formal theory whose axioms correspond to the reasoner´s current object level assumptions. The adoption of a new default assumption corresponds to replacing one formal theory, whose axioms are the current object level assumptions, by a second formal theory whose axioms are those of the first plus the default assumption. We represent the metalevel reasoning that leads to this change in theories in yet another formal theory, one defined in a language in which we can talk about the sentences corresponding to the reasoner´s object level assumptions.

If the adoption of a default assumption is viewed in the terms just described, then we can think of default reasoning as inferring at the metalevel that the set of assertions representing the current object-level assumptions along with the new default assumption can be accepted together as the new current assumptions. We would expect the formal theory representing metalevel reasoning to allow us to carry out such an inference. In particular, it is natural to expect the informal notion of a rule for introducing a default assumption to be interpreted as an axiom giving the conditions under which the default assumption along with a set of current assumptions constitute a new set of current assumptions. In stating our formal definition we will actually impose a weaker requirement. We will only insist that it be possible to conclude in the formal theory for metalevel reasoning that the axiom set of each of the formal theories representing possible current object level assumptions is an acceptable set of assumptions and that this conclusion be possible for no other set of object-level assertions.

Let us summarize the general idea for our model. It will consist of a collection of formal theories representing the various sets of current object-level assumptions that the reasoner might arrive at if starting with given initial object-level assumptions and certain given rules about the introduction of new default assumptions. These theories will be expressed in some given formal language. Rules for introducing default assumptions will be contained in an additional formal theory. This theory will be expressed in a language which can serve as a metalanguage with respect to the one in which the other theories are expressed. The reasoning done to justify the introduction of a default assumption will be represented by inference within this metatheory about the sentence corresponding to the default assumption.

The theories representing the various possible current assumption sets can be any formal theories, but the theory in which we are to do default reasoning must behave in the way we intend and allow us to reason correctly about those objects we wish to reason about, i.e., sets of sentences representing sets of assumptions and individual sentences representing individual assumptions. In other words, it is necessary that the theory be capable of being given a formal interpretation that accords with our intuitive interpretation. We insure this by requiring that an interpretation always be given for this theory and that the given interpretation meet certain conditions.

Let us now turn to our definition. We first define a concept to be employed in defining the notion of a two-level system. The idea of a structure for, or model of, the formulas of a formal language is well known (see for example, [B]). A structure for the language L consists of a domain of objects and a collection of mappings assigning suitable interpretations over the domain to the constant, predicate, and function symbols of L.

As usual, given a term t from L and a structure for L we think of each function s from the variables of L to a subset of the domain of discourse as assigning a meaning to the variables, each such function being called an assignment. Also as usual, we can define for each term t of L a function $\hat{t}$ which maps assignments to elements of the domain. For a given t, $\hat{t}$ is defined as follows (see [B]):

1. If t is a constant symbol c, then $\hat{t}(s) = \hat{c}$ for all s where $\hat{c}$ is the element of the domain which c is interpreted as;
2. If t is a variable v, then $\hat{t}(s) = s(v)$ for all s;
3. If t is the term $f(t_1,\ldots,t_n)$ then, for all s, define
$$\hat{t}(s) = \hat{f}(\hat{t}_1(s),\ldots,\hat{t}_n(s))$$
where $\hat{f}$ is the interpretation of f.

The following lemma is obvious from the definition of the mapping $\hat{t}$.

Lemma 4.1

If t is a closed term of L, then the mapping $\hat{t}$ has constant value for all assignments.

For any closed term t of L the element of the domain which is the value of the mapping $\hat{t}$ is the same no matter what the value of s. Let us also call this _element_ $\hat{t}$. We will then say that t _denotes_ the element $\hat{t}$.

A _two-level_ system consists of:

1. A set _S_ of sets of wffs in a language L called the _possible axiom sets_. One possible axiom set is distinguished as the _initial axiom set_. All other possible axiom sets must be supersets of the initial possible axiom set. The set of theories generated by the possible axiom sets is called the set of _object theories_.

2. A theory in a language L´ called the _metatheory_.

3. A structure for L´ called the _intended interpretation_ of L´.

We require that the domain of the intended interpretation be such that it includes the wffs of L and the possible axiom sets. We also require that L´ and the intended interpretation be such that:

1. For each wff α of L there is at least one closed term t of L´ such that t denotes α.

2. For each possible axiom set $S \in \underline{S}$ there is at least one closed term t of L' such that t denotes S.

3. There is a binary predicate symbol of L', say $\in$, which is interpreted as set membership.

4. There is a unary predicate symbol of L', say A_P, such that A_P is interpreted as the set $\underline{S}$. That is, $A_P(t)$ will hold for some assignment s just if $\hat{t}(s) \in \underline{S}$.

Finally, we require that the axioms of the metatheory be such that:

1. For each possible axiom set S there is a closed term t denoting S such that $A_P(t)$ is provable.

2. For each closed term t of L' if $A_P(t)$ is provable, t denotes a possible axiom set.

3. If $\alpha \in S$, then there are t_1, t_2 denoting α and S such that $t_1 \in t_2$ and $A_P(t_2)$ are provable.

4. If $t_1 \in t_2$ is provable for closed terms t_1, t_2, then t_2 denotes a set and t_1 denotes a member of that set.

The possible axiom sets are intended to represent the various sets of current beliefs that the reasoner might hold at some point during the reasoning process given that he begins with the initial axiom set. Thus, the object theories represent the sets of formulas that the reasoner could infer using conventional deductive inference. The requirements given for the metatheory's axioms are intended to insure that we can prove in the metatheory that the possible axiom sets are indeed the possible axiom sets using the predicate A_p (for "possible axiom set").

Intuitively, the introduction of a default assumption would be represented in a two-level system by proving that the set consisting of the possible axiom set representing the current assumptions along with the given default assumption is also a possible axiom set. A formula would be inferred from the current assumptions by constructing a proof of the formula using only current assumptions (this would be a proof in one of the object theories). It would also be necessary to prove in the metatheory that the current assumptions constituted a possible axiom set and that the members used in the proof of the formula belonged to that set. Our definition of a two-level system includes systems satisfying this intuitive picture as well as some which do not. We choose the form of definition given both because it is simple and for its generality.

4.3 The FOL System

In [We] Weyhrauch describes a computer reasoning system which he calls FOL. This system makes use of certain ideas that are similar to those underlying the definition of a two-level system. In particular FOL can be used to represent a finite number of first order object theories as well as a metatheory in which to reason about the object theories. Each theory is represented by a description of a language, an object called a simulation structure, and a set of axioms in the language. The simulation structure is essentially a fragment, which can be implemented on a machine, of an interpretation for the language. The axioms are to be true in the simulation structure.

FOL and the notion of a two-level system were developed independently. With respect to the problem of modelling default reasoning there are several differences. First, although the general idea of treating default reasoning as a metalevel process has been independently recognized by others including Weyhrauch, the concept has not been developed by him (or anyone else). Thus, although FOL presents a possible framework for treating default reasoning in the manner provided by the notion of a two-level system, it does not contain any explicit definitions corresponding to those to be found within the two-level system concept. (In fact, no one, to our knowledge, has gone so far as to work out an informal theory of default reasoning such as the one we presented above.)

A second point of difference concerns the handling of the metatheory. In FOL the metatheory may be used as its own metatheory. This arrangement appears to preclude the extension of FOL in any fashion similar to the extension of a two-level system to an n-level system as will be done in Chapter 6.

Finally, FOL is a computer system rather than a general mathematical model. It is therefore possible to represent certain systems as two-level systems which could not be represented in FOL even if definitions similar to those that make up the notion of a two-level system were developed for FOL.

5. The Relation of Default Theories to Two-Level Systems

In this chapter we show certain relationships between default theories and two-level systems. On the basis of these we argue that the nonmonotonicity of default theories as well as their lack of a notion of "default inference rule" are the result of certain hypotheses about default reasoning that need not be accepted. We also argue that a two-level system is at least as suitable as a default theory for a model of default reasoning.

5.1 Arbitrary Closed Default Theories

We begin by showing that the notion of a two-level system subsumes that of a default theory in the sense that for any closed default theory there is an equivalent two-level system.

Suppose that (D,W) is an arbitrary closed default theory in the language L. Let E be an extension of (D,W) and let

$$D(E) = \{\beta \mid \beta \in E \text{ and } \alpha:M\alpha_1,\ldots,M\alpha_k/\beta \in D \text{ for some } \alpha,\alpha_1,\ldots,\alpha_k\}.$$

It is shown in [R] that $E = Th(W \sqcup D(E))$. We will use this fact to define our two-level system. First we define a metalanguage L'. Let L' consist of:

1. A constant symbol, say α' for each wff α in L;
2. A constant symbol, say S', for each set S of wffs of L;
3. One unary predicate symbol A_p and one binary predicate symbol $\in$.

Let AP be the unary relation over the sets of wffs of L defined by:

1. $W \sqcup D(E) \in AP$ for each extension E of (D,W);
2. Nothing else is in AP.

We can now define a structure for L' which will serve as the intended interpretation of our two-level system's metatheory. The domain of the structure consists of:

1. The wffs of L;
2. The sets of wffs of L.

Each constant α' of L' is interpreted as the corresponding wff α of L. Each constant S' is interpreted as the corresponding set S. The predicate symbol $\in$ is interpreted as set membership while $A_p(x)$ is interpreted to mean that x is in AP.

The possible axiom sets of the two-level system are the members of AP. The axioms of the metatheory are:

1. $\alpha' \in S'$ for each α' and each S' such that $\alpha \in S$;
2. $A_p(S')$ for each S' such that $S \in AP$.

We must show that these definitions of a metatheory, an interpretation for the metatheory, and a set of possible axiom sets satisfy the requirements for a two-level system. The domain of the given structure certainly includes the wffs of the object language and the possible axiom sets. We must also show that $A_p(t)$ is provable in the metatheory if and only if t denotes a possible axiom set. Finally, if β is provable from a possible axiom set S, then for each member α of S occurring in the proof, we must show that $\alpha \in t$ is provable in the metatheory for closed t such that t denotes S. Let us call the system just defined Σ.

Lemma 5.1

If t is a closed term in L' and $A_p(t)$ is provable in the metatheory of Σ, then t denotes a member of AP.

Proof:

Obviously, the axioms of Σ's metatheory are satisfied by Σ's structure. Therefore, if $A_p(t)$ is provable, it must also be satisfied by the structure. Since A_p is interpreted as AP, this can only be the case if t denotes a member of AP.[]

Lemma 5.2

If $S \in AP$ then there is a closed term t of L' such that t denotes S and $A_p(S)$ is provable in Σ's metatheory.

Proof:

For every set S in AP the constant symbol S' denotes S and $A_p(S')$ is an axiom.[]

Lemma 5.3

Suppose t is a closed term of L'. Then $\alpha' \in t$ is provable iff t denotes a set and α is a member of the set.

Proof:

If $\alpha' \in t$ is provable, then $\alpha' \in t$ must be satisfied by the structure so t must denote a set of which α is a member.

Since there are no function symbols, the only terms denoting sets are constant symbols. If $t = S'$ and $\alpha \in S$, then $\alpha' \in S'$ is an axiom.[]

Thus, Σ is indeed a two-level system. From the definition of the intended interpretation we see that the set of sentences provable from a possible axiom set is an extension of (D,W) and that every extension can be generated from some possible axiom set. Thus, in an obvious sense, Σ is equivalent to (D,W).

Although we have shown that the extensions of a closed default theory can be accounted for in terms of the object theories of a two-level system, we did so by introducing a two-level system with a metatheory which is not especially illuminating. It allows us to conclude that possible axiom sets are indeed possible axiom sets but only because an axiom asserting A_p for each such set is included. This arrangement bears little resemblance to the intuitive view of default reasoning expressed in the previous chapter. There we took default reasoning to be a process of adding a default assumption to a set of axioms representing the reasoner´s current beliefs and pictured the proposed metatheory as containing rules by which the reasoner would determine whether a given default assumption could be added to a given set of current assumptions.

In fact, the problem with our first two-level system seems to be the default theory which it represents. In Chapter 2 it was noted that there appears to be no notion of "default inference rule" that can be associated with the definition of a default theory. One cannot speak of inferring a default assumption in a given default theory but only of its membership in some extension. In effect, although the definition of an extension supplies a description of the result of having introduced a certain set of default assumptions, for the general case there is no notion of a process that does the introducing. Thus, we have a notion (extension of a default theory) corresponding to the result of default reasoning but no notion of default reasoning itself.

The structure of the metatheory in the above two-level system essentially mirrors the default theory's lack of rules. Given our intuitive view of default reasoning, we would expect the metatheory to allow us to represent a procedure of sequentially adding one default assumption at a time to build a series of sets of "current" object-level assumptions. Instead, we simply have an axiom asserting the predicate A_P for each of the axiom sets that generate a Reiter extension. Thus, we have the same problem as with the corresponding default theory. The possible axiom sets represent the result of default reasoning but the proofs of the metatheory do not represent the process of default reasoning. We will see in the next section that the problem stems from the hypotheses underlying the definition of default theories.

5.2 Closed Normal Default Theories

A normal default theory is one in which all defaults are of the form $\alpha:M\beta/\beta$. That is, the only wff that must be consistent with "what is known" is the one to be assumed. In the previous chapter we discussed the difference between supposing that the requirement in a default reasoning argument of not being contrary to what the reasoner knows always refers only to the default assumption, and supposing that such a requirement could refer to other assertions as well. We argued that the first of these possibilities is more intuitive. In terms of the notion of a default theory, our argument would call for allowing only normal default theories. In this section we show that Reiter´s "normal form" is not only more intuitive but also allows a well-defined notion of rules for introducing default assumptions.

Let (D,W) be an arbitrary closed normal default theory in the language L. The following results about (D,W) will show that in the case of a closed normal default theory defaults can be interpreted as inference rules in a natural way. They will also be used in the next section as a basis for defining a natural form of two-level system that is equivalent to a closed normal default theory.

Let E be a fixed set of closed wffs in the language L of the closed normal default theory (D,W). Consider the defaults of D to be given in some order. Suppose that α_j, β_j are the wffs occurring in the jth default.

Let

$$F_0 = W$$

$$F_{i+1} = \bigsqcup \{F_{i+1,j} \mid 0 \leq j \leq i+1\}$$

where

$$F_{i+1,0} = F_i$$

$$\begin{aligned} F_{i+1,j+1} &= F_{i+1,j} \sqcup \{\beta_j\} \text{ if } F_{i+1,j} \vdash \alpha_j, \\ &\qquad \text{and } {\sim}\beta_j \notin E \\ &= F_{i+1,j} \text{ otherwise.} \end{aligned}$$

Let $E^* = \bigsqcup F_i$ for $i = 0$ to ∞.

F in effect represents the result of building a (possibly infinite) set of assumptions starting with the initial assumptions and adding default assumptions one at a time. By showing that F generates E just if E is an extension we can show that the set of default assumptions contained in an extension may be built up in the natural way.

Recall that Reiter's notion of extension is defined in terms of the operator Γ. In particular, the set of wffs E is an extension of the default theory (D,W) just if $E = \Gamma(E)$. For our discussion of default theories we will use Th(S) to mean the set of sentences provable from S.

Lemma 5.4

Th(E*) = Γ(E). Hence, E is an extension for (D,W) iff E = Th(E*).

Proof:

Since E is an extension iff E = Γ(E), it suffices to show that Th(E*) = Γ(E). First we will show that Th(E*) satisfies the conditions which must be true of Γ(E):

Condition 1

W $\subseteq$ Th(E*) by the definition of E*.

Condition 2

Obviously, Th(E*) = Th(Th(E*)).

Condition 3

Suppose for some member of D, say α_k:Mβ_k/β_k, $\alpha_k \in$ Th(E*) and $\sim\beta_k \notin$ E. Since $\alpha_k \in$ Th(E*), there is a least i, say i$'$, such that $F_i \vdash \alpha_k$. Suppose k $\leq$ i$'$. Since $F_{i'} \vdash \alpha_k$, $F_{i'+1,k} \vdash \alpha_k$ because $F_{i'} \subseteq F_{i'+1,k}$. Also, $\sim\beta_k \notin$ E so by the definition of E*, $\beta_k \in$ F. Now suppose k > i$'$. Since $F_{i'} \vdash \alpha_k$ and $F_{i'} \subseteq F_i$ for all i > i$'$, $F_i \vdash \alpha_k$ for all i > i$'$. Therefore, $F_k \vdash \alpha_k$ and so $F_{k+1,k} \vdash \alpha_k$. Also, $\sim\beta_k \notin$ E so again $\beta_k \in$ E*.

Thus, $Th(E^*)$ satisfies the three conditions.

We can show $Th(E^*) \subseteq \Gamma(E)$ by showing by induction that $E^* \subseteq \Gamma(E)$. Obviously, $F_0 = W \subseteq \Gamma(E)$. Suppose $F_i \subseteq \Gamma(E)$ and consider $\beta \in F_{i+1}$. Either $\beta \in F_i$, in which case $\beta \in \Gamma(E)$ by assumption, or $\beta = \beta_j$ where β_j is such that $F_{i+1,j+1} = F_{i+1,j} \sqcup \{\beta_j\}$, $F_{i+1,j} \vdash \alpha_j$, and $\sim\beta_j \notin E$. Suppose β is such a β_j. Since $F_i \subseteq \Gamma(E)$ by assumption, $F_{i+1,0} = F_i \subseteq \Gamma(E)$. Suppose $F_{i+1,j} \subseteq \Gamma(E)$. Then since $F_{i+1,j} \vdash \alpha_j$, $\alpha_j \in \Gamma(E)$ because $\Gamma(E) = Th(\Gamma(E))$. Also, $\sim\beta_j \notin E$ so by the definition of $\Gamma(E)$, $\beta_j \in \Gamma(E)$. It follows that for $j = 0$ to $i+1$, $F_{i+1,j} \subseteq \Gamma(E)$ and therefore, $F_{i+1} \subseteq \Gamma(E)$. Thus, $E^* \subseteq \Gamma(E)$ and so $Th(E^*) \subseteq \Gamma(E)$.

Thus, we have that $Th(E^*)$ satisfies the three conditions on $\Gamma(E)$ and that $Th(E^*) \subseteq \Gamma(E)$. Therefore, $Th(E^*) = \Gamma(E)$.[]

The point of Lemma 5.4 is that an extension of a closed normal default theory could be constructed by starting with W and adding one default assumption at a time - a view of an extension that accords much better with the natural picture of the default reasoning process than does the definition of an extension as a fixed point. Because of the way in which the set E* is defined the particular ordering chosen for D has no importance.

The purpose served by the ordering is to allow us, at say the ith stage in the construction of E*, to try over again all those defaults which have been tried in the previous i - 1 stages. The reason for this is that a default which did not apply before may apply at the ith stage. (Since we want to add default assumptions one at a time, we have to try defaults one at a time.) For example, if $W = \{\alpha\}$ and $D = \{\alpha:M\beta/\beta, \beta:M\gamma/\gamma\}$, then the second of D´s two defaults could be applied but only after the first had been applied. Since we have no way to tell when a default will become applicable, we keep trying them over again.

Consider the class of sets consisting of W and all sets of the form $W \sqcup \{\alpha_1,\ldots,\alpha_k\}$. We define a unary relation AP over this class as follows:

1. $W \in AP$.
2. For any set A in AP and any default $\alpha:M\beta/\beta \in D$, if $A \vdash \alpha$, and $A \nvdash {}^{\sim}\beta$, then $(A \sqcup \{\beta\}) \in AP$.
3. No other members of the domain are members of AP.

Lemma 5.5

Suppose E is an extension of (D,W). Then for the sequence of sets $\{F_i\}$ defined above, $F_i \in AP$ for $i = 0$ to ∞.

Proof:

Clearly, $F_0 \in$ AP. Suppose $F_k \in$ AP and consider F_{k+1}. From the definition we can see that $F_{k+1} = F_{k+1,k+1}$. Since $F_{k+1,0} = F_k$, $F_{k+1,0} \in$ AP. Suppose $F_{k+1,i} \in$ AP for $i < k+1$ and consider $F_{k+1,i+1}$. If $F_{k+1,i+1} \neq F_{k+1,i}$, then $F_{k+1,i+1} = F_{k+1,i} \sqcup \{\beta_j\}$ where $\sim\beta_j \notin E$ and $F_{k+1,i} \vdash \alpha_j$. Since $\sim\beta_j \notin E$, B_j is consistent with $F_{k+1,i}$ by the previous lemma. Therefore, F_{k+1}, α_j, and β_j satisfy the conditions of the definition of AP and $F_{k+1,i+1} \in$ AP. It follows that $F_{k+1,i} \in$ AP for $i = 0$ to $k+1$. Thus, $F_k \in$ AP and by induction $F_i \in$ AP for all i.[]

Theorem 5.1

Let E be an extension of (D,W). Then there is a sequence of sets in AP, say $A_0, A_1, \ldots$, such that for each i, $A_i \subseteq A_{i+1}$, and if $A = \bigsqcup A_i$ for $i = 0$ to ∞, then $E = Th(A)$.

Proof:

By Lemma 5.4 $E = Th(E^*)$ where $E^* = \bigsqcup F_i$ for $i = 0$ to ∞ and by the definition of the sequence $\{F_i\}$ and Lemma 5.5, the F_i's satisfy the other conditions of the theorem.[]

Actually, the above results would carry through for any closed default theory. However, the following results do not hold for other than the normal case.

For any closed normal default theory (D,W) and any ordering of the defaults of D let

$$E_0 = W$$
$$E_{i+1} = \bigsqcup \{E_{i+1,j} \mid 0 \leq j \leq i+1\}$$

where

$$E_{i+1,0} = E_i$$
$$\begin{aligned} E_{i+1,j+1} &= E_{i+1,j} \sqcup \{\beta_j\} \text{ if } E_{i+1,j} \vdash \alpha_j, \\ &\qquad \text{and } E_{i+1,j} \nvdash {}^{\sim}\beta_j \\ &= E_{i+1,j} \text{ otherwise.} \end{aligned}$$

Here we again assume that α_j and β_j are the wffs occurring in the jth default for the given ordering. Let E = Th(E´) where E´ = $\bigsqcup E_i$, i = 0 to ∞ and let E* be defined as above in terms of this set E. The construction of E is based on interpreting a default as a rule. We do this in the way that is natural according to our informal theory of default reasoning. Given a default, say α:Mβ/β, we add β to the set constructed so far if α is provable from that set and β is consistent with that set. We do not concern ourselves with whether β is going to be consistent with the entire set that will eventually be constructed. The interesting fact about the normal form is that treating defaults in this manner leads to the same sets as are generated by Reiter´s fixed point definition as will be seen below.

Lemma 5.6

E is an extension for (D,W).

Proof:

If W is inconsistent, E = L the entire language. It is easy to see from the definition of the set E* that if W is inconsistent, it has the unique extension L. Thus, E is an extension in this case.

Suppose W is consistent. Then by construction E' is consistent and so is E. We already know that E is an extension iff E = Th(E*). Therefore, we will show that $E' = E^*$.

Obviously, $E_0 = F_0$. Suppose $E_i = F_i$. Then $E_{i+1,0} = F_{i+1,0}$ so assume that $E_{i+1,j} = F_{i+1,j}$ and that β_j is such that $E_{i+1,j} \vdash \alpha_j$ and $\beta_j \notin E_{i+1,j}$ (and hence, $F_{i+1,j} \vdash \alpha_j$ and $\beta_j \notin F_{i+1,j}$). If $E_{i+1,j} \nvdash {\sim}\beta_j$, then ${\sim}\beta_j \notin E_i$ or $E_{i+1,k}$, k = 0 to j and furthermore $\beta_j \in E_{i+1}$ and hence also E_k for all k > i+1. Therefore, since E is consistent, ${\sim}\beta_j \notin E$. Thus, $F_{i+1,j+1} = F_{i+1,j} \sqcup \{\beta_j\} = E_{i+1,j} \sqcup \{\beta_j\} = E_{i+1,j+1}$ in this case. If ${\sim}\beta_j \notin E$, (so that $F_{i+1,j+1} = F_{i+1,j} \sqcup \{\beta_j\}$), then we also have $E_{i+1,j} \nvdash {\sim}\beta_j$ and so $E_{i+1,j+1} = E_{i+1,j} \sqcup \{\beta_j\}$. Thus, $E_{i+1,j} = F_{i+1,j}$, j = 0 to i+1 and it follows that $E_{i+1} = F_{i+1}$. Therefore, by induction, $E' = E^*$.[]

In the definition of the sets E_i given before Lemma 5.6 an ordering of the defaults of D is assumed as was done in constructing the set E* for Lemma 5.4. Note that the lemma does not tell us that we could construct an extension by simply applying each default as a rule in the order given by the ordering. That is, if we apply the first rule to W, then apply the second rule to the set of wffs which results from applying the first, and so on, the resulting set of wffs need not be a complete extension. Lemma 5.6 does let us show that if we take the defaults in any order and attempt to apply each default once, the resulting set will be a subset of some extension.

Also, the definition of the sets E_i induces another ordering, possibly distinct from the assumed ordering. In this ordering, defaults actually applied in constructing some E_i come before those which never apply, and applied defaults are ordered according to the order of their application. Using this second ordering, we could construct an extension simply by applying the defaults in the order given by the ordering. Similarly, the definition of the sets F_i given before Lemma 5.4 induces an ordering of the defaults for any given extension E such that E can be constructed by applying the defaults according to their position in the ordering.

Theorem 5.2

Suppose $A \in AP$. Then there is an extension E of (D,W) such that $Th(A) \subseteq E$.

Proof:

Since $A \in AP$, either $A = W$ or $A = A_0 \sqcup \ldots \sqcup A_k$ where $A_0 = W$ and for each i $A_{i+1} = A_i \sqcup \{\beta_i\}$ for some β_i such that for some α_i $\alpha_i : M\beta_i / \beta_i \in D$ and A_i, α_i, and β_i satisfy the conditions of the definition of AP. Order the defaults of D so that the first k defaults are those used to form $A_1, \ldots, A_k$. We will show that $A \subseteq E'$ where E' is the set defined prior to Lemma 5.6.

Clearly $A_0 = E_0$. Suppose that $A_i = E_i$ for $i < k$ and consider A_{i+1} and E_{i+1}. $E_{i+1} = \sqcup E_{i+1,j}$, j = 0 to i+1. Also, $E_{i+1,0} = E_i = A_i$. $E_{i+1,j+1} = E_{i+1,j} \sqcup \{\beta_j\}$ if $E_{i+1,j} \vdash \alpha_j$ and $E_{i+1,j} \nvdash {\sim}\beta_j$. But for $j < i$ $\beta_j \in A_i$. Therefore, for j = 0 to i $E_{i+1,j} = A_i$. Hence, for j = i $E_{i+1,j} \vdash \alpha_j$ and $E_{i+1,j} \nvdash {\sim}\beta_j$. Therefore, $E_{i+1,i+1} = A_i \sqcup \{\beta_i\} = A_{i+1}$, and it follows that $E_{i+1} = A_{i+1}$ also. Hence, $A \subseteq E$ which is an extension by Lemma 5.5.[]

Theorem 5.3

A sentence β is a member of some extension E of (D,W) iff $\beta \in Th(A)$ for some $A \in AP$.

Proof:

If $\beta \in Th(A)$, then $\beta \in E$ for some extension E by the previous theorem.

If $\beta \in E$ for some extension E, then $\beta \in Th(E^*)$ by Lemma 5.4 (where E* is as in Lemma 5.4). Since the proof of β from E* must be finite, there is an i such that $F_i \vdash \beta$. By Lemma 5.5 $F_i \in AP$.[]

The above results show that for the normal case we can give a characterization of an extension which is different from that given by Reiter who defines an extension to be a fixed point of the operator Γ. The set of theorems of an ordinary formal theory is in a sense a fixed point too, but we can also think of this set as being produced incrementally by the inference process from the axioms of the theory. Analogously, we would like to think of an extension of a default theory (D,W) as being produced from W by a process involving ordinary inference rules and the defaults of D treated as rules for introducing default assumptions. As we have seen in Chapter 2, this does not appear to be possible in general. In the case of a closed normal default theory the situation is different. The above results show that it is possible to generate an extension incrementally in a way similar to that in which one could generate the theorems of a formal theory by constructing proofs. In fact, in the normal case we can supply an interpretation of a default as a rule for introducing default assumptions and a notion of "proof".

In Chapter 2 we discussed the natural interpretation of a default as a kind of inference rule, finding that it would not work in general. Let us consider this interpretation as it would apply to a normal default, say $\alpha:M\beta/\beta$. Recall that the idea was to interpret the consistency requirement of the default with respect to the set of assumptions, initial plus default, that were accepted by the reasoner at the time of the default's application. The above results show that we can interpret a normal default in this way. For example, if α is provable (by the rules of the predicate calculus) from W and β is consistent with W (again, in terms of the inference rules of the predicate calculus), then β is a member of some extension of (D,W). Furthermore, one can begin with the initial assumptions of a closed normal default theory and derive any member of an extension by a sequence of applications of defaults and conventional inference rules. In the process a sequence of assertions is produced which can naturally be defined as a proof of that member of an extension.

However, since the consistency requirement of a default is interpreted in terms of the inference rules of the predicate calculus, we are not treating defaults in the same way as the other rules of the system. Thus, this approach would require altering Reiter´s initial postulate that default assumptions are themselves consequences of a logic. Implicit in that view is a further assumption by Reiter that consequences of the logic, whether default assumptions or assertions derived by conventional deductive inference, will have the same status. Our interpretation of a default causes default assumptions to be distinct from assertions derived by conventional inference. In determining whether a default assumption is consistent with the set of current assumptions we do not consider other default assumptions which might be "derived" from that set, only those assertions that can be derived by conventional inference. Thus, Reiter´s assumption is in effect altered by our interpretation. The fact that we can make this alteration in the case of a closed normal theory without changing the contents of the sets called extensions is one reason for arguing that Reiter´s original stronger assumption is unnecessary.

Reiter comes near to discovering the possibility of interpreting normal defaults as rules when he shows that for closed normal default theories any member of an extension has what he calls a default proof.

Given a normal default rule $\alpha{:}M\beta/\beta$, call α the prerequisite of the rule and β the consequent. For any set of normal default rules D let P(D) be the set of prerequisites occurring in D and let C(D) be the set of consequents. A default proof of γ from a closed normal default theory (D,W) is a finite sequence of finite subsets of D, say $\{D_i\}$ for i = 1 to k, such that

1. For each $\alpha \in P(D_1)$, $W \vdash \alpha$;
2. For each i, i = 1 to k-1 and each $\alpha \in P(D_{i+1})$,
   $W \sqcup C(D_i) \vdash \alpha$;
3. $W \sqcup C(D_k) \vdash \gamma$;
4. $W \sqcup C$ is consistent where $C = \sqcup C(D_i)$, i = 1 to k.

Reiter shows that γ is a member of some extension for (D,W) if and only if γ has a default proof.

While it is easy to see that Reiter´s result follows from our result, the notion of a default proof does not seem as appealing as the notion of "proof" (i.e., derivation of a member of an extension) presented above. A default proof still affords no interpretation of a rule for introducing a default assumption, nor does it fit as well with one´s intuitive idea of a proof as a sequence of assertions, each one derived from those preceding it.

5.3 A Natural Two-Level System for a Closed Normal Default Theory

We noted in the previous section that our interpretation of normal defaults as rules relied on treating them as separate from the inference rules referred to in the interpretation. The obvious role for rules of the type that defaults become in this view is as metalevel rules. Defaults could then be examples of the sort of rules envisioned in our informal theory of default reasoning. In fact, we can define a two-level system that is equivalent to a given closed normal default theory and is such that the meta-axioms for A_p correspond directly to the defaults. We now proceed to do this.

Suppose (D,W) is a closed normal default theory in the language L. We first define a metalanguage, L´, consisting of:

1. A constant symbol for each wff, α, of L, say α';
2. A constant symbol for W, say W´;
3. One binary function symbol, say ad;
4. The usual connectives and quantifiers and an infinite supply of variables;
5. The binary predicate symbols $\in$ and Pr and the unary predicate symbols S, and A_p.

Next, we define a structure for L'. The domain of discourse consists of $A \sqcup B \sqcup \{W\}$ where A is the set of wffs of L and B is the set of sets of the form $W \sqcup \{\alpha_1,\ldots,\alpha_k\}$. Thus, the domain consists of the wffs of L, the set W, and all sets consisting of the union of W and some finite set of wffs of L. Symbols of the form α' from L' are interpreted as the corresponding wff α of L. The symbol W' is interpreted as the set W. We interpret $\in$ as the standard membership relation. S(x) is interpreted to mean that x is a set while Pr(x,y) is interpreted to mean that the wff y is provable from the set of wffs x. A_P is interpreted as the unary relation AP defined in the previous section.

The function adj is defined as follows:

If $x = W$ or $x = W \sqcup \{\alpha_1,\ldots,\alpha_k\}$ and y is a wff of L,
then let $\text{adj}(x,y) = x \sqcup \{y\}$
else let $\text{adj}(x,y) = d$.

Here, d is some fixed wff of L. Thus, if x is one of the sets in the domain and y is one of the wffs, then adj(x,y) is the union of x and $\{y\}$. Otherwise adj(x,y) is a wff. The function symbol ad is interpreted as the function adj. This completes our interpretation of L'.

Our metatheory must allow us to deal to a certain extent with sets of wffs of L. We must be able to handle taking the union of a set of wffs and a singleton and we must be able to show that members of these sets are indeed members. However, we need not get bogged down in the machinery of set theory since we do not need anything so powerful. We thus introduce the function adj and its corresponding symbol ad as well as the symbol S and its interpretation. The axioms for ad and ∈ given below allow us the necessary ability to manipulate sets. The axioms for S allow us to distinguish terms denoting sets from those denoting wffs.

We also wish to keep our metatheory first order. For many of the axioms below it would be most natural to quantify over sets of wffs but this would result in a second order theory. We therefore use countably infinite sets of axioms in these cases, one axiom for each finite set of wffs. Finally, we introduce axioms asserting both provability and unprovability statements for wffs in L. In the case of provability our purpose is to keep the metatheory simple. In the case of unprovability we of course have no choice.

We can now state the axioms of the metatheory to be employed in the two-level system we wish to define. They are as follows:

In the following we write $ad(W',\alpha'_1,\ldots,\alpha'_k)$ for $ad(\ldots ad(W',\alpha'_1),\ldots,\alpha'_k)$.

1. $\alpha' \in W'$ for each $\alpha \in W$.

2. $\sim(\alpha' \in W')$ for each $\alpha \notin W$.

3. $S(W')$.

4. $\sim S(\alpha')$ for all constants of L' other than W'.

5. $\forall x \forall y(S(x)\ \&\ \sim S(y) \Leftrightarrow S(ad(x,y)))$.

6. $\forall x \forall y \forall z(x \in ad(y,z) \Leftrightarrow (S(y)\ \&\ \sim S(z))$
   $\&\ (x \in y \vee x = z))$.

7_0. $Pr(W',\alpha')$ for each α such that $W \vdash \alpha$.

$\vdots$

7_n. $Pr(ad(W',\alpha'_1,\ldots,\alpha'_n),\beta')$ for each $\alpha'_1,\ldots,\alpha'_n$, β' such that $W \sqcup \{\alpha_1,\ldots,\alpha_n\} \vdash \beta$.

$\vdots$

8_0. $\sim Pr(W',\alpha')$ for each α such that $W \nvdash \alpha$.

$\vdots$

8_n. $\sim Pr(ad(W',\alpha'_1,\ldots,\alpha'_n),\beta')$ for each $\alpha'_1,\ldots,\alpha'_n$, β' such that $W \sqcup \{\alpha_1,\ldots,\alpha_n\} \nvdash \beta$.

$\vdots$

9. $A_p(W')$.

9_0. $Pr(W',\alpha')\ \&\ \sim Pr(W',\sim\beta') \rightarrow A_p(ad(W',\beta'))$ for each $\alpha{:}M\beta/\beta \in D$.

$\vdots$

9_n. $\forall x_1\ldots\forall x_n(A_p(ad(W',x_1,\ldots,x_n))\ \&$

$$Pr(ad(W',x_1,\ldots,x_n),\alpha') \ \&$$
$$\sim Pr(ad(W',x_1,\ldots,x_n),\sim\beta) \rightarrow$$
$$A_P(ad(W',x_1,\ldots,x_n,\beta))) \text{ for each } \alpha:M\beta/\beta \in D.$$

$\vdots$

As we will see below, these axioms represent true statements concerning the function and predicate symbols of L′ as we have interpreted them.

To define the desired two-level system we take the above axioms as the axioms of the metatheory. The possible axiom sets are just the members of AP. We take the structure defined above as the intended interpretation of the metatheory.

The metatheory we have defined would not be recursively axiomatizable in general and we will discuss this point below. However, the set of axioms we have defined for the metatheory is countable. Let us call the system just defined Σ. We must now show that Σ meets the requirements for a two-level system.

Lemma 5.7

Let t be a term of L′ of the form

$$ad(W',\alpha'_1,\ldots,\alpha'_k).$$

Then t denotes $W \sqcup \{\alpha_1,\ldots,\alpha_k\}$.

Proof:

We use induction on k. If k = 1, then t = $ad(W',\alpha'_1)$ and by our definition of "denotes" t denotes $adj(W,\alpha_1) = W \sqcup \{\alpha_1\}$.

Suppose the lemma is true for k = n and consider k = n+1. By assumption $ad(W',\alpha'_1,\ldots,\alpha'_{k-1})$ denotes $W \sqcup \{\alpha_1,\ldots\alpha_{k-1}\}$ so again by the definition of "denotes" $ad(W',\alpha'_1,\ldots,\alpha'_k)$ denotes $W \sqcup \{\alpha_1,\ldots,\alpha_k\}$.[]

Lemma 5.8

If S = W or $S = W \sqcup \{\alpha_1,\ldots,\alpha_k\}$, then there is a closed term t of L' such that t denotes S.

Proof:

W is denoted by W'. By Lemma 5.7 $W \sqcup \{\alpha_1,\ldots,\alpha_k\}$ is denoted by $ad(W',\alpha'_1,\ldots,\alpha'_k)$.[]

Lemma 5.9

Let t be a closed term in which ad occurs such that t is not of the form $ad(W',\alpha'_1,\ldots,\alpha'_k)$. Then t denotes d, the arbitrary wff specified in the definition of adj.

Proof:

Since the only function symbol is ad, t must be of the form $ad(t_1,t_2)$.

Suppose t_1 and t_2 are constant symbols. Then either $t_1 \neq W'$ or $t_2 = W'$ (otherwise t is of the wrong form). In either case, t denotes d.

Suppose the claim is true for terms containing k occurrences of ad and consider t containing k+1 occurrences of ad. If t_1 denotes a set and t_2 denotes a wff, then t is of the wrong form. Thus, either t_1 does not denote a set or t_2 does not denote a wff so again, t denotes d.[]

Lemma 5.10

A closed term t of L' denotes a set iff $t = W'$ or t is of the form $ad(W',\alpha'_1,\ldots,\alpha'_k)$.

Proof:

Since the only function symbol is ad, t must either be a constant symbol or a term of the form $ad(t_1,t_2)$. By the interpretation of the constant symbols, only W' denotes a set. By Lemma 5.9 if ad occurs in t, then t denotes a set iff t is of the form $ad(W',\alpha'_1,\ldots,\alpha'_k)$.[]

Lemma 5.11

The axioms of Σ's metatheory are satisfied by the given structure.

Proof:

Axioms of the form $\alpha' \in W'$ and $\alpha' \notin W'$ are obviously satisfied by the structure as are $S(W')$ and axioms of the form $\tilde{}S(\alpha)$.

If x is a set in the domain of discourse, then x is either W or the union of W and a finite set of wffs of L. If y is not a set, then y is a wff of L. Thus, by the definition of adj, adj(x,y) is a set. Conversely, if adj(x,y) is a set, then x must be a set and y a wff. Hence, the axiom

$$\forall x\forall y(S(x) \mathbin{\&} {\sim}S(y) \Leftrightarrow S(ad(x,y)))$$

is satisfied by the structure and similarly

$$\forall x\forall y\forall z(x \in ad(y,z) \Leftrightarrow (S(y) \mathbin{\&} {\sim}S(z)) \mathbin{\&} (x \in y \vee x = z))$$

is also satisfied.

Axioms of the form $Pr(W',\alpha')$ and ${\sim}Pr(W',\alpha')$ are obviously satisfied by the structure. By Lemma 5.7 any term of the form $ad(W',\alpha'_1,\ldots,\alpha'_k)$ denotes the set $W \sqcup \{\alpha_1,\ldots,\alpha_k\}$ so it is also clear that axioms of the form $Pr(ad(W',\alpha'_1,\ldots,\alpha'_k),\beta')$ and ${\sim}Pr(ad(W',\alpha'_1,\ldots,\alpha'_k),\beta')$ are satisfied.

The axiom $A_P(W')$ is also obviously satisfied. If for any $\alpha{:}M\beta/\beta \in D$ $W \vdash \alpha$ and $W \nvdash {\sim}\beta$, then $adj(W,\beta) \in AP$. Therefore, axioms of the form

$$Pr(W',\alpha') \mathbin{\&} {\sim}Pr(W',\beta') \rightarrow A_P(ad(W',\beta')$$

are satisfied by the structure. Similarly, axioms of the form

$$\begin{aligned}\forall x_1\ldots\forall x_n(A_P(ad(W',x_1,\ldots,x_n)) \mathbin{\&}& \\ &Pr(ad(W',x_1,\ldots,x_n),\alpha') \mathbin{\&} \\ &{\sim}Pr(ad(W',x_1,\ldots,x_k),\beta') \rightarrow\end{aligned}$$

$$A_p(ad(W',x_1,\ldots,x_k,\beta')))$$

are satisfied. To see this we note that by Lemma 5.10 $ad(W',a_1,\ldots,a_n)$ denotes a set iff $a_1,\ldots,a_n$ are constants, say $\alpha'_1,\ldots,\alpha'_n$, denoting wffs of L. But if $A_p(ad(W',\alpha'_1,\ldots,\alpha'_n))$, $Pr(ad(W',\alpha'_1,\ldots,\alpha'_n),\alpha')$, and $\sim Pr(ad(W',\alpha'_1,\ldots,\alpha'_n),\beta')$ are satisfied by the structure, then $adj(W \sqcup \{\alpha_1,\ldots,\alpha_n\},\beta)$ belongs to AP and $A_p(ad(W',\alpha'_1,\ldots,\alpha'_n,\beta'))$ is also satisfied.[]

Theorem 5.4

If $S \in AP$ then there is a closed term t of L' denoting S such that $A_p(t)$ is provable in Σ's metatheory.

Proof:

For $S = W$ we have $A_p(W')$ as an axiom.

For S a union of W and a finite set R of wffs it is obvious that the members of R can be ordered, say as $\beta_1,\ldots,\beta_k$, such that $W \sqcup \{\beta_1\} \in AP$, $W \sqcup \{\beta_1,\beta_2\} \in AP,\ldots,$ $W \sqcup \{\beta_1,\ldots,\beta_k\} \in AP$. For $S \neq W$ we will show that if $S = W \sqcup \{\beta_1,\ldots,\beta_k\}$ where $\beta_1,\ldots,\beta_k$ are ordered in the way just described, then $A_p(ad(W',\beta'_1,\ldots,\beta'_k))$ is provable. By Lemma 5.7 this will satisfy the theorem's claim.

Suppose $S = W \sqcup \{\beta_1\}$. Then by the definition of AP there is $\alpha_1{:}M\beta_1/\beta_1 \in D$ where $W \vdash \alpha_1$ and W is consistent with β_1. Thus, there is an instance of axiom schema 9_0 in which α'_1 and β'_1 occur. Furthermore, $Pr(W',\alpha'_1)$ and

$\sim Pr(W', \sim\beta'_1)$ are instances of axiom schemas 7_0 and 8_0 respectively. Thus, $A_p(ad(W',\beta'_1))$ is provable.

Suppose that for $S = W \sqcup \{\beta_1,\ldots\beta_k\}$ with $\beta_1,\ldots,\beta_k$ ordered as above $A_p(ad(W',\beta'_1,\ldots,\beta'_k)$ is provable. Consider $S = W \sqcup \{\beta_1,\ldots\beta_{k+1}\}$ where again we assume the β_j's are ordered as above. Then there must be $\alpha_{k+1}{:}M\beta_{k+1}/\beta_{k+1} \in D$ such that $W \sqcup \{\beta_1,\ldots,\beta_k\} \vdash \alpha_{k+1}$ and is consistent with β_{k+1}. Therefore, there is an instance of axiom schema 9_k in which α'_{k+1} and β'_{k+1} occur. Furthermore, there are instances of axiom schemas 7_k and 8_k of the form $Pr(ad(W',\beta'_1,\ldots,\beta'_k),\alpha'_{k+1})$ and $\sim Pr(ad(W',\beta'_1,\ldots,\beta'_k),\sim\beta'_{k+1})$. By hypothesis we have $A_p(ad(W',\beta'_1,\ldots,\beta'_k))$ so $A_p(ad((W',\beta'_1,\ldots,\beta'_{k+1}))$ is also provable.[]

Theorem 5.5

If $A_p(t)$ is provable in Σ's metatheory for a closed term t, then t denotes a member of AP.

Proof:

By Lemma 5.11 the axioms of the metatheory are satisfied by the structure defined for Σ. Therefore, we can make the same argument as for Lemma 5.1.[]

Theorem 5.6

For any closed term t of L', $\alpha' \in t$ is provable in

Σ's metatheory iff t denotes a set and α is a member of the set.

Proof:

Suppose $\alpha' \in t$ is provable. Then $\alpha' \in t$ must be satisfied by Σ's structure since the axioms are. Therefore, t must denote a set and α must be a member of it.

Suppose t denotes a set and α is a member of the set. By Lemma 5.10 t is either W' or of the form $ad(W',\alpha'_1,...,\alpha'_k)$. If t is W' then $\alpha' \in W'$ is an axiom. Otherwise, $\alpha' \in t$ is provable by repeated applications of axiom 5.[]

Thus, Σ is a two-level system which is equivalent to a closed normal default theory in the sense that the sentences provable from each possible axiom set are contained in an extension and every extension corresponds to the set of sentences provable from an increasing sequence of possible axiom sets. The set of axioms of the metatheory is countable and the axioms for A_p correspond directly to the defaults of the default theory. As a result, Σ corresponds well with our intuitive view of default reasoning as a process of introducing a new assumption because it is justified by our current assumptions. We were able to define the meta-axioms for A_p in a natural way because, unlike the case for arbitrary closed default theories, the extensions of a closed normal default theory may be defined in terms of a sequence of increasing sets of assumptions where each set contains only finitely many more wffs than its predecessor.

5.4 The Effect of the Nonmonotonicity of a Default Theory on the Equivalent Two-Level System

The relation between default theories and two-level systems leads to two observations about the nonmonotonicity of default theories. In the following result we consider nonmonotonically related closed normal default theories to illustrate these two points.

Recall that it is possible to have two default theories, (D,W) and (C,V) such that $D \subseteq C$, $W \subseteq V$ and yet have a formula α such that some extension E of (D,W) contains α but no extension F of (C,V) contains α. Both D and W are viewed as representing the axioms of (D,W). Thus, the default theories (D,W) and (C,V) are nonmonotonically related.

Suppose (D,W), (C,V) are closed normal default theories such that $D \subseteq C$ and $W \subseteq V$. Suppose also that α is a wff such that $\alpha \in E$ where E is some extension of (D,W) and $\alpha \notin F$ where F is any extension of (C,V). Let Σ and Σ' be the two-level systems generated by (D,W) and

(C,V). We first show in what sense the theories of Σ are monotonically related to those of Σ' and then discuss the apparent nonmonotonicity of (D,W) and (C,V). Recall that the default assumptions of a default theory (D,W) are those formulas β for which there is some default in D of the form α:Mβ/β.

We say that formal theory A is a proper extension of formal theory B if all theorems of B are also theorems of A. Also, a model for theory B is a submodel of a model for theory A if the domain of B's model is contained in the domain of A's model, and all constant, function, and relation symbols that have a given interpretation in B's model have the same interpretation in A's model.

Theorem 5.7

Let Σ, Σ' be as above.

a) If the metatheory of Σ' is a proper extension of Σ, then the intended interpretation of Σ is not a submodel of the intended interpretation of Σ'.

b) There exists a finite set $\{\alpha_1,\ldots,\alpha_k\}$ of default assumptions such that for some possible axiom set, A, of Σ $\{\alpha_1,\ldots,\alpha_k\} \subseteq$ A but $\{\alpha_1,\ldots,\alpha_k\}$ is not a subset of any possible axiom set of Σ'.

Proof:

Part a.

Since α is a member of an extension of (D,W) iff α is provable from some possible axiom set of Σ, the assumption that $\alpha \in E$ but $\alpha \notin F$ is equivalent to assuming that there is a possible axiom set of Σ, say A, such that $A \vdash \alpha$ but that α is not provable from any possible axiom set of Σ'.

Suppose W = V. Then A is a possible axiom set of Σ' since we have the same initial object level axiom set as for Σ and we have among the axioms of Σ''s metatheory all the axioms of the metatheory of Σ. Therefore, if W = V, the possible axiom sets of Σ are also possible axiom sets of Σ'. Thus, in this case α would be a member of some extension of Σ' as well as of Σ. Hence, for α to exist we must have that $W \neq V$. But then the metatheory of Σ' contains the axiom $A_p(V')$ instead of the axiom $A_p(W')$ contained in the metatheory of Σ where V' must be interpreted as V and W' as W. For the metatheory of Σ' to be an extension of that of Σ we would have to have $W' = V'$. But then the intended interpretation of Σ is not a submodel of Σ''s interpretation.

Part b.

Since α is a member of an extension of (D,W), there is a possible axiom set of Σ, say A, such that $A \vdash \alpha$. If there were a possible axiom set of Σ', say B, such that $A \subseteq B$, then we would have that $B \vdash \alpha$ contradicting the assumption that α is not a member of any extension of Σ'. Therefore, there must be some finite subset of A, say $\{\alpha_1,\ldots\alpha_k\}$, such that α_i occurs in the proof of α for each i and $\{\alpha_1,\ldots,\alpha_k\}$ is not a subset of any possible axiom set of Σ'. Since $W \subseteq V$ and V is a subset of every possible axiom set of Σ', α_i must be a default assumption for each i.[]

The above result states that Σ''s metatheory cannot be an extension of Σ's metatheory in any meaningful way. The meaning of the meta-axiom $A_p(W')$ in Σ is that W represents the set of all initial assumptions about which the reasoner may reflect. The point we wish to make about this axiom is that it represents an assertion which in effect must actually be employed by a reasoner doing default reasoning. The reasoner must know what he knows if he is going to apply rules of the form: If α is consistent with what I know... The assertion describing what the reasoner knows initially cannot be deduced in a two-level system but must be assumed. If a new assumption is added to the set about which the reasoner reflects, then the meta-level assumption has also changed.

It now is made for a new, larger set of assumptions.

We argue that an assumption about what is initially known is implicitly present in the informal interpretation of a default as stated by Reiter. If $\alpha:M\beta/\beta$ is to mean: Infer β if α follows from what is known and β is consistent with what is known, then the reasoner in applying such a rule must be taking cognizance of what he knows. The fact that the reasoner initially knows the contents of the set W cannot be deduced within the default theory.

We have shown that for a closed normal default theory "what is known" can be identified initially with the initial axiom set and subsequently with the union of this set and the set of default assumptions introduced up to the point when the rule is applied. Thus, the assertions of the predicate A_p in the corresponding two-level system are just an explicit representation of the reasoner's assertions concerning what is known. Although these assertions are made explicit by the meta-axioms of Σ and Σ', we argue that they are implicit in (D,W) and (C,V). Otherwise Reiter's intended interpretation of a system like (D,W) as a set of initial assumptions and rules for introducing new assumptions does not make sense.

The second part of the result shows the connection between the nonmonotonicity of default theories and the hypothesis that default assumptions are logical consequences of the reasoner´s initial assumptions.

Consider the wff α which is assumed to belong to an extension of (D,W) but not to any extension of (C,V) in the above example. Recall that any extension of a default theory is just the (conventional) deductive closure of the union of W and the set of default assumptions belonging to the extension. Because of this fact, α must either be a default assumption or its proof must depend on a default assumption. Otherwise, α would be provable from W and hence be a member of an extension of (C,V).

This observation along with part b of Theorem 5.7 in effect tells us that (D,W) and (C,V) are nonmonotonically related only because we choose to consider default assumptions to be logical consequences. It is the default assumptions which can be made to disappear by adding information. In a default theory a default assumption is a logical consequence. In a two-level system a default assumption is a type of axiom, not a consequence. Thus, the nonmonotonicity of default theories depends on the hypothesis that default assumptions are logical consequences rather than on any intrinsic property of default reasoning.

5.5 Conclusions

The results of this chapter show certain important connections between the properties of default theories and the hypotheses about default reasoning on which their definition is based. First, we found that the lack of a notion of rule in a default theory depends on a combination of what we might call the "non-normal form" hypothesis and the postulate that default assumptions are logical consequences. Second, we found that nonmonotonicity also depends on the second of these two hypotheses. Thus, we find two unusual properties of default theories depending on hypotheses about default reasoning that are not intuitively appealing.

The results of this chapter also provide evidence for our claim that the notion of a two-level system serves at least as well for a model of default reasoning as does the concept of a default theory. For any closed default theory there is an equivalent two-level system. Because of this fact we can claim that the definition of two-level system subsumes that of default theory in a formal sense. We also claim (and this point seems more important) that the notion of two-level system does at least as well in capturing the informal concept of default reasoning.

For an arbitrary closed default theory the equivalent two-level system turns out to be intuitively unappealing. However, the two-level system's lack of appeal is just a reflection of a corresponding lack in the default theory. The notion of an arbitrary default theory is a generalization which has no basis in the informal examples considered in [R]. The motivation for the notion of a default theory is clearly the desire to model arguments like:

> If x is a bird and nothing contrary to the assumption that x can fly is known, then it is reasonable to assume that x can fly.

Such arguments are better represented by the notion of a normal default theory than by the more general notion of arbitrary default theory. But in the case of a closed normal default theory we find that there is an equivalent two-level system which is actually more appealing than the default theory it corresponds to. This two-level system not only generates the same sets of consequences as Reiter expects to get through default reasoning arguments like the above example, it also allows us actually to represent the arguments - something that cannot be done in a default theory.

6. A Comparison of Nonmonotonic Theories and Two-Level Systems

Because of the nature of the definition of a nonmonotonic theory it is not appropriate to attempt the same sort of comparison with two-level systems as we made between default theories and two-level systems. It is possible to show that for any nonmonotonic theory there is a two-level system such that the theorems of the nonmonotonic theory are just the theorems of the two-level system´s only object-level theory. However, this result is not useful for two reasons.

First, the wffs of the two-level system´s object language must be those of L_M, the language of the nonmonotonic theory. Now it must be possible to give a meaning to the object level theorems of a two-level system if that system is to be of interest. As we have already pointed out, however, the question of interpreting the symbol M occurring in the wffs of L_M is problematic. We could employ the interpretation defined in [M] which relies on an underlying modal system, but this would be begging the question.

In the case of a default theory, showing that there existed a two-level system which defined the same sets of consequences as the default theory constituted evidence for our claim that a two-level system is as satisfactory a model as a default theory. This is because the meaning of the consequences does not depend on the notion of a default theory. Instead the consequences are ordinary first order formulas which could be given an interpretation in the usual way. In the case of a nonmonotonic theory we would instead have to rely on a notion of semantics which is essential to the model we are considering. Thus, in effect we would not have a significantly different explanation of default reasoning.

Another factor which makes the formal comparison of nonmonotonic theories and two-level systems of little use is the lack of a "normal" nonmonotonic theory. We believe that our claim of the comparability of default theories and two-level systems is made much more convincing by the fact that in the case of a normal default theory the equivalent two-level system can actually be said to do better at capturing the informal concept of default reasoning than the default theory does. Since there is no "normal" nonmonotonic theory corresponding to a normal default theory, the only sort of two-level system we can define that is equivalent to a nonmonotonic theory will share with it the problem of having no rules for the adoption of default assumptions.

We will thus take a different approach to the comparison of nonmonotonic theories and two-level systems. We first study the notion of a nonmonotonic theory in order to isolate the motivations for its form. We then argue that the notion of a two-level system could be extended to include the concept which underlies the main difference between the definitions of nonmonotonic theory and default theory. To illustrate our claim we present an example of such an extended system.

6.1 Distinctions Between Nonmonotonic and Default Theories

We have argued that the notion of a two-level system does at least as well in satisfying the motivation behind the definition of a default theory as does the default theory itself. The definition of a nonmonotonic theory is intended to serve the same purpose as that of a default theory: to model informal default reasoning arguments. However, there are differences between the points of view of the two models.

First, the definition of a nonmonotonic theory does not include an explicit notion of default. Second, while a default theory may define any number of extensions there is only one set of theorems for a nonmonotonic theory. Third, the symbol M is a part of the language of a nonmonotonic theory.

The absence of an explicitly defined class of syntactic objects corresponding to defaults is not as great a difference as it may seem to be, because both default and nonmonotonic theories are intended to represent the same intuitive concept. In both cases the basic idea is that the reasoner accepts, say, β because it is not contrary to what the reasoner knows (and because there is some reason to think β is likely to be true, of course). In terms of the assumptions used by Reiter and by McDermott and Doyle this is to say that β is inferred by the reasoner because it is consistent with what is known and because some known facts make β likely. At the same time, it is desired that the symbol M, introduced in L_M, should mean "is consistent with what is known" which is just what the symbol M was intended to mean in a default. Thus, any default has a corresponding wff in L_M.

For example, the default α:Mβ/β corresponds in its intended meaning to the intended meaning of the wff α & Mβ → β of L_M. Furthermore, the intended meaning of a default appears to characterize very well the form that one would want a "default inference rule" to take if one began with the hypotheses about default reasoning shared by

Reiter and McDermott and Doyle. Thus, although there occur in L_M many wffs containing M that do not correspond to a default, they do not fulfill any apparent role in modelling default reasoning. Certainly, no such role is demonstrated for these formulas by McDermott and Doyle. A possible exception to this conclusion is the class of wffs of L_M which contain nested occurrences of M, for example, $MM\alpha$. We will consider the significance of such formulas below.

In any case, it seems clear that most if not all wffs of L_M that contain M, but that do not contain nested occurrences of M, could be naturally expressed as metalevel formulas in a two-level system just as defaults can be so expressed. Given the intended meaning of a formula such as $\alpha \rightarrow (M\beta \ \& \ \gamma)$, a corresponding metalevel formula in a two-level system could be something like $Pr(S,\alpha) \rightarrow {\sim}Pr(S,{\sim}\beta) \ \& \ Pr(S,\gamma)$ in the notation of Chapter 5. The wff of L_M is supposed to mean "α implies β is consistent with what is known and γ". The metalevel wff of the two-level system would mean "if α is provable from S then β is consistent with S and γ is provable from S". Considering the differences between the underlying hypotheses of the two approaches, this seems to be the appropriate translation.

The only apparent problem with rendering formulas of

L_M containing unnested occurrences of M into metalevel formulas of a two-level system occurs with L_M formulas of the form $\forall x M\alpha$. The counterpart of this formula in a two-level system is not obvious. Although the meaning intended for such a formula by McDermott and Doyle is unclear, if we assume what seems the most likely interpretation there does appear to be a way to deal with such formulas in a two-level system.

For example, consider $\forall x MP(x)$. It seems most likely that this formula is supposed to mean that for any term t it is consistent to believe P(t). It should be possible to express the same sort of thing in a two-level system with an appropriate metatheory. In the two-level systems considered in Chapter 5 object-level wffs can only be treated as atomic objects in the metatheory. However, instances of metatheories in which the wffs of the object language can be treated as structured objects are well known. It seems likely, therefore, that a two-level system could be constructed in which object level formulas can be manipulated so as to allow one to express the apparent intended meaning of $\forall x MP(x)$.

Recall that given a set A of wffs of L_M

$$As_A(S) = \{M\beta \mid {\sim}\beta \notin S\} - Th(A)$$

$$NM_A(S) = Th(A \sqcup As_A(S))$$

and that FP(A) is the class of all sets S such that S = NM_A(S) while TH(A) is the intersection of the members of FP(A). Thus, the components of the definition of a nonmonotonic theory that correspond most closely to the extensions of a default theory are the members of FP(A). However, instead of taking Reiter´s approach and viewing each member of FP(A) as a "possible world" that the reasoner could accept if he began with the initial assumptions of A, only those formulas in the intersection of the members of FP(A) are treated as consequences of A. No argument is given in either [MD] or [M] as to why this course is chosen instead of an approach similar to Reiter´s.

Reiter´s allowance of multiple extensions appears to stem from an informal conception of the default reasoning process similar to the one presented in Chapter 4. There we imagined the basic default reasoning step to be one of adding to one´s current assumptions a new assertion which does not follow from the current assumptions by conventional deductive inference. Each time a new assertion is added the possibility of adding certain other assumptions at some later point in the process is ruled out. For example, the addition of a given assertion precludes the later addition of its negation. This view of default reasoning leads naturally to the notion of various sequences of assertions. Each sequence represents one possible set of choices made in incrementally adding default assumptions to a given initial set. Although two such sequences might contain the same assertions in different order, in general they would contain different assertions and represent incompatible chains of default reasoning. The (conventional) deductive closure of each such sequence is an extension. As we saw in Chapter 5, this view can be formalized, though not in the way in which Reiter attempted to do so.

The basis for multiple extensions, then, is the notion that the reasoner's acceptance of a given default assumption could rule out others that might themselves be accepted if the given assumption had not already been adopted. One possible way for such a situation to occur is when the reasoner is faced with a default assumption, say α, such that there is evidence for the acceptance of either α or $\sim\alpha$ and both are consistent with what he knows. The natural response to such a case is to accept neither α or $\sim\alpha$ and wait for further evidence. However, it might be the case that the reasoner is forced to accept one or the other. For example, we might be faced with having to decide whether or not to continue funding the search for extraterrestrial life. We can view our decision as an implicit acceptance of either the existence or nonexistence of such life, a default assumption in either case. However, other explanations of this example seem possible.

In fact, this question, which apparently divides Reiter´s approach from that of McDermott and Doyle, of whether or not accepting a default assumption can rule out other equally acceptable default assumptions remains open. The approach taken in defining the notion of a two-level system can accommodate either possibility. We saw in Chapter 5 that a two-level system can be defined which generates multiple incompatible sequences of default assumptions, but it is equally possible to define a two-level system whose rules never result in incompatible default assumptions. That is, if more than one sequence of default assumptions could be generated, they would differ only in order. The set of default assumptions that could be generated would always be the same. Furthermore, such a system would have a more satisfying nature than a set of assertions defined, in the manner of McDermott and Doyle, as the intersection of a collection of fixed points.

It is also interesting to note that when McDermott considers in [M] the problem of doing default reasoning he suggests that the reasoning agent must "be brave" and work within a single fixed point anyway instead of somehow trying to determine that a default assumption occurs in all fixed points. However, the problem of an effective process for deriving the members of even one fixed point remains unsolved.

Because the symbol M is part of the language L_M, one can construct all sorts of wffs in this language which do not resemble Reiter's defaults. We have already noted that such formulas, in the case that they do not contain nested occurrences of M, could be represented in a natural way as part of a two-level system. We now consider the importance of nested occurrences of M.

The intended meaning of a formula such as $MM\alpha$ would be something like "It is consistent to believe that it is consistent to believe α". What would the ability to construct such assertions have to do with default reasoning? The examples of informal default reasoning that we have previously considered do not require such assertions. They all took the form "If there is evidence for α and α is not contrary to what is known, then it is reasonable to assume α". Even if we interpret "not contrary to" as consistent with, we only need to be able to assert that α is consistent with what is known. This would just be $M\alpha$ under the intended interpretation of M. Of course, if α itself turned out to be, say, $M\beta$, then we would find ourselves wanting to assert $MM\beta$. However, in each of our examples α in fact is a simple first-order assertion containing no claims about consistency. Why might a default assumption include M? Is the meaning of an informally stated default assumption ever such that its translation into a formula of L_M would require the symbol M? No examples of such default assumptions are given in either [MD] or [M], and indeed, it is difficult to conceive of any.

Because we are assuming here that the phrase "not contrary to" from informal default reasoning rules is to be interpreted as consistent with, there are two possible ways in which M could occur in the formalization of a default assumption. In particular, a given instance of a default assumption containing M might or might not be a formula which itself could be viewed as a rule for inferring a default assumption. For example, the formula α & $M\beta \rightarrow \beta$ would constitute a rule for introducing the default assumption β under McDermott's and Doyle's intended interpretation. On the other hand, a formula such as $M\alpha$, which is merely supposed to mean that it is consistent to believe α but does not allow the inference of α, would not be such a rule.

Default reasoning involves what we may call a local criterion and a global criterion. The local criterion can be expressed as "There is evidence for α" while the global criterion is "α is not contrary to what is known". The use of consistency as the global criterion for accepting a default assumption hides the fact that the two example formulas above reflect two possibly distinct questions about default reasoning. In the case of α & $M\beta \rightarrow \beta$, we are actually considering the possibility that one can do default reasoning about default reasoning. To do so would involve reasoning about an assertion that refers to a global property of the reasoner´s knowledge, that property being consistency according to the hypotheses of McDermott and Doyle. In the case of $M\alpha$, we are considering the more general case of doing default reasoning about an assertion stating some global property of the reasoner´s knowledge. In general, this global property need not have anything to do with default reasoning, and the assertion might just as well refer to some other property than consistency. In both examples, then, the real question concerns assertions about global properties of the reasoner´s knowledge.

Although no explanation is given in either [MD] or [M] as to why the symbol M is included in the language, the main motivation is obviously the assumption that default reasoning may require the ability to reason about assertions which themselves contain references to consistency. The focus on consistency, however, apparently stems from the assumption that consistency is the criterion for acceptance of a default assumption. We see that when the consistency hypothesis is dropped the problem is really a more general one of dealing with default reasoning about assertions that contain global reference to the reasoner´s knowledge.

How default reasoning can involve reasoning about assertions of such global properties, if at all, is another open problem. McDermott and Doyle do not give any actual examples of such reasoning, nor do we know of any. We can easily see that if such examples do exist, they cannot be characterized either by a default theory or a two-level system. However, we will see below that the notion of a two-level system can be extended to handle this problem as it appears in the context of our hypotheses about default reasoning.

We have noted three major distinctions between the notions of default theory and nonmonotonic theory. Concerning both the absence of an explicit notion of default and the single set of theorems defined by a nonmonotonic theory, we have argued that when each formalism is viewed as an attempt to model default reasoning the distinction in question is not of real interest. We have also argued that, anyway, a two-level system can accommodate the way in which each of these matters is handled in a nonmonotonic theory. In the third case, the distinction appears to be potentially of greater importance. There would not be any natural way to handle the nonmonotonic theory approach in this instance within a two-level system. We must therefore consider an extension of the notion of a two-level system.

6.2 The Notion of an N-Level System

Let us consider the effect of our view of default reasoning on the problem of default reasoning about an assertion that refers to global properties. The essential idea is that one might wish, for example, to justify a default assumption that refers to some global property of the reasoner´s knowledge. We have taken the view that a reasoner, in deciding whether a default assumption is contrary to his knowledge, does not in fact consider all his knowledge but some part of it. We have further assumed that that part of his knowledge can be represented by an object theory while the default reasoning process that concerns this knowledge can be represented by an inference process within a corresponding metatheory. Thus, the particular global property of the reasoner´s knowledge that must be asserted to hold in a default reasoning argument becomes a property of an object theory. A natural generalization is that any global property should be treated as a property of an object theory.

In other words, our view of default reasoning leads us to consider the reasoner´s knowledge to be divided into components which we represent by various object theories and a metatheory. Once the reasoner´s knowledge is so divided we no longer need refer to vague global properties of some kind; instead we can refer to properties that apply to some component. However, if we have only two levels, assertions referring to properties of components can only refer to properties of the object theories on the first level and can only exist themselves in the metatheory on the second level. Therefore, in a two-level system we cannot do default reasoning about such assertions. We can at most employ them in the default reasoning process.

Thus, if we want to do default reasoning about assertions referring to what correspond to global properties in our view, the obvious solution is to add another level. If we have a third level, it is then possible to reason in the theory of the third level about assertions at the second level, including those that happen to assert properties of the theories at the first level. Furthermore, we can arrange to have more than one theory at the second level, thus allowing the possibility of default reasoning about assertions at the second level. However, this does not mean that we must contemplate adding infinitely many levels.

Adding a third level would be intended to represent the ability of the reasoner to reason about that part of his knowledge represented at the second level. The second level is itself intended to represent his knowledge about that part of his knowledge represented by the first level. Finally, the first level is intended to represent the reasoner's knowledge about matters other than his own knowledge. There is no *a priori* reason for us to suppose that the reasoner can somehow reason about *everything* that he knows. It is just as plausible to postulate that there is certain information that can be employed by the reasoner but which he cannot reason about. Also, if we are attempting to represent how a machine might do default reasoning, having only a finite number of levels seems more promising. We therefore propose to define a notion of n-level system, analogous to a two-level system, that assumes only a finite number of levels.

Let n be an integer greater than 1. An *n-level system* is defined to be a collection S of formal theories and a collection of interpretations S' such that:

1. Each theory of S is assigned to some <u>level</u> j where $1 \leq j \leq n$, and each level j is assigned at least one member of S with level n being assigned exactly one theory.

2. Each theory assigned to level j, $j \geq 2$ has a unique interpretation in S´, and each interpretation in S´ is an interpretation for a unique theory in S.

3. If T is a member of S assigned to level j, $2 \leq j \leq n$, then there is a two-level system Σ such that T is the metatheory of Σ, each of the object theories of Σ is a member of S assigned to level j - 1, and T´s interpretation in S´ is the intended interpretation of the metatheory of Σ.

4. If T is a member of S assigned to level j, $1 \leq j \leq n-1$, then there is a two-level system Σ such that T is an object theory of Σ, all other object theories of Σ are among the members of S assigned to level j, the metatheory of Σ is a member of S, say T´, assigned to level j + 1, and the interpretation of T´ in S´ is the intended interpretation of Σ´s metatheory.

This definition, which depends on our previous definition of a two-level system, simply collapses to the earlier definition if n is 2. If n is larger than 2 we have a system in which each theory except for the one assigned to level n is an object theory of some theory assigned to the next higher level. Similarly, each theory except for those assigned to level 1 is a metatheory for some collection of theories assigned to the next lower level. In addition the relation between any given metatheory and its object theories is the same as that established in the definition of a two-level system.

We sketch an example of a three-level system based on the two-level system Σ defined in section 5.3. The level-1 theories will be the object level theories of Σ. (Recall that the object theories of a two-level system are defined by the system's possible axiom sets.) The level-2 theories will be defined from the metatheory axioms of Σ. Let us call the level-2 theories T_j, $0 \leq j \leq \infty$. We define T_0 to be the metatheory axioms of Σ labelled 1 through 9 (note that this is an infinite set). For each j greater than 0 we define T_j to consist of the axioms of T_{j-1} plus the metatheory axiom of Σ labelled 9_{j-1}.

A collection of two-level systems can be defined using the theories of level 2 as their metatheories in the following way. Let Σ_0 be the two-level system whose metatheory is T_0 and whose only possible axiom set is W, the initial possible axiom set of Σ. (An appropriate interpretation of T_0 could be defined as the intended interpretation.) For j greater than 0 let Σ_j be the two-level system whose metatheory is T_j and whose possible axiom sets are those of Σ_{j-1} along with those axiom sets which would be generated by applying axiom 9_{j-1} to the possible axiom sets of Σ_{j-1}. (Again, include an appropriate interpretation for the metatheory.) It is easy to see from the definition of Σ that each Σ_j is a two-level system. Furthermore, the theories of levels 1 and 2 are seen to satisfy the above conditions for an n-level system.

Finally, we can define a theory for level 3 which would be similar to the metatheory of Σ (though not in the same language as Σ). This theory then serves as the metatheory of a two-level system, say Σ^*, whose possible axiom sets are defined by the theories of level 2, the initial possible axiom set being the axioms of T_0.

The three-level system that we have sketched would allow default reasoning about rules for doing default reasoning. The system is an artificial one. There is no reason to suppose that the rules for default reasoning defined in the metatheory of Σ in Section 5.3 are incompatible and cannot be accepted simultaneously. However, our example does illustrate the possibility of going beyond a two-level system to allow the representation of default reasoning about assertions referring to global properties of the reasoner´s knowledge.

6.3 Conclusions

Unlike the case of default theories, a formal comparison of nonmonotonic theories and two-level systems does not provide useful evidence for our claim that the notion of a two-level system makes as satisfactory a model of default reasoning as does the notion of a nonmonotonic theory. Instead we have argued that a two-level system can formally express, although in a different way than a nonmonotonic theory, the concepts underlying the differences between the definitions of default theory and nonmonotonic theory. In fact, for the concept underlying the most significant of these differences, the presence of the symbol M in the language of nonmonotonic theories, we cannot give an adequate formalization in terms of a two-level system. We can, however, begin to formalize this concept by extending the notion of two-level system to n-level system. The need for the notion of an n-level system remains in doubt since no actual examples of default reasoning which would necessitate such a system have as yet appeared.

7. Heuristically Based Default Reasoning Systems

In [W] Winograd gives a survey of what he terms extended modes of inference. A number of computational systems are described and an attempt is made to determine their common characteristics. This leads him to hypothesize that computer systems can be devised to perform certain types of inference which he claims are not formalizable in standard logical terms. The various forms of inference considered turn out to be essentially default reasoning based on certain types of heuristic rules. It is easy to see that the various sorts of rules discussed cannot be modelled by either default or non-monotonic theories, let alone ordinary formal theories. However, we argue in this chapter that such rules can be modelled by two-level systems, which is really to say that they can be modelled in standard logical terms.

Four categories of procedures for performing heuristically based inferences are given, each representing a principle that forms the basis of a type of default reasoning:

1. Procedures which infer a formula as the result of the presence or absence of certain formulas in

memory.

2. Procedures which infer a formula if a finite deductive procedure fails to prove a certain formula.

3. Procedures which attempt inferences in a certain order.

4. Procedures which infer a formula if a resource-limited procedure fails to prove a certain formula.

Note that the fourth category is a special case of the second since resource-limited procedures are a subclass of the class of all finite procedures.

We first examine the definitions of Categories 1, 2, and 3, to determine their relationship to the notion of default reasoning and to isolate for each the reason for the claim that it represents a form of inference not reducible to conventional inference. We then show how the concept of default reasoning procedures underlying each category can be modelled in a two-level system.

7.1 Category 1

We begin with an example discussed by Winograd. If we are asked whether the Mekong River is longer than the Amazon, we might conclude that it is not, since the Mekong being longer than the Amazon would be such a significant fact that we would know it if it were true. Here the word "know" clearly means something like "be aware of the truth of". Letting α be an arbitrary assertion, we can generalize this example to a rule of the form: If we are not aware of the truth of α and α is such that if it were true we would likely be aware of it, then it is reasonable to assume that α is false. Such a rule represents an intuitively valid justification for the assertion that an assumption is reasonable and is, in fact, a form of default reasoning. For an assertion such that we would be likely to be aware of its truth if it were true, not being aware of its truth constitutes absence of evidence contrary to the negation of the assertion.

There are two problems in constructing a computer reasoning system employing the above form of default reasoning. One is how one should define the set of assertions of whose truth the system will be "aware" at any point in a computation. The second problem is how the system is to decide for a given assertion whether it would likely be aware of the assertion's truth if it were true.

Suppose we wish to construct a reasoning system which employs conventional inference along with the above rule. One way, and the way suggested by Winograd´s characterization of Category 1, to define the set of assertions of whose truth the system is aware at any point during a computation is to define it as those assertions occurring explicitly in the system´s memory at that point. The set would therefore consist of those assertions assumed true initially or through the default rule along with those assertions that had previously been logically inferred by the system from its assumptions. The significant feature of this definition for our purposes is that it makes a distinction between assertions which are logical consequences of the system´s assumptions and have already been explicitly inferred by the system and those which are also logical consequences but have not been explicitly inferred. Within a formal theory there is no way to distinguish those theorems for which a proof has been constructed from those for which we have no proof. They are all equally theorems of the formal theory. Therefore, any rule making use of the distinction between proved and unproved theorems cannot be part of a formal theory, and we find here the basis for the claim that a procedure belonging to the first category defines a form of inference which cannot be explained in terms of inference in a formal theory. The definition of the Category 1 is a generalization of the idea of defining the assertions that a reasoning system knows to be those that occur in memory.

The notion of generating an assertion as the result of the presence of certain formulas in memory need not concern us. The procedure generating an assertion from those in memory is effective since it is part of a computer system. Thus, this case is just conventional inference, though the procedure may represent some nonstandard inference rule. Our task, therefore, will be to demonstrate the possibility of modelling by a two-level system the behavior of a system which asserts formulas from the absence of assertions in memory.

7.2 Category 2

The definition of the second category is motivated by the observation that often after we have expended a certain amount of effort to infer an assertion and have failed, we decide that the assertion is probably false. If, having arrived in this way at the conclusion that an assertion is probably false, we then assume the negation of the assertion, we have introduced a default assumption. A general rule would be: If an attempt has been made to infer α and the attempt has failed, and if α is such that if it were true the attempt would have been likely to succeed, then it is reasonable to assume that α is false.

As in the case of the first category, there are two problems in constructing a computer reasoning system employing this form of default reasoning, one of which is addressed by the definition of the second category. The first problem is how to handle the notion of an attempt to infer α. The second is how to decide when α belongs to the class of assertions which may be assumed false after the failure of an attempt to prove them.

In a computer reasoning system, an attempt to infer an assertion would just be the execution of some procedure that would take as its input the assertion to be proved along with the assertions accepted as true by the system. Here, we are considering an inference attempt that ends at some point and so corresponds to a total procedure, that is, one which halts for every input. Thus, we have the notion of a computer reasoning system which generates an assertion as the result of the failure of a (total recursive) subprocedure to infer the negation of the assertion. The definition of the second category is a generalization of this notion.

In general we cannot express within a formal theory a procedure for inferring theorems of that theory. Hence we cannot include a rule of the type described above. Of course, such a rule introduces new assumptions and so cannot be part of the theory formed by introducing those assumptions in any case. In a subsequent section of this chapter we show how such a rule can be represented in a two-level system.

7.3 Category 3

The third category is illustrated by the example of default reasoning about the properties of birds. In that example we considered the operation of introducing a default assumption asserting that a tern can fly given that the reasoner already held an assumption that most birds can fly, along with assumptions stating that certain individual species of birds cannot fly. In this example we have an assertion stating that a property is generally true of the members of a class as well as several assertions stating that the property is false for certain specifically named members of that class. A general rule expressing the type of default reasoning done in the example would be: If property P is true of most members of a class and nothing is known contrary to the assumption that P is true of a specific member of the class, then it is reasonable to assume that P is true of that individual.

The problems in constructing a computer reasoning system for the above form of default reasoning are making precise the notion of a property holding for most members of a class and handling the idea of not knowing anything contrary to the assumption that that the property holds for some individual. As before, only the first of these problems is involved in the definition of the third category.

To avoid the difficulty of giving a precise definition to a quantifier like "most" computer systems have been constructed in which assertions of the form "Most members of class C have property P" have been replaced by "All members of class C have property P". These systems also include assertions of the form, "I is a member of C and does not have property P". It is left to the program performing inference to deal with the existence of such inconsistent assumptions in the system.

One method for dealing with the kind of inconsistency illustrated is to choose a subset of the assumptions contained in the system and attempt to prove the assertion from these. If the attempt fails after finitely many steps, a new set of assumptions is chosen. The program´s criteria for choosing sets of assumptions are such that no set including a universally quantified assertion is employed until all sets containing only those assertions about individuals which the program considers relevant have been tried. Thus, the program chooses axioms from which to reason in a certain order. In doing so it behaves according to the above rule with "nothing known to the contrary" interpreted as "no proof from a relevant set of assertions about individuals has been found". The intention is that the program be defined in such a way as to preclude any attempt to construct a proof from an inconsistent set of assumptions. It is not known whether such procedures exist for other than trivial cases. However, it is often claimed that humans actually maintain inconsistent sets of beliefs of the sort described above and somehow avoid getting into trouble in their reasoning by using a procedure similar to the one described.

In a later section we consider a hypothetical computer reasoning system of the sort just described and show how to define a two-level system in which certain of the possible axiom sets represent the sets of assumptions which could be chosen by the computer reasoning system.

7.4 Two-Level Systems Based on Heuristic Rules: Some Background Considerations

Before discussing the definition of a two-level system modelling the concept motivating each of the three categories, we first establish some necessary conventions. We also explain the sense in which we will claim to have captured these concepts.

We assume for our discussion that the assertions treated by these systems are expressed in some formal language L. This ignores the question of whether the languages used by some of the systems discussed by Winograd can be considered formal languages, but the significant characteristics of the heuristic default rules being considered do not depend on the choice of language. We also assume that, with the exception of some given default inference rule, all inference rules employed by a system are conventional. This assumption also does not effect the properties of default inference rules that we wish to consider.

The notion of a two-level system relies on the ability to reason about sets of wffs in a language. Each category defined by Winograd represents a heuristic principle of default reasoning. For example, the first category represents the principle of introducing an assumption because its negation is not in memory rather than because it is consistent with current assumptions. We will argue that the principle represented by each category can be thought of as a principle for reasoning about sets of wffs in a language and so can be incorporated in a two-level system.

First we relate the notion of reasoning about sets of wffs to the sort of computations done by computer reasoning systems. Typically, a system of the sort considered by Winograd begins with an initial set of assumptions in memory. All wffs in memory at any time during a computation are considered true. (Actually, the wffs to be considered true might be explicitly marked or located in a particular part of memory, the truth value of wffs not so marked or located being left indeterminate, but this refinement could easily be handled by slightly complicating our argument, so we ignore it.) Thus a default assumption is introduced simply by placing the formula in memory. It is also usual for the system to add each newly inferred formula to memory as it is inferred. At any point during a computation, therefore, memory contains the initial assumptions, any default assumptions which have been introduced up to that point, and any formulas which have so far been inferred from other formulas in memory. (We will ignore here the possibility of deleting assumptions and of adding new assumptions which are not default assumptions.) The computations of such a system can thus be described by a finite or infinite sequence of finite sets of wffs of L, say $S_1, S_2, \ldots,$ that has the following properties:

1. S_1 is the initial set of axioms (note that S_1 is finite);

2. If S_j and S_{j+1} are consecutive members of the sequence then $S_{j+1} = S_j \sqcup \{\alpha_j\}$ where either α_j is a default assumption whose introduction is justified by applying the system's default inference rule to the members of S_j or α_j is the result of applying a conventional deductive inference rule to members of S_j;

3. For each i and j, $S_i \neq S_j$ if $i \neq j$.

The members of such a sequence represent the contents of the system's memory at successive stages of the computation. Let us call such a sequence a *memory-state sequence*. The collection of all memory-state sequences determined by a particular set of rules and initial assumption set can be thought of as representing the set of all computations which might be performed by a system using these rules and initial assumptions.

Given the notion of a memory-state sequence it is reasonable to say that the set of wffs which can be accepted as true by the system at any point during a computation is just the deductive closure of the set representing the contents of memory at that point. We can also reasonably say that a two-level system accounts for the type of reasoning done by our hypothetical computer system if it is the case that each object theory corresponds to the closure of some member of a memory-state sequence, and the closure of each member of a memory-state sequence corresponds to some object theory. Thus, our approach will be to define two-level systems meeting these conditions. The system defined will also be such that the meta-axioms for A_P bear a natural relation to the heuristic default reasoning rule being considered.

7.5 Systems Based on "Memory Contents Rules"

The basis of the definition of the first category is the notion of asserting that an assumption is reasonable because of the absence of some formula from memory. The most natural example of this form of justification is the case of asserting the reasonableness of an assertion because of the absence of the assertion´s negation.

Consider a computer reasoning system which employs a rule of the form: If α is such that if it were true then it would already be in memory and, if α is not in memory, then it is reasonable to assume $\sim\alpha$. Since it is the computer system which determines whether a wff is already in memory, the set of potential default assumptions, those wffs which can be assumed if their negations do not occur in memory, must be recursively enumerable. Let us call this set of wffs PA. Thus, the memory-state sequences of this system would be all finite or infinite sequences of sets $S_1,\ldots,S_k,\ldots$ such that:

1. $S_1 = I$ where I is the initial set of assumptions;
2. For each j, $S_{j+1} = S_j \sqcup \{\alpha_j\}$ where either $\alpha_j \in PA$ and $\sim\alpha_j \notin S_j$ or $S_j \vdash \alpha_j$;
3. $S_i \neq S_j$ for distinct i and j.

We can think of PA as defining a predicate, say P, where $P(\alpha)$ is true just if $\alpha \in PA$. In the two-level system defined below we will simply include all true instances of $P(\alpha)$ as axioms. Of course, the problem of which formulas should actually belong to PA is likely to be difficult and is an important issue in its own right. However, it is the notion of introducing an assumption because of the absence of some formula from memory that Winograd contends is outside the concepts of conventional deductive logic, and this contention makes no reference to a realistic approach to handling PA.

In considering the general idea of handling at a metalevel any rule which requires the assertion of the absence of a wff from memory, it is natural to think of defining a predicate which is true of those sets of wffs that could be the contents of the system´s memory at some point during execution. One would then be inclined to attempt to use such a predicate directly in expressing the required rule. However, our definition of a two-level system was made in terms of possible axiom sets. While we would expect the system´s memory always to include the current axioms, we would not expect it to contain only axioms. Thus, we are faced with the problem of handling the notion of the current contents of memory within a system that is defined in terms of the current axioms, usually just a subset of total current memory contents. We will see below that this can in fact be done.

By defining two-level systems in terms of axiom sets (actually, formal theories) we arrive at a model that is based on conventional concepts of logic. If we tried to replace object-level theories in our notion of a two-level system with, say, object-level memory sets, we would cause the difference between default reasoning based on memory contents and conventional inference to appear to be greater than it is.

Let us suppose that the language used by the computer system is L and that the initial set of assumptions is I. Consider the class W of sets consisting of I and all sets of the form $I \sqcup \{\alpha_1,\ldots,\alpha_k\}$ where each α_i is a wff of L. We define two relations by simultaneous recursion over this class. The first, M (for "memory set"), is unary; the second, MA (for "memory set axioms") is binary.

1. $I \in M$;
2. For all sets S and R belonging to W, if $S \in M$, $(S,R) \in MA$, $S \vdash \alpha$, and $\alpha \notin S$, then $S \sqcup \{\alpha\} \in M$;
3. For all sets S and R belonging to W, if $S \in M$, $(S,R) \in MA$, $\alpha \in PA$, $\alpha \notin S$, and $\sim\alpha \notin S$, then $S \sqcup \{\alpha\} \in M$;
4. Nothing else is in M.

In the definition of M, α is any wff of L.

1. $(I,I) \in MA$;
2. For all sets S and R belonging to W, if $S \in M$, $(S,R) \in MA$, $S \vdash \alpha$, and $\alpha \notin S$, then $(S \sqcup \{\alpha\},R) \in MA$;
3. For all sets S and R belonging to W, if $S \in M$,

(S,R) $\in$ MA, $\alpha \in$ PA, $\alpha \notin$ S, and $\sim\alpha \notin$ S, then (S $\sqcup$ $\{\alpha\}$,R $\sqcup$ $\{\alpha\}$) $\in$ MA;

4. Nothing else is in MA.

A third relation, AP, will serve the same purpose as those previously defined with this name:

1. AP is the range of MA. That is, R $\in$ AP if there is a set S belonging to the class such that (S,R) $\in$ MA.
2. Nothing else is in AP.

Each member of M represents a set of wffs that could be the contents of memory at some point during a computation, as we will see below. MA associates with each member of M a set of axioms that generates the same theorems as M. We will also see below that if (S,R) belongs to MA then Th(S) = Th(R). The result of these observations will be that the memory-state sequences can be characterized by sequences of members of AP while every member of AP is a subset of some member of a memory-state sequence. This correspondence will allow us to define the desired two-level system.

Lemma 7.1

S $\in$ M iff there is R such that (S,R) $\in$ MA.

Proof:

Only if:

We use induction on the cardinality of $S - I$. Suppose $S = I$. Then $(I,I) \in MA$ by definition.

Assume that if the cardinality of $S - I$ is n, then there is R such that $(S,R) \in MA$. Consider S such that the cardinality of $S - I$ is $n+1$. We are assuming $S \in M$, which must be as a result of either condition two or three of the definition. Thus, there are $\hat{S}$ and α such that $S = \hat{S} \sqcup \{\alpha\}$, $\alpha \notin \hat{S}$, and $\hat{S} \in M$. By the induction hypothesis there is $\hat{R}$ such that $(\hat{S},\hat{R}) \in MA$.

If $S \in M$ by condition two, then $\hat{S} \vdash \alpha$, and thus $(S,\hat{R}) \in MA$. If $S \in M$ by condition three, then $\alpha \in PA$ and $\sim\alpha \notin \hat{S}$ so $(S,\hat{R} \sqcup \{\alpha\}) \in MA$.

If:

Given S, suppose there is R such that $(S,R) \in MA$. If $S = I$, then $S \in M$. If $S \neq I$, then $(S,R) \in MA$ by condition two or three. Therefore, there is α such that either $S - \{\alpha\}$, α, and R satisfy the prerequisites of condition two or $S - \{\alpha\}$, α, $R - \{\alpha\}$ satisfy the prerequisites of condition three. In either case the prerequisites of the corresponding condition of the definition of M are satisfied and $S \in M$.[]

Lemma 7.2

If $(S,R) \in MA$, then $R \subseteq S$ and $Th(S) = Th(R)$.

Proof:

We use induction on the cardinality of $S - I$. Suppose $S = I$. Since $(S,R) \in MA$ by condition two or three requires that $S = \hat{S} \sqcup \{\alpha\}$ where $\alpha \notin \hat{S}$ and $\hat{S} \in M$, $(I,R) \in MA$ only if $R = I$.

Assume the claim is true for $(S,R) \in MA$ where the cardinality of $S - I$ is n and consider $(S,R) \in MA$ such that the cardinality of $S - I$ is $n + 1$. (S,R) must be in MA by condition two or three. Thus, there must be $\hat{S}$ and *a such that $S = \hat{S} \sqcup \{\alpha\}$, and $\alpha \notin \hat{S}$. If $(S,R) \in MA$ by condition two, then $(\hat{S},R) \in MA$ and $\hat{S} \vdash \alpha$. By the induction hypothesis $R \subseteq \hat{S}$ and $Th(\hat{S}) = Th(R)$. Thus, $Th(S) = Th(R)$ and $R \subseteq S$. A similar argument applies if $(S,R) \in MA$ by condition three.[]

Lemma 7.3

$S \in M$ iff S is a member of a memory-state sequence.

Proof:

Only if:

We use induction on the cardinality of $S - I$. Suppose $S = I$. I is a member of every memory-state sequence.

Assume that if $S \in M$ and the cardinality of $S - I$ is n, then S is a member of a memory-state sequence. Consider $S \in M$ such that the cardinality of $S - I$ is $n + 1$. We know that there are $\hat{S}$ and α such that $S = \hat{S} \sqcup \{\alpha\}$ and $\hat{S}$ and α satisfy either condition two or three of the definition of M. By the induction hypothesis there is a memory-state sequence, say $S_1,\ldots,S_k,\ldots$ such that $\hat{S} = S_k$. Define a new finite sequence, say $Q_1,\ldots,Q_{k+1}$ where $Q_i = S_i$ for $i = 1$ to k and $Q_{k+1} = S$. Then $Q_1,\ldots,Q_{k+1}$ is a memory-state sequence with S as a member.

If:

Let $S_1,\ldots,S_k,\ldots$ be a memory-state sequence. Since $S_1 = I$, $S_1 \in M$. Suppose $S_k \in M$ and consider S_{k+1}. By Lemma 5.15, there is R such that $(S_k,R) \in MA$. Furthermore, $S_{k+1} = S_k \sqcup \{\alpha\}$ where either $S \vdash \alpha$ or $\alpha \in PA$ and $\sim\alpha$, $\alpha \notin S_k$. Therefore, S_k, R, and α satisfy either condition two or three of the definition of M and $S_{k+1} \in M$.[]

Theorem 7.1

If S is a member of a memory-state sequence, then there is $R \in AP$ such that $Th(S) = Th(R)$ and $R \subseteq S$.

Proof

Suppose S is a member of a memory-state sequence.

Then by Lemma 7.3, $S \in M$. By Lemma 7.1 there is R such that $(S,R) \in MA$. By Lemma 7.2 $Th(S) = Th(R)$.[]

Theorem 7.2

If $R \in AP$, then there is S such that S is a member of a memory-state sequence, $Th(S) = Th(R)$, and $R \subseteq S$.

Proof:

Suppose $R \in AP$. Then there is S such that $(S,R) \in MA$. By Lemma 7.1, $S \in M$. By Lemma 7.3, S is a member of a memory-state sequence. By Lemma 7.2, $Th(R) = Th(S)$ and $R \subseteq S$.[]

The above results show that the members of AP as determined by the relations M and MA are just the sets of assumptions (initial and default) which generate the deductive closures of the members of the memory-state sequences. We will next define a two-level system whose intended interpretation includes M and MA. The axioms of the system's metatheory will be such that they allow deduction of assertions corresponding to the true instances of $S \in M$ and $(S,R) \in MA$. This will result in the system's object theories being just the deductive closures of the members of all memory-state sequences as desired.

We now define a two-level system Σ. L', the metalanguage of Σ, consists of:

1. A constant symbol α' for each $\alpha \in L$;
2. A constant symbol I';
3. Four unary predicate symbols: S, M, P, and A_p;
4. Three binary predicate symbols: $\in$, Pr, MA;
5. One binary function symbol, ad;
6. An infinite supply of variables and the usual quantifiers and connectives.

For the intended interpretation we define a structure whose domain consists of $A \sqcup B \sqcup \{I\}$ where A is the set of wffs of L and B is the set of all sets of the form $I \sqcup \{\alpha_1, \ldots \alpha_k\}$. The symbols S, Pr, $\in$, and ad are interpreted over this domain in the same manner as for the two-level system defined for closed normal default theories. The symbols P, M, MA, and A_p are interpreted as the relations PA, M, MA, and AP respectively.

The axioms of the metatheory are as follows:

1. $M(I')$.
2. $Pr(I',\alpha') \;\&\; \alpha' \notin I' \rightarrow M(ad(I',\alpha'))$.
3. $P(\alpha') \;\&\; \alpha \notin I' \;\&\; {\sim}\alpha \notin I' \rightarrow M(ad(I',\alpha'))$.
4. For each $n \geq 1$ and each $k \leq n$,

$$\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall_k (M(ad(I',x_1,\ldots,x_n))\,\&$$
$$MA(ad(I',x_1,\ldots,x_n),ad(I',y_1,\ldots,y_k)) \;\&$$

$\alpha' \notin ad(I',x_1,\ldots,x_n)$ &

$Pr(ad(I',x_1,\ldots,x_n),\alpha') \rightarrow$

$M(ad(I',x_1,\ldots,x_n,\alpha'))$.

5. For each n ≥ 1 and each $k \leq n$,

$\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall_k (M(ad(I',x_1,\ldots,x_n))$ &

$MA(ad(I',x_1,\ldots,x_n),ad(I',y_1,\ldots,y_k))$ &

$\alpha' \notin ad(I',x_1,\ldots,x_n)$ &

$\sim\alpha' \notin ad(I',x_1,\ldots,x_n)$ & $P(\alpha') \rightarrow$

$M(ad(I',x_1,\ldots,x_n,\alpha'))$.

6. $MA(I',I')$.

7. $Pr(I',\alpha')$ & $\alpha' \notin I' \rightarrow MA(ad(I',\alpha'),I')$.

8. $P(\alpha')$ & $\alpha \notin I'$ & $\sim\alpha \notin I' \rightarrow$

$MA(ad(I',\alpha'),ad(I',\alpha'))$.

9. For each n ≥ 1 and each $k \leq n$,

$\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall_k \cdot (M(ad(I',x_1,\ldots,x_n))$ &

$MA(ad(I',x_1,\ldots,x_n),ad(I',y_1,\ldots,y_k))$ &

$\alpha' \notin ad(I',x_1,\ldots,x_n)$ &

$Pr(ad(I',x_1,\ldots,x_n),\alpha') \rightarrow$

$MA(ad(I',x_1,\ldots,x_n,\alpha'),ad(I',y_1,\ldots,y_k))$.

10. For each n ≥ 1 and each $k \leq n$,

$\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall_k (M(ad(I',x_1,\ldots,x_n))$ &

$MA(ad(I',x_1,\ldots,x_n),ad(I',y_1,\ldots,y_k))$ &

$\alpha' \notin ad(I',x_1,\ldots,x_n)$ &

$\sim\alpha' \notin ad(I',x_1,\ldots,x_n)$ & $P(\alpha') \rightarrow$

$MA(ad(I',x_1,\ldots,x_n,\alpha'),ad(y_1,\ldots,y_k))$.

11. $\alpha' \in I'$ for each $\alpha \in I$.

12. $\alpha' \notin I'$ for each $\alpha \notin I$.

13. $\alpha' \neq \beta'$ for each distinct pair of constants α, β.

14. $P(\alpha')$ for each $\alpha \in PA$.

15. Axioms for S, Pr, and ad as in the system for a closed normal default theory.

16. $A_p(I')$.

17. For each $n \geq 1$ and each $k \leq n$,

$$\forall x_1 \ldots \forall x_k (A_p(ad(I', x_1, \ldots, x_k)) \Leftrightarrow \exists y_1, \ldots, \exists y_n MA(ad(I', y_1, \ldots, y_n), ad(I', x_1, \ldots, x_k))).$$

Finally, the possible axiom sets of Σ are just the members of AP. This completes our definition of Σ. We must now show that Σ satisfies the requirements for a two-level system. We do this in the same way as for the system defined for closed normal default theories. The following lemmas are analogous to lemmas already stated in Chapter 5.

Lemma 7.4

Let t be a term of L′ of the form

$$ad(I', \alpha'_1, \ldots, \alpha'_k).$$

Then t denotes $I \sqcup \{\alpha_1, \ldots, \alpha_k\}$.

Lemma 7.5

If $S = I$ or $S = I \sqcup \{\alpha_1, \ldots, \alpha_k\}$, then there is a closed term t of L′ such that t denotes S.

Lemma 7.6

Let t be a closed term in which ad occurs such that t is not of the form $ad(I', \alpha'_1, \ldots, \alpha'_k)$. Then t denotes d, the arbitrary wff specified in the definition of adj.

Lemma 7.7

A closed term t of L′ denotes a set iff $t = I'$ or t is of the form $ad(I', \alpha'_1, \ldots, \alpha'_k)$.

Lemma 7.8

The axioms of Σ′s metatheory are satisfied by the given structure.

Proof:

Axiom one is obvious.

By the definitions of the relations M and MA axioms of type two or three are satisfied.

For an axiom of type four suppose that $M(ad(I', a_1, \ldots, a_n))$ and $MA(ad(I', a_1, \ldots, a_n), ad(I', b_1, \ldots, b_k))$ are satisfied by some assignment. Then by Lemma 5.21, for each i a_i and b_i must be constant symbols, say α'_i and β'_i respectively, denoting wffs of L. Thus, $I \sqcup \{\alpha_1, \ldots, \alpha_n\} \in M$ and

$(I \sqcup \{\alpha_1,\ldots,\alpha_n\}, I \sqcup \{\beta_1,\ldots\beta_k\}) \in MA$. If the wffs $\alpha \notin ad(I',\alpha'_1,\ldots,\alpha'_n)$ and $Pr(ad(I',\alpha'_1,\ldots,\alpha_n),\alpha')$ are also satisfied, then by definition of the relation M $I \sqcup \{\alpha_1,\ldots,\alpha_n,\alpha\} \in M$. Thus, $M(ad(I',\alpha'_1,\ldots,\alpha'_n,\alpha'))$ is also satisfied.

For axioms of type five the argument is similar to that for axioms of type four.

Axiom six is obvious.

Axioms of types seven and eight are similar to axioms of type two.

Axioms of types nine and ten are similar to those of type four.

Types eleven, twelve, thirteen, and fourteen are obvious.

The axioms for S, Pr, and ad are similar to those given above.

Axiom sixteen is obvious.

For axioms of type seventeen the argument is similar to that for axioms of type four.[]

Lemma 7.9

Let t be a closed term of L' of the form

$$ad(I',\alpha'_1,\ldots,\alpha'_k).$$

Then $\alpha' \notin t$ is provable in Σ's metatheory iff $\alpha \notin I \sqcup \{\alpha_1,\ldots,\alpha_k\}$.

Proof:

Only if:

If $\alpha' \notin t$ is provable, it must be satisfied by the intended interpretation. Since t denotes $I \sqcup \{\alpha_1,\ldots,\alpha_k\}$ it must be that $\alpha \notin I \sqcup \{\alpha_1,\ldots,\alpha_k\}$.

If:

By applying the axioms for $\notin$, $\neq$, and ad we can prove in the metatheory:

$\alpha' \notin I', \alpha' \notin ad(I',\alpha'_1),\ldots,\alpha' \notin ad(I',\alpha'_1,\ldots,\alpha'_k)$.[]

Lemma 7.10

If $S \in M$, then there is a closed term s of L′ denoting S such that M(s) is provable in Σ′s metatheory and for every R such that $(S,R) \in MA$ there is a closed term r such that MA(s,r) is provable.

Proof:

If $S = I$, then $M(I')$ is an axiom. Furthermore, I is the only set such that $(I,I) \in MA$ and $MA(I',I')$ is also an axiom.

Suppose the claim is true for all S such that the cardinality of $S - I$ is n. Consider $S \in M$ such that the cardinality of $S - I$ is $n + 1$. S must be in M by condition two or three. In either case there are $\hat{S}$, $\{\alpha\}$, and R such that $\hat{S} \in M$, $(\hat{S},R) \in MA$, and $S = \hat{S} \sqcup \{\alpha\}$. By hypothesis there are closed terms $\hat{s}$ and r denoting $\hat{S}$ and R such that $M(\hat{s})$ and $MA(\hat{s},r)$ are provable. Since $\alpha \notin S$, $\alpha \notin \hat{s}$ is provable, by Lemmas 7.7 and 7.9 If $S \in M$ by condition two, then $\hat{S} \vdash \alpha$ so $Pr(\hat{s},\alpha')$ is also provable. It follows that $M(ad(\hat{s},\alpha'))$ is provable and since $\hat{s}$ denotes $S - \{\alpha\}$, $ad(\hat{s},\alpha')$ denotes S. A similar argument applies in the case that $S \in M$ by condition three.

Let R be any set such that $(S,R) \in MA$. Then since $(S,R) \in MA$ by condition two or three there are $\hat{S}$ and α such that $S = \hat{S} \sqcup \{\alpha\}$ and either $(\hat{S},R) \in MA$ or $(\hat{S},\hat{R}) \in MA$ where $R = \hat{R} \sqcup \{\alpha\}$. By an argument similar to that given above we have that either $MA(ad(\hat{s},\alpha'),r)$ is provable where r denotes R or $MA(ad(\hat{s},\alpha'),ad(\hat{r},\alpha'))$ is provable where $ad(\hat{r},\alpha')$ denotes r.[]

Lemma 7.11

If M(s) or M(s,r) are provable in Σ's metatheory for s and r closed terms of L', then s denotes S and r denotes R such that $S \in M$ and $(S,R) \in MA$.

Proof:

Similar to Theorem 5.5.[]

The previous two lemmas along with Lemma 7.1 tell us also that M(s) is provable for closed s if and only if there is closed r such that MA(s,r) is provable. Thus, Σ

correctly characterizes the relations M and MA.

Theorem 7.3

If S ∈ AP, then there is a closed term t of L′ denoting S such that $A_p(t)$ is provable in Σ′s metatheory.

Proof:

If S ∈ AP, then there is R such that (R,S) ∈ MA. By Lemma 7.1, R ∈ M. Therefore, by Lemma 7.10 there are closed r and s denoting R and S such that MA(r,s) is provable. It follows that $A_p(s)$ is also provable.[]

The last two results needed are the same as results stated for the case of a closed normal default theory considered in Chapter 5.

Theorem 7.4

If $A_p(t)$ is provable in Σ′s metatheory for a closed term t, then t denotes a member of AP.

Theorem 7.5

For any closed term t of L′, α′ ∈ t is provable in Σ′s metatheory iff t denotes a set and α is a member of the set.

Thus, we see that Σ is indeed a two-level system. Furthermore, the definition of Σ directly translates a heuristic default inference rule relying on the notion of the current contents of a system's memory into a (recursive) set of meta-level axioms. This fact provides evidence to support the claim that a heuristic default inference employing the principle of testing memory for the absence of a formula can be modelled within a two-level system.

It is important to note that we can make the same observations about nonmonotonicity as we did in Chapter 5 for default theories. Our definition of a memory-state sequence is such that we can obviously define the provable formulas of the system to be just those that occur as members of some member of a memory-state sequence. For such a system it is possible that if we replace the initial axiom set I by J where I $\subseteq$ J there will be a formula α such that α was provable starting with I but is not provable starting with J. Here we again appear to have nonmonotonic behavior, but, as with default theories, we argue that in addition to the assumptions represented by I and J there are implicit assumptions which must be made explicit before comparing the two systems.

Let us suppose we have two two-level systems Σ and Σ' as defined in the previous section with initial axiom sets I and J where I $\subseteq$ J. Suppose also that there is a wff α such that α is provable in Σ and not in Σ'. We can then show a result similar to that stated in Theorem 5.7.

Theorem 7.6

a) If the metatheory of Σ' is an extension of Σ, then the intended interpretation of Σ is not a submodel of Σ'.

b) There exists a finite set $\{\alpha_1,\ldots,\alpha_k\}$ of default assumptions such that for some possible axiom set, A, of Σ $\{\alpha_1,\ldots,\alpha_k\} \subseteq A$ but $\{\alpha_1,\ldots,\alpha_k\}$ is not a subset of any possible axiom set of Σ'.

Proof:

Similar to Theorem 5.7.[]

Thus, we have a result analogous to the one shown for default theories and can make similar arguments.

7.6 Systems Based on Recursive Deductive Procedures

The definition of Winograd's second category relies on the notion of a total recursive procedure which, given a set of assumptions and an assertion as input, attempts to find a proof of the assertion from the given assumptions and returns "yes" or "no" depending on whether a proof is found. An obvious example of a default reasoning rule employing such a procedure is: If procedure f fails to find a proof of α from the current assumptions and α is such that if it were true f would have been likely to succeed, then it is reasonable to assume $\sim\alpha$.

As in the case of the first category, we can consider a computer reasoning system based on the above rule in combination with conventional inference. The class of potential default assumptions would be recursively enumerable just as the corresponding class was for the first category, and, as we did for the first category, we can treat this class in terms of a predicate, say P. The procedure f which we are postulating is just a realization of a recursive function which we may also call f. Thus, the above rule as it would be implemented in a hypothetical computer system can be thought of as stating that if $P(\alpha)$ is provable and $f(\alpha,A)$ is "no" (where A is the system's current set of assumptions), then $\sim\alpha$ can be introduced as an assumption.

In the case of this system we could provide axioms for P as we did in the previous example. We could also introduce axioms of the form $f(\alpha,A)$ = "yes" for each instance of α and A such that f would return "yes" and of the form $f(\alpha,A)$ = "no" for each instance of α and B such that f would return "no". Since f is recursive the set of such axioms would be recursive. Thus, it is obvious that the approach to be taken in defining a two-level system to account for the type of default reasoning represented by the second category is to employ P and f in the definition of the axiom set predicate A_P. It is easy to see how such a system could be defined in a form similar to the two-level system defined for the first category.

7.7 Systems Based on Inconsistent Sets of Assumptions

Finally, the third category is concerned with the notion of asserting that some property holds for all individuals of a class while also asserting the negation of that property for some members of the class. As mentioned in section 7.3, it is sometimes claimed that humans actually maintain such contradictory assertions. It seems questionable that humans actually believe, for example, that all birds fly. However, it might be useful in a computer reasoning system to use such assertions rather than deal with quantifiers like "most". (But note that this same problem could apparently be dealt with as an instance of default reasoning in other ways that we have already discussed.) If such contradictory assertions are to be employed, an alternative, and to us more reasonable, view of such an arrangement would be to consider the machine as accepting at any one time only a subset of the set of all possible axioms available to it. Thus, the machine might accept either, "All birds can fly" or, "A penguin is a flightless bird", but not both at the same time.

To illustrate this notion let us suppose we have a computer reasoning system employing contradictory assumptions and relying on an algorithm which handles these assertions in the manner described in Section 7.3. (Recall that such algorithms are not known to exist for other than trivial cases.) We can divide the assumptions of the system into currently accepted assumptions and potential assumptions. The currently accepted assumptions are those from which the system is currently attempting a proof. Corresponding to this view, we would say that the set of formulas accepted as true by the system at any time is the deductive closure of the set of current assumptions.

We can define a trivial two-level system which can be employed by the computer system´s algorithm in exactly the same way we suppose the algorithm to manipulate the given axioms of the system.

We let the axiom set of the initial object theory of the two-level system be empty. The other object theories are defined to be the deductive closures of all subsets of the computer system's axiom set. The metatheory simply defines the predicate A_p to be provable for a term denoting the empty set and also specifies that if $A_p(s)$ is provable for a term s denoting some set of wffs S and if α is a member of the computer system's axiom set, then $A_p(ad(s,\alpha'))$ can be inferred. Thus, A_p will be provable for every subset of the computer system's axiom set. It is easy to see that such a two-level system can be defined and that α is provable from a consistent subset of the computer system's axioms just if it is a theorem of the object theory whose axioms are that same subset. Since we assume that the given system's algorithm never constructs an inconsistent set of current assumptions, the sets of formulas which could be accepted as true by the system would be just the deductive closures of the consistent possible axiom sets. Here, we have simply made use of the algorithm's assumed ability to always choose a consistent subset of the system's axioms and noted that since the algorithm is assumed to employ at any time only a proper subset of the given set of axioms, we can treat each such subset as the axioms of a separate theory.

The important point here is not the particular two-level system whose definition we have just sketched. After all, it relies on an algorithm which may not exist.

Rather, it is the idea that the presence of contradictory assertions in a computer reasoning system need not mean that the system represents some unusual form of inference. It is perfectly possible to define a two level system including object theories whose axioms, if taken together, would be inconsistent. If it is useful to construct such computer systems, then it seems much more reasonable to view them in terms of a two-level system.

8. Summary and Conclusions

We have seen that the definitions of both default and nonmonotonic theories are based on three hypotheses about the nature of default reasoning. Let us call these the nonstandard inference hypothesis, the consistency hypothesis, and the nonnormality hypothesis. Examination of these hypotheses and comparison to alternatives showed that they were not intuitively compelling. The definition of a two-level system is, on the other hand, based on the more intuitive, alternative hypotheses that we have considered. In subsequent chapters we compared the notion of a two-level system to the notions of a default theory and a nonmonotonic theory, presenting evidence that a two-level system constitutes as good a model of default reasoning as does either a default or a nonmonotonic theory. Thus, one conclusion that we draw is that the unintuitive hypotheses about default reasoning in the definitions of default and nonmonotonic theories do not result in a more suitable formal model.

Default and nonmonotonic theories can also be nonmonotonic whereas a two-level system is not. Again, since a two-level system serves at least as well to formalize default reasoning as the other two models, we conclude that there is no reason to suppose that default reasoning is inherently nonmonotonic.

A third conclusion is related to the notion of rules for default reasoning. Examples of informal default reasoning, as well as our observations in developing an informal theory of the process, show that it is natural to view default reasoning in terms of some sort of rules for justifying the default assumptions. In [R], [MD], and [D], the authors continually speak in terms of such rules and discuss their models as though they actually included such rules. However, neither model can represent a well defined notion of a rule, but a two-level system can. Thus, there is no reason to think that the lack of rules in default or nonmonotonic theories indicates anything about default reasoning itself. In fact, in the case of default theories, we were able to see that dropping the nonstandard inference and nonnormality hypotheses in favor of the alternatives suggested by us allowed the definition of rules as well as removing nonmonotonicity. This result suggests that both nonmonotonicity and the absence of rules are directly related to these two particular hypotheses.

Both default and nonmonotonic theories involve consistency as a condition on default assumptions. The informal notion of rule discussed by the authors of both these models involves consistency as well. Thus, even if one of these models could represent rules, the rules that the authors have in mind are not effective. Indeed we saw that although a two-level system equivalent to a closed normal default theory contained well defined rules at the metalevel, those rules were not effective. We also observed that although consistency does not seem to be a sufficient condition for introducing a default assumption, it does appear to be necessary. Thus, the question arises as to whether there can be effective rules for justifying default assumptions. In fact, it is easy to see that there are two-level systems based on consistency which are such that the axioms of their metatheories represent effective rules. In such a system there will be an effective procedure for enumerating the theorems of the object theories. To demonstrate this fact we can use the two-level system that is equivalent to a particular closed normal default theory.

The decision problem for classes of first-order formulas can be stated as: Given a class of formulas, is there a procedure for deciding whether or not a formula in the class is satisfiable? The fact that there are classes of formulas for which a decision procedure exists allows us to give an example of the desired type of two-level system.

Consider the two-level system generated by a closed normal default theory (D,W). Recall that each default of D leads to a corresponding set of axiom schemas in the metatheory, each set containing a schema for each natural number n. Let us assume these sets are given some order and call the jth set d_j. Also, let us call the wff names α' and β', which occur in each member of d_j and are determined by the corresponding default, α'_j and β'_j. Given a metalanguage term t that denotes a possible axiom set, if we wish to show that $ad(t,\beta'_j)$ also denotes a possible axiom set, we must have $\sim Pr(t,\sim\beta'_j)$ as a meta-axiom. Suppose that W is finite. Then $\sim Pr(t,\sim\beta'_j)$ has the interpretation $\gamma_1,\ldots,\gamma_k \nvdash \sim\beta$ where $\gamma_1,\ldots,\gamma_k$ are the members of the set denoted by t. This is equivalent to $\nvdash \gamma_1 \& \ldots \& \gamma_k \rightarrow \sim\beta_j$ and this last formula is not provable just if $(\gamma_1 \& \ldots \& \gamma_k)$ & β_j is satisfiable since the language L over which (D,W) is defined is chosen by Reiter to be first order. Thus, each default and each possible axiom set lead to a formula which must be satisfiable in order for the meta-axiom corresponding to the default to be applied to a term denoting the possible axiom set. Let us suppose that there is a decision procedure for the set S of all such formulas for the two-level system generated by (D,W). Under these conditions we can give a procedure for enumerating the members of all extensions of (D,W).

Theorem 8.1

Suppose (D,W) and S satisfy the above conditions. Suppose also that Σ is the two-level system corresponding to (D,W). Then there is a procedure for enumerating the theorems of the object theories of Σ.

Proof:

Let us call the possible axiom sets determined by (D,W) the A-sets. Let us call the jth theorem in an enumeration of the theorems of a set of axioms, A, t_A^j. We define a procedure as follows:

Maintain the following lists:

L, a list of the A-sets enumerated so far;

For each A_j on L, L_j, a list of the theorems of A_j enumerated so far.

```
$A_1 = W$
put $A_1$ on L
put $t_{A_1}^1$ on $L_1$
for k = 1 to $\infty$ do
  for each $(\alpha_j,\beta_j)$ , $j \leq k$ do
    for each i such that $A_i$ is on L do
      if $\alpha_j$ is on $L_i$ then
        if $\beta_j$ is consistent with $A_i$ then
```

```
          let $A_m$ be the last element of L
          $A_{m+1} = A_i \sqcup \{\beta_j\}$
          put $A_{m+1}$ on L
        end
      end
    end
  end
  for each i such that $A_i$ is on L do
    Let $t^n_{A_i}$ be the last element of $L_i$
    put $t^{n+1}_{A_i}$ on $L_i$
  end
end
```

By our assumptions we can test the consistency of β_j and A_i by determining whether the appropriate wff is satisfiable or not. It is easy to show by induction on the length of L that if A is on L, then A is an A-set.

Suppose A is an A-set. Then $A = A_1 \sqcup \ldots \sqcup A_h$ where $A_1 = W$ and $A_{i+1} = A_i \sqcup \{\beta\}$ for i = 1 to h-1 where A_i is consistent with β and for some α $A_i \vdash \alpha$ and for some m $\alpha = \alpha_m$ and $\beta = \beta_m$.

W is on L. Suppose A_i is on L for $1 \leq i \leq h-1$. Then since $A_i \vdash \alpha_m$ eventually α_m is added to A_i, say as $t^n_{A_i}$. Also, eventually k becomes such that $k \geq m$ and $k \geq n$. Thus, A_{i+1} would be added to L.[]

The above procedure enumerates the theorems of Σ's object theories. It does so by employing the rules of Σ's metatheory that correspond to the defaults of the original default theory. Thus, it is possible to define rules based on consistency that are effective. Systems employing such rules might prove suitable for the construction of computer reasoning systems for particular applications. However, there is no reason to think that such systems would be sufficient to model default reasoning in general.

Since it seems unlikely in any case that humans check the logical consistency of default assumptions with their current knowledge, one is led to consider the possibility of heuristic rules. In Chapter 7 we have provided evidence that the notion of a two-level system can serve as a framework for modelling heuristic rules as well as serving to model definitions of correctness.

An important part of the study of default reasoning is the problem of techniques for the performance of default reasoning in mechanical systems. The notion of a two-level system appears to offer potential as a model of computer reasoning systems that do default reasoning. A syntactic approach to a model of mechanical systems seems appropriate since machines are syntactic in nature. Our particular model offers the advantage of allowing the definition of effective rules, especially heuristic rules. The following simple result illustrates the sort of information one might hope to gain from a model of computer default reasoning systems.

Theorem 8.2

There is a two-level system Σ such that the problem of deciding whether a given wff is a theorem of some object theory of Σ is NP complete.

Proof:

We let Σ be the two-level system equivalent to a certain closed normal default theory (D,W). The language of the wffs of (D,W) is the language of the propositional calculus. W is the empty set, and D is the set of all defaults of the form $M\alpha/\alpha$ where α is any proposition.

Suppose α is satisfiable. Then since $\emptyset$, the empty set, is the initial possible axiom set of the corresponding two-level system, $\{\alpha\}$ is a possible axiom set. Thus, every satisfiable proposition is a member of a possible axiom set and hence, a member of some extension.

Suppose that α is provable from some possible axiom set. Since each possible axiom set is consistent, α must be satisfiable also. Thus, every member of every extension is satisfiable. The union of the object theories of

Σ is thus the set of all satisfiable propositions. Hence, the problem of deciding whether α is a theorem of an object theory of Σ is just the problem of deciding whether α is satisfiable, which is an NP complete problem.[]

Throughout this study we have discussed only the introduction of assumptions. What about a reasoning process that involves deleting assumptions? From our point of view there are two types of assumptions: the initial assumptions of the system and the default assumptions introduced during the reasoning process. Deleting an initial assumption, like introducing a new initial assumption, simply changes the definition of the system. The other possibility would be to delete a default assumption after it has been introduced while maintaining the same initial assumptions.

For a case like the two-level system generated by a closed normal default theory (where the consistency hypothesis is retained), once a default assumption is introduced there can never be any reason to delete it. In such systems no default assumption can be introduced unless it is consistent with the assumptions made so far. Thus, the assumptions remain consistent throughout. This fact constitutes one of the reasons that mechanical systems based on rules requiring consistency would be desirable. On the other hand, for a system like the one used to model Winograd's memory contents rule, it would be possible to arrive at a set of inconsistent assumptions. There does not seem to be any good solution to this problem. It seems to be the price paid for an effectively computable default inference rule.

In summary, we have presented a formal model that is based on quite different hypotheses about default reasoning from those underlying the two established models. Comparisons show our model to be at least as promising as the other two. In fact, our notion of a two-level system has several advantages. It is based on a more intuitive view of default reasoning. It allows the representation of well defined rules for justifying default assumptions. Finally, it allows the modelling of heuristic rules for default reasoning as well as definitions of correctness.

References

[A] Aczel, P., "An Introduction to Inductive Definitions", *Handbook of Mathematical Logic*, Barwise, (ed.), North-Holland Publishing Co., Amsterdam, 1977, pgs. 739-782.

[B] Barwise, J., "An Introduction to First-Order Logic", *Handbook of Mathematical Logic*, Barwise, (ed.), North-Holland Publishing Co., Amsterdam, 1977, pgs. 5-46.

[D] Davis, M., "The Mathematics of Nonmonotonic Reasoning", *Artificial Intelligence*, vol. 13, nos. 1, 2, April, 1980, pgs. 73-80.

[MD] McDermott, D., and Doyle, J., "Nonmonotonic Logic I", *Artificial Intelligence*, vol. 13, nos. 1, 2, April, 1980, pgs. 41-72.

[M] McDermott, D., "Nonmonotonic Logic II: Nonmonotonic Modal Theories", *Journal of the ACM*, vol. 29, no. 1, January, 1982, pgs. 33-57.

[R] Reiter, R., "A Logic for Default Reasoning", *Artificial Intelligence*, vol. 13, nos. 1, 2, April, 1980, pgs. 81-132.

[RC] Reiter, R., and Criscuolo, C., "Some Representational Issues in Default Reasoning", Technical Report 80-7, Computer Science Dept., Univ. of British Columbia, August, 1980.

[We] Weyhrauch, R., "Prolegomena to a Theory of Mechanized Formal Reasoning", *Artificial Intelligence*, vol. 13, nos. 1, 2, April, 1980, pgs. 133-169.

[W] Winograd, T., "Extended Inference Modes in Reasoning by Computer Systems", *Artificial Intelligence*, vol. 13, nos. 1, 2, April, 1980, pgs. 5-26.