THE RELATION BETWEEN CONVENTIONAL INFERENCE AND SOME APPROACHES TO DEFAULT REASONING
(Revised)

by

Kenneth Whitebread

Computer Sciences Technical Report #422

December 1981

The Relation Between Conventional Inference and Some Approaches to Default Reasoning

Kenneth Whitebread

Abstract

Default reasoning has become an important topic of research in Artificial Intelligence. It has been claimed that this process is not understandable in terms of conventional inference. We study three theories of default reasoning agreeing with this claim.

A characterization of the default reasoning process as an alternating sequence of executions of proof procedures for a conventional metatheory and a related set of object theories is presented in the form of a formal machine definition. We then prove some results showing the relation of systems stemming from the three theories considered to our characterization of default reasoning. These results are used as evidence to justify the hypothesis that our machine definition represents a generalization of those systems defined by the theories which can actually be mechanized.

We also show that the property of nonmonotonicity is, in a certain sense, not required for default reasoning as it is characterized by these theories. Some results are given which state conditions for the mechanizability of some of the systems studied.

Contents

1. Introduction

The attempt to develop a suitable theory of default reasoning and to build mechanical reasoning systems capable of doing default reasoning has become an important area of research in Artificial Intelligence. The fact that default reasoning is concerned with introducing new assumptions during the reasoning process while the assumptions of a formal theory are fixed by the definition of the particular theory, suggests that default reasoning cannot be explained by the notion of inference in a formal theory. Several investigators have concluded that default reasoning is not explicable in terms of the conventional concepts of logic. In particular, it has been claimed (see [8] or [3] for example) that a system for default reasoning will have a property called nonmonotonicity which ordinary formal theories do not have.

The purpose of this paper is to present a formal definition of a reasoning system based on conventional inference and argue that the system thus defined subsumes several approaches to default reasoning all of which appear to involve nonmonotonicity. We begin in section 2 with a discussion of default reasoning. In section 3 we give the motivation for our definition and the definition itself which characterizes a class of systems that we call two-level systems. In section 4 we describe several approaches to default reasoning, specifically, those of McDermott and Doyle [5], Reiter [6], and Winograd [8] while in section 5 we analyze these approaches in terms of the notion of a two-level system.

The approaches of [5] and [6] involve formally defined systems. We show that two-level systems can be defined that are equivalent in a reasonable sense to these systems. Winograd's characterization yields several principles each of which is the basis of a form of heuristic default reasoning rule. For each of these principles we consider a heuristic rule exemplifying the principle. We present a two-level system and show that it has the properties to be expected of a model of a reasoning system based on the given rule.

In section 6 we consider the problem of nonmonotonicity. We argue that the systems of [5], [6], and [8] are not in fact non-monotonic. To support our claim we analyze the behavior of the two most important two-level systems defined in section 5, showing the precise sense in which they are monotonic. We also show that there are mechanizable two-level systems that are equivalent to certain examples of the systems defined in [6]. Conditions for mechanizability are not given in [6]. Finally, we show that there are among these mechanizable systems examples for which no efficient algorithm is known.

2. The Process of Default Reasoning

In the study of logic, a formal system or formal theory is defined as a set of inference rules and a set of axioms specified in a formal language. A statement in the language is then a theorem of the formal theory if it can be proved by the inference rules from the axioms of the theory. Although human reasoning may not proceed by constructing proofs, the statements which a human would consider to be the logical consequences of a set of assumptions seem to be well modelled by the theorems of a formal system using those assumptions as axioms. However, humans seem to frequently introduce new assumptions during the reasoning process, a procedure which is not allowed for in the definition of a formal theory. A statement, determined by some process to be plausible or reasonable, is assumed true and treated as such until and unless discovered to be false. Such a statement may then be used either as the basis for making further reasonable assumptions or as the basis for making logical inferences. Yet the statements generated by such means are not necessarily true and sometimes must be discarded when additional facts are learned.

For example, if our car ran properly when we used it yesterday, we expect 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 our 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 ran 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 introduction of this assumption may be incomplete. Some condition which makes the car inoperable may actually be true and simply unknown to us.

Another example of the idea of introducing new assumptions 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 in the same way as the previous one. Since most birds fly and since we don't know anything to the contrary, 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 since 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. The process of introducing new assumptions in this way is called default reasoning because of the observation that the assumption is dependent on the absence of certain information. The assumptions thus introduced are called default assumptions.

It appears that in general, humans combine 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 which are not among those currently known but which can be inferred. One might also infer statements from those currently known or from default assumptions.

The theories of default reasoning which have previously been proposed treat this process as one which begins with a set of axioms and from these generates assertions whose justification may involve any combination of default assumptions and statements inferred by conventional inference rules. The reasoning systems defined by these theories all share a property which is not true of conventional formal theories. In general, if we modify an example of one of these systems by adding an axiom to the originally specified set, we may find that a statement which could be justified on the basis of the original axiom set cannot be justified from the extended axiom set. For a formal theory, the statements which can be justified on the basis of the axioms are, of course, the theorems of the theory. In contrast to the systems for default reasoning, if a statement is provable from the axiom set of a formal theory, that statement is also provable from any extension of the axiom set.

A reasoning system which is such that a statement justifiable from a given set of axioms may not be justifiable from an extension of that set is said to be nonmonotonic. The term is meant to contrast the behavior of such systems with the behavior of a formal theory which can be said to be monotonic in the sense that the set of theorems never decreases as the axiom set is extended. Nonmonotonicity is an important and unfortunate property of the reasoning systems introduced so far as models of default reasoning. Intuitively, adding new beliefs to those which one has previously held should not have the effect of making it impossible to justify an assertion which previously could be justified. In a later section we will argue that nonmontonicity is a fictitious property for the systems of [5], [6], and [8].

3. Motivation and Definition of Two-Level Systems

In this section we will introduce the notion of a two-level system. To motivate the definition we begin by considering the problem we face in attempting to account for default reasoning in terms of conventional inference.

3.1 Motivation

Default reasoning can be viewed in terms of the set of assumptions held by the reasoner at any point during the process. Before any default assumptions are introduced, the reasoner holds some assumptions which he is certain are true. At any time thereafter, the assumptions held by the reasoner include these initial assumptions as well as any default assumptions generated up to that point.

We clearly cannot view default reasoning as an inference process within a single formal theory because the process introduces new assumptions whereas the assumptions of a formal theory, the axioms, are fixed by definition. Our problem then, is that default assumptions must be viewed as axioms if we wish to account for them within the framework provided by the notion of a formal theory, but default assumptions are introduced during the reasoning process while the axioms of a formal theory are specified before any reasoning is done within the formal theory.

If we add an axiom to a formal theory, we have defined a new formal theory, an extension of the original one. Thus, if we consider the assumptions with which the default reasoning process begins as the axioms of a formal theory, then the first time a new assumption is introduced, a new formal theory has been created. Furthermore, if the process goes on to generate statements by conventional inference using the new assumption as well as the orignal ones, the statements thus generated are theorems of the second theory but possibly not of the first.

The above considerations suggest that we formulate our definition in terms of multiple formal theories. Each set of assumptions which may be held by the reasoner could be represented as the axioms of a formal theory (we would assume the same rules of inference for all theories and that all assumptions and any statements generated from them could be expressed in a formal language). The introduction of a default assumption would then be equivalent to introducing a new formal theory. Thus, we would account for the introduction of default assumptions by postulating that the reasoner reasons in more than one formal theory. Reasoning would begin in a theory whose axioms consist of the reasoner's initial assumptions. Each time a default assumption was introduced, the reasoner would, in effect, cease to reason in the theory in which he had been operating and begin reasoning in the extension of that theory determined by adding to its axioms the newly introduced default assumption. However, such a formulation would not account for the process of justifying the introduction of default assumptions. To do this, we will take the idea of reasoning in more than one theory a step further.

Let us return to our example of default reasoning about a car. Consider the statement, "Since the car ran yesterday and since we do not know any information to the contrary, it is reasonable to assume the car will run today". This statement is our justification for assuming the car will run today. We do not use it to deduce the sentence, "The car will run today", but we do use it to deduce an assertion about this sentence. We are, in effect, asserting that among the statements we accept as true is the sentence, "The car ran yesterday". We are also asserting that we are unable to conclude from the statements we accept as true that the car will not start today. From these two assertions about our current beliefs and the statement giving conditions under which it is reasonable to assume the car will run we deduce the assertion, ""The car will run today", is a reasonable assumption", and therefore, we assume the car will run today. Thus, it is reasonable to hypothesize that we justify the introduction of a default assumption by deducing an assertion about it. Such an assertion would have to be in a metalanguage for the language in which the default assumption was stated. Therefore, our definition of a default reasoning system based on conventional inference will be formulated in terms of multiple formal theories in two formal languages.

Our view of default reasoning is that the reasoner maintains two sets of assumptions during the reasoning process. One set, which we call the set of current assumptions may be replaced at any time during the process by a new set consisting of the union of the old set and one new assumption. This set of assumptions represents those beliefs about which the reasoner can reflect at any time for the purpose of introducing new beliefs. The other set of assumptions does not change during the process and represents the reasoner's rules for introducing new beliefs.

The definition of a two-level system reflects our notion of default reasoning. An instance of two-level system will be a collection of theories in two languages. All but one of these theories will be defined in one language and will represent the various sets of assumptions which the reasoner could hold as current at different stages of the reasoning process including a set representing the initial beliefs about which he can reason. One theory will be defined in the second language. This theory is intended to allow us to reason about sets of assumptions in the other language. It should allow us to assert that a given assumption set A and a default assumption α which could be introduced if the current assumptions were those of A form a new current assumption set $A \sqcup \{\alpha\}$.

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

3.2 Definition of a Two-Level System

Let us now turn to our definition. We first give a definition of a concept which we will employ 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, [1]). A structure for the language L consists of a domain of objects and a collection of mappings assiging 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 [1]):

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 3.1

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 <u>element</u> $\hat{t}$. We will then say that t <u>denotes</u> the element $\hat{t}$.

A <u>two-level</u> system consists of:

1. A set <u>S</u> of sets of wffs in a language L called the <u>possible axiom sets</u>. One possible axiom set is distinguished as the <u>initial axiom set</u>. All other possible axiom sets must be supersets of the initial possible axiom set. The set of theories generated by by the possible axiom sets is called the set of <u>object theories</u>.

2. A theory in a language L´ called the <u>metatheory</u>.

3. A structure for L´ called the <u>intended interpretation</u> 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 $A_p(t)$ 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. 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 the possible axiom sets through the presence of 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 intuitve picture as well as some which do not. We choose the form of definition given both because it is simple and for its generality.

The use of metatheories and their accompanying object theories in computer reasoning systems has been investigated by

Kowalski [4] and Weyhrauch [7]. Our definition of a two-level system was developed independently of their work and makes use of the notion of defining a set of (possibly contradictory) theories in the object language whereas the work of both Kowalski and Weyhrauch is based on the more usual idea of a metatheory-object theory pair. Our definition also serves a somewhat different purpose. Kowalski and Weyhrauch are concerned with constructing actual computer systems while we are concerned with investigating the relation to logic of certain concepts from artificial intelligence.

4. Three Theories of Default Reasoning

Let us now turn to a description of the three theories of default reasoning which we will analyze in terms of our model. These three approaches are due to Winograd [8], Reiter [6], and McDermott and Doyle [5].

4.1 Heuristically Based Default Reasoning Systems

Winograd's paper 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 results in a hypothesis by Winograd that computer systems can be devised to perform certain types of inference which he claims are not formalizable in standard logical terms. Four categories of procedures for performing such 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 inference procedure fails to prove a certain formula.

We will now examine the definitions of these categories to determine their relationship to the notion of default reasoning and to isolate for each of them the reason behind the claim that it represents a form of inference which cannot be accounted for in terms of conventional inference.

4.1.1

In order to understand the definition of the first category we will 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 be likely to 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 which is such that we would be likely to be aware of its truth if it were true, not being aware of the assertion's truth represents absence of information 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 problem 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 be likely to 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 to define the set of assertions of whose truth the system is aware at any point during a computation would be to define it to consist of those assertions occuring in the system's memory at that point. The set would therefore consist of those assertions which had been assumed true initially or through the default rule along with those assertions which had previously been 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 inferred by the system and those which are logical consequences but have not been inferred.

There is no way to distinguish within a formal theory 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. The definition of the first category is a generalization of the idea of defining the assertions which a reasoning system knows to be those which occur in memory. It does not address the other problem mentioned above.

As we have just seen, the problem with including the notion of the set of assertions known by a reasoning system in a system which reasons in a formal theory is that the distinction between already proven and unproven theorems cannot be made within the theory. Therefore, any rule making use of this distinction cannot be part of the 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 notion of generating an assertion as the result of the presence of certain formulas in memory need not concern us. The formulas in memory at any time are those which the system has already accepted as true. The procedure generating an assertion from these 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.

4.1.2

The definition of the second category is motivated by the observation that after we have made a certain amount of effort to infer an assertion and failed we often decide that the assertion is likely to be false. If, having arrived in this way at the conclusion that an assertion is likely to be false, we then assume the negation of the assertion, we have actually 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 prove or infer an assertion would just be the execution of some procedure which 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 the next section we will show how such a rule can be represented in a two-level system.

4.1.3

The third category is illustrated by the previously given 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 indiviual 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 inconsistent assumptions in the system.

The method followed by the program in attempting to prove an assertion 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 those sets of 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 axioms. It is not known whether such procedures exist for other than trivial cases.

In the next section we will 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 system.

4.1.4

The definitions of these categories exemplify Winograd's contention that mechanical reasoning systems, to be successful, must use some knowledge about their own inference procedures and must take into account the fact that a practical system cannot in any reasonable sense be viewed as "knowing" all the statements which follow from its assumptions. Such contentions are quite reasonable. We wish to argue, however, that it is possible to define reasoning systems based on conventional inference in formal theories which display such properties.

4.2 Default Theories

The second approach to be considered is that of Reiter and is based on the definition of what he calls a default theory. 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 expressions called 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 <u>closed</u>. Since Reiter deals mainly with closed default theories and treats nonclosed theories by relating each one to a certain closed theory, we will consider only those which are closed.

From the definition of default theory we make the definition of the <u>extensions</u> 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.

The formulas of W represent the initial assumptions of the system while the members of D are intended to represent rules for introducing new (default) assumptions. The intuitive interpretation of $\alpha:M\alpha_1,\ldots,M\alpha_k/\beta$ is that if α can be inferred from "what is known" and $\alpha_1,\ldots,\alpha_k$ are consistent with "what is known" then α can be assumed. However, the phrase "what is known" presents an obvious difficulty in making this interpretation precise. Reiter skirts this problem by introducing the definition of an extension of a default theory.

An extension is any fixed point of the operator Γ. Examining the definition of Γ, we see that an extension E is deductively closed and includes W, the initial assumptions. In addition E contains various default assumptions. If $\alpha:M\alpha_1,\ldots,M\alpha_k/\beta \in D$, $\alpha \in E$, and $\alpha_1,\ldots,\alpha_k$ are consistent with E, then $\beta \in E$ also.

Intuitvely we think of the reasoner as beginning with the initial assumptions and applying both conventional inference rules and "default inference rules". The result would be a set of wffs closed under both ordinary and default inference rules just as the result of applying ordinary inference rules alone to a set of axioms is a deductively closed set. Unfortunately, defaults cannot actually be used as the default inference rules suggested by intuition.

An extension is intended to be the closed set of formulas which would result if we could use defaults as default inference rules. 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.

According to Reiter an extension is to be viewed as "an acceptable set of beliefs that one may hold about the incompletely specified world W". Reiter calls a sentence which is a member of some extension believable. Thus, given a wff and a default theory, one question we would like to be able to answer is whether the wff is a member of some extension of the default theory. We will show in the next section that for any default theory a two-level system can be defined such that a wff is a member of an extension just if it is provable from some possible axiom set of the two-level system.

4.3 Nonmonotonic Theories

Our last example of a theory for default reasoning is due to McDermott and Doyle. In this case, we begin with a special 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

$$As_A(S) = \{M\beta : \sim\beta \notin S\} - Th(A);$$
$$NM_A(S) = Th(A \sqcup As_A(S)).$$

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

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

Finally, we define the set of wffs TH(A) to be the intersection of all sets in the class FP if FP is not empty and L_M if FP 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. Intuitively, it can be thought of as the set of default assumptions generated by the wffs of S and the given initial assumptions of A. Here, the idea is to introduce $M\alpha$ if α is a reasonable assumption rather than introducing α itself. 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 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. We will discuss this difference between default theories and nonmonotonic theories in the next section.

One purpose behind the introduction of the symbol M into the language in which reasoning is to be done is to allow assertions of the form: $MM\alpha$. Such an assertion would mean intuitively that it is reasonable to assert that it is reasonable to assert α. Such statements cannot be made in default theories yet it seems plausible that humans may do such default reasoning about default reasoning.

Unfortunately, as Davis [2] points out, the definition of a model theory for nonmonotonic theories given by McDermott and Doyle does not actually explain how to give a meaning to formulas of the form $M\alpha$. Thus, though syntactically well defined, nonmonotonic theories have no clear interpretation. We consider them in this paper because they represent the only attempt known to us to define a system which allows default reasoning about default reasoning. Furthermore, it may be possible to give a clear interpretation either to the syntactic definition as it stands or to a somewhat restricted form. The intended meaning of the assertion $M\alpha$ is that it is reasonable to assert α which is the same as saying that it is reasonable to assume α. Thus, the intended meaning of M is similar to our notion of a predicate in the metatheory of a two-level system which is provable of a set of object language formulas just if it is reasonable to assume those formulas simultaneously. This similarity suggests that one way to approach the problem of a formally interpretable version of the notion of nonmonotonic theory would be by defining, instead of a two-level system, say, an n-level system where each pair of levels corresponds to a two-level system.

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

We now turn to the relation between two-level systems and the theories of default reasoning introduced in the previous section. Each of these theories will be examined in turn. Our general approach will be to define in each case a two-level system whose object theories bear a certain relation to the sets of "believable" formulas determined by an arbitrary example of the systems defined by the the default reasoning theory under consideration. The exact nature of this relation will depend on the theory in question, but in each case we will be able to claim that the defined two-level system is equivalent to the given default reasoning system in a reasonable sense.

5.1 Default Theories

We begin with the default theories of Reiter. First, we will define a trivial two-level system for an arbitrary closed default theory to show that conventional logical concepts do account for such systems. We will then turn to a special form of default theory called a normal default theory, discuss its importance and show how to define a more interesting two-level system for closed normal default theories.

5.1.1 Arbitrary Closed Default Theories

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 [6] 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 interpeted 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 and that if β is provable from a possible axiom set S, then for each member of S, say α, occuring in the proof $\alpha \in t$ is provable in the metatheory for closed t such that t denotes S and $A_p(t)$ is provable. 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 statisfied 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 the 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).

5.1.2 Closed Normal Default Theories

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 very satisfying. It allows us to conclude that possible axiom sets are possible axiom sets but only because an axiom asserting A_p for each such set is included. Furthermore, the set of these axioms may be uncountably large. (We also have uncountably many constant symbols, of course.)

It is therefore natural to ask whether a two-level system can be defined such that the object theories correspond to the extensions of a closed default theory, but with a countable axiom set for the metatheory and a more reasonable way of handling A_p. In particular, we would like the metatheory to agree with the intuitive basis for the notion of a two-level system as presented in a previous section. There we viewed default reasoning as a process of adding a default assumption to a set of axioms representing the reasoner's current beliefs. We will see below that this can essentially be done for a special type of closed default theory called a closed normal default theory.

A <u>normal</u> default theory is one in which all defaults are of the form $\alpha:M\beta/\beta$. That is, the only wff which must be consistent with "what is known" is the one to be assumed. Reiter argues in [6] that most natural default rules are of this form. He also 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 occuring in D and let C(D) be the set of consequents. A <u>default</u> 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 for each $\alpha \in P(D_{i+1})$,
   $W \cup C(D_i) \vdash \alpha$;
3. $W \cup C(D_k) \vdash \gamma$;
4. $W \cup C$ is consistent where $C = \cup C(D_i)$, 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. A similar result for arbitrary closed default theories using an analogous definition of default proof fails to hold indicating a shortcoming in the general notion of a default theory.

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 which introduces these assumptions. Thus, we have a notion (extension of a default theory) corresponding to the result of default reasoning but no notion of default reasoning itself. Obviously, defaults were intended to be a sort of inference rule for introducing default assumptions. We will see below, however, that they cannot actually be treated as such in general. On the other hand, the notion of a default proof represents a way of describing a default reasoning process for normal default theories. In the course of defining a two-level system for the normal case we will see that for this case defaults can be treated as inference rules, and it is this fact which underlies Reiter's result on default proofs. Default proofs are easy to account for in the context of a two-level system.

5.1.2.1

Let (D,W) be an arbitrary closed normal default theory in the language L. The following results about (D,W) will be used in defining a two-level system for (D,W) and will show why the defaults of D can be thought of as inference rules for introducing default assumptions.

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 occuring in the jth default.

Let

$$F_0 = W$$
$$F_{i+1} = \bigsqcup F_{i+1,j} \text{ for } j = 0 \text{ to } i+1$$

where

$$F_{i+1,0} = F_i$$
$$F_{i+1,j+1} = F_{i+1,j} \bigsqcup \{\beta_j\} \text{ if } F_{i+1,j} \vdash \alpha_j, \text{ and } \sim\beta_j \notin E$$
$$= F_{i+1,j} \text{ otherwise.}$$

Let $F = \bigsqcup F_i$ for $i = 0$ to ∞.

For our discussion of default theories we will use Th(S) to mean the set of sentences provable from S.

Lemma 5.4

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

Proof:

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

Condition 1

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

Condition 2

Obviously, Th(F) = Th(Th(F)).

Condition 3

Suppose for some member of D, say $\alpha_k : M\beta_k / \beta_k$, $\alpha_k \in$ Th(F) and

$\sim\beta_k \notin E$. Since $\alpha_k \in Th(F)$, 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 F, $\beta_k \in F$. 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 $\beta_k \in F$.

Thus, Th(F) satisfies the three conditions.

We can show $Th(F) \subseteq \Gamma(E)$ by showing by induction that $F \subseteq \Gamma(E)$. 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 $F_{i+1,j+1} = F_{i+1,j} \cup \{\beta_j\}$, $F_{i+1,j} \vdash \alpha_j$, 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, $F \subseteq \Gamma(E)$ and so $Th(F) \subseteq \Gamma(E)$.

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

Consider the class of sets consisting of W and all sets of the form $W \cup \{\alpha_1, \ldots, \alpha_k\}$. 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 $(A \cup \{\beta\}) \in AP$.

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 $A = \sqcup A_i$ for i = 0 to ∞, then E = Th(A).

Proof:

By Lemma 5.4 E = Th(F) where $F = \sqcup F_i$ 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.[]

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

$$E_0 = W$$
$$E_{i+1} = \bigcup E_{i+1,j},\ j = 0 \text{ to } i+1$$

where

$$E_{i+1,0} = E_i$$
$$E_{i+1,j+1} = E_{i+1,j} \cup \{\beta_j\} \text{ if } E_{i+1,j} \vdash \alpha_j,$$
$$\text{and } E_{i+1,j} \nvdash \sim\beta_j$$
$$= E_{i+1,j} \text{ otherwise.}$$

Here we again assume that α_j and β_j are the wffs occuring in the jth default for the given ordering. Let $E = Th(E')$ where $E' = \bigcup E_i$, $i = 0$ to ∞ and let F be defined as above in terms of this set E.

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 F 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 $E = Th(F)$. Therefore, we will show that $E' = F$.

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 $F_{i+1,j+1} = F_{i+1,j} \cup \{\beta_j\}$), 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, $E' = F$.[]

In the definition of the sets E_i given before Lemma 5.6 an ordering of the defaults of D is assumed. It is important to note that the lemma does not tell us that we could construct an extension by simply attempting to apply each default as a rule in the order given by the ordering. That is, we cannot attempt to apply the first rule to W, then attempt to apply the second rule to the set of wffs which results from applying the first rule, and so on. It is easy to see that such a procedure need not result in an extension. However, the definition of the sets E_i induces another ordering, possibly distinct from the assumed ordering. In this ordering, defaults which are actually applied in constructing some E_i come before those which never apply and those defaults which are applied are ordered in the order in which they apply. 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 in the order given by the ordering.

Lemma 5.6 also lets us show in the next theorem that if we treat the defaults as rules, then any set of formulas contstructed from W by applying defaults will be a subset of some extension.

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 above.

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 $\beta \in Th(F)$ by Lemma 5.4 (where F is as in Lemma 5.4). Since the proof of β from F 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 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. However, we cannot do this in general.

Reiter's intuitive description of the meaning of a default employs the phrase "what is known". To actually use a default as a rule for introducing an assumption we must make the definition of this phrase precise. It is natural to try to interpret "what is known" to mean the set of sentences accepted as true at the time the default is applied. Thus, W would initially represent what is known and each time an assumption was introduced we would add the assumption to the set representing what is known. We could then use defaults as rules for introducing assumptions by saying that if $\alpha:M\alpha_1,\ldots M\alpha_k/\beta$ is a member of D, then β may be added to the set of formulas representing what is currently known if α is provable from that set and $\alpha_1,\ldots,\alpha_k$ are consistent with that set. Unfortunately this approach fails.

Suppose that the default $\alpha:M\alpha_1,\ldots,M\alpha_k/\beta$ 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.

For example, let $W = \{P(a)\}$ and $D = \{P(a):MQ(b)/R(a), P(a):MR(a)/\sim Q(b)\}$. 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 defintion of an extension is such that the justification for the occurance in an extension of any default assumption depends on all other default assumptions occuring in the extension. Because of this there does not appear to be any way to treat defaults as rules in the case of an arbitrary default theory.

5.1.2.2

For a closed normal default theory the above results show that we can treat the defaults as rules for introducing assumptions in the manner described above. Because of this Reiter's result concerning default proofs is not surprising. Each default proof corresponds to an initial segment of one of the sequences of sets $\{E_i\}$ defined above in an obvious way. Furthermore, because we can treat the defaults of a closed normal default theory as rules for introducing assumptions, we can define a two-level system for such a default theory where 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 defined above.

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 $adj(x,y) = x \sqcup \{y\}$
else let $adj(x,y) = d$.

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 one of these sets are indeed members. However, we do not wish to 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 $\in$ 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 emmployed 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. $\forall x \forall y \forall z (x \in ad(y,z) \Leftrightarrow (S(y) \ \& \ \sim S(z)) \ \& \ (x \in y \lor x = z))$.

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

⋮

7_n. $Pr(ad(W',\alpha'_1,\ldots,\alpha'_n),\beta')$ for each $\alpha'_1,\ldots,\alpha'_n,\beta'$ 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 $\alpha'_1, \ldots, \alpha'_n, \beta'$ 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. $\forall x_1 \ldots \forall x_n (A_p(ad(W', x_1, \ldots, x_n))$ & $Pr(ad(W', x_1, \ldots, x_n), \alpha')$ & $\sim Pr(ad(W', x_1, \ldots, x_n), \sim\beta) \rightarrow A_p(ad(W', x_1, \ldots, x_n, \beta)))$ 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 two-level system we wish to consider 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 occurances of ad and consider t containing k+1 occurances 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 $\sim S(\alpha)$.

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

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

is satisfied by the structure and similarly

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

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

$$Pr(W',\alpha') \; \& \; \sim Pr(W',\beta') \rightarrow A_p(ad(W',\beta')$$

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

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

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 $A_p(ad(W',\alpha'_1,\ldots,\alpha'_n))$, $Pr(ad(W',\alpha'_1,\ldots,\alpha'_n),\alpha')$, $\sim Pr(ad(W',\alpha'_1,\ldots,\alpha'_n),\beta')$ 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 $ad(W',\alpha'_1,\ldots,\alpha'_k)$. 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. This fact is not obvious from the results of [6].

5.2 Nonmonotonic Theories

For the nonmonotonic theories of McDermott and Doyle we can also supply a trivial two-level system. Recall that

$$As_A(S) = \{M\beta : \neg\beta \notin S\} - Th(A)$$

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

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).

Lemma 5.12

TH(A) is closed under ordinary dedcution.

Proof:

First, note that if $S \in FP(A)$, then S is deductively closed. This is true because $S = Th(A \sqcup As_A(S))$.

Suppose $TH(A) \vdash \alpha$. Then α is provable from some finite subset B of $TH(A)$. By the definition of $TH(A)$ $B \subseteq S$ for all $S \in FP(A)$. Therefore, since for all $S \in FP(A)$ S is deductively closed, $\alpha \in S$ for all such S. Thus, $\alpha \in TH(A)$.[]

Given a nonomonotonic theory $TH(A)$ in the language L_M we define a system Σ. We begin by defining a metalanguage L'_M consisting of:

1. A constant symbol, say α', for each wff $\alpha \in L'_M$;
2. A constant symbol, say S';
3. A unary predicate symbol, say A_p, and a binary predicate symbol, say $\in$.

We next define a structure for L'_M. The domain of the structure consists of the wffs of L_M and the set $TH(A)$. Each symbol α' is interpreted as the corresponding wff α. The symbol S' is interpreted as the set $TH(A)$. The predicate symbol $\in$ is interpreted as set membership while $A_p(x)$ is interpreted to mean that x is $TH(A)$.

We can now define the axioms of Σ's metatheory:

1. $\alpha' \in S'$ for each $\alpha \in TH(A)$;
2. $A_p(S')$.

To complete our definition of Σ we define $TH(A)$ to be the only possible axiom set of Σ. It is easy to show that Σ satisfies the requirements for a two-level system.

Lemma 5.13

Σ is a two-level system.

Proof:

Requirements one and three of the definition of a two-level system are obviously satisfied. For requirements two and four we use the fact that the meta-axioms of Σ are obviously satisfied by the structure and argue as in Lemmas 5.1 and 5.3[]

From Lemma 5.12 we see that Σ is equivalent to TH(A) in the sense that α is a member of Σ's only object theory if and only if α is a member of TH(A). Note that here we have a two-level system with a meaningful metatheory and a meaningless object theory.

As with closed default theories, we introduce the above two-level system simply to show that a nonmonotonic theory can be treated in terms of formal theories. Unlike default theories we do not have a more promising special case to investigate. However, we can take note of one possible variation on the definition of nonmonotonic theory.

In the section 4 we noted that the definition of the class FP(A) is similar to the definition of the class of extensions of a default theory, but where each extension is considered a seperate "possible world", so to speak, a wff of L_M is not considered believable uless it occurs in every member of FP(A). Of these two approaches to the definition of the believable formulas, the treatment of extensions seems to us to more closely reflect the intuitive idea of default reasoning. A natural modification to the definition of a nonmonotonic theory, therefore, would be to discard the definition of TH(A) and instead to consider each member of FP(A) to be a set of assertions which could be accepted as simultaneously true.

In the above outline of a two-level system equivalent to a nonmonotonic theory we used TH(A) as the only object theory and treated it as its own axiom set. In fact there does not seem to be a natural set of axioms for TH(A). This is not true, however, for the inidvdual members of FP(A). Suppose that a set of wffs A is given and suppose that S is a set of wffs with the following properties:

1. $A \subseteq S$;
2. Every wff in S - A is of the form $M\beta$ where β is consistent with S;
3. If γ is a wff such that γ is consistent with S, then $M\gamma \in S$.

As we show below, if S satisfies properties one, two, and three, then $Th(S) \in FP(A)$. On the other hand if $S' \in FP(A)$, then there is a set S satisfying properties one through three such that $S' = Th(S)$. We now show these claims. Given a set A of wffs in L_M define the class C_A of sets of wffs of L_M as follows:

$$C_A = \{S \mid S \text{ satisfies properties 1, 2, and 3}\}.$$

Recall that

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

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

where Th(S) is the set of wffs provable from S. Notice that if we let $Bs_A(S) = \{M\beta : {\sim}\beta \notin S\}$ then $NM_A(S)$ is also equal to $Th(A \sqcup Bs_A(S))$.

Lemma 5.14

Suppose $S \in FP(A)$ and suppose $R \in C_A$. Then

a. $A \sqcup Bs_A(S) \in C_A$.

b. $Th(R) \in FP(A)$.

Proof:

Part a.

Clearly $A \subseteq (A \sqcup Bs_A)$. If $\alpha \in Bs_A$, then $\alpha = M\beta$ such that ${\sim}\beta \notin S$. If ${\sim}\beta \notin S$, then $S \nvdash {\sim}\beta$ since $S = Th(S)$ by the definition of S. Thus, $(A \sqcup Bs_A) \nvdash {\sim}\beta$.

Suppose there exists a β such that $(A \sqcup Bs_A) \nvdash {\sim}\beta$. Since $S = Th(A \sqcup Bs_A)$, $S \nvdash {\sim}\beta$. Thus, ${\sim}\beta \notin S$ and therefore, $M\beta \in Bs_A$ by definition. Hence, $A \sqcup Bs_A \in C_A$.

Part b.

Suppose that $R \in C_A$. To show $Th(R) \in FP(A)$ we must show that $Th(R) = Th(A \sqcup Bs_A(Th(R)))$. We will show $R = A \sqcup Bs_A(Th(R))$. By definition, $Bs_A(Th(R)) = \{M\beta \mid {\sim}\beta \notin Th(R)\}$. But ${\sim}\beta \notin Th(R)$ iff $R \nvdash {\sim}\beta$. Therefore, by the definition of C_A, $M\beta \in R$. Thus, $Bs_A(Th(R)) \subseteq R$, and since $A \subseteq R$, $(A \sqcup Bs_A) \subseteq R$.

If $\alpha \in R$, then either $\alpha \in A$ or $\alpha = M\beta$ for some β such that $R \nvdash {\sim}\beta$. If $\alpha = M\beta$, then, since $R \nvdash {\sim}\beta$, $M\beta \in Bs_A(Th(R))$. It follows that $R \subseteq (A \sqcup Bs_A(Th(R)))$. Thus, $R = A \sqcup Bs_A(Th(R))$.[]

Thus, a member of FP(A) can be thought of as the deductive closure of a set consisting of the union of an initial set of axioms and a set of default assumptions. Of course, the problem of interpreting the symbol M remains.

5.3 Systems Based on Heuristic Rules

Let us now turn to the systems characterized by Winograd's categories. We will 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 considered by Winograd can be thought of as formal languages, but the significant characteristics of the heuristic default rules being considered do not depend on the choice of language. We will 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 study.

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 will 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. 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 or 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 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 or state of the system's memory at successive stages of the computation. Let us call such a sequence a <u>memory-state sequence</u>. 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 therefore 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 a two-level system 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.

5.3.1 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.

5.3.1.1

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 would already be 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. However, it is the notion of introducing an assumption because of the absence of some formula from memory which Winograd is presenting as being outside the concepts of conventional logic, and this contention is not dependent on a realistic approach to handling PA.

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 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 the class 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 the class 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 (S,R) belonging to the class 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 (S,R) belonging to the class 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.

We will see that sets belonging to M can be thought of as representing the states of memory which could occur during a computation. We will also see 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 5.15

$S \in M$ 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 definiton of M are satisfied and $S \in M$.[]

Lemma 5.16

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 α such $S = \hat{S} \sqcup \{\alpha\}$, 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 5.17

$S \in M$ 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 memeber.

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 5.7

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 5.17, $S \in M$. By Lemma 5.15 there is R such that $(S,R) \in MA$. By Lemma 5.16 $Th(S) = Th(R)$.[]

Theorem 5.8

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 5.15, $S \in M$. By Lemma 5.17, S is a member of a memory-state sequence. By Lemma 5.16 $Th(R) = Th(S)$ and $R \subseteq S$.[]

5.3.1.2

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 $n \geq 1$ and each $k \leq n$,

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

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

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

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

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 $n \geq 1$ and each $k \leq n$,

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

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

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

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

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

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

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

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

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

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 above.

Lemma 5.18

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 5.19

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 5.20

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 5.21

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 5.22

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 $M(ad(I',a_1,\ldots,a_n))$, $MA(ad(I',a_1,\ldots,a_n), ad(I',b_1,\ldots,b_k))$ 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 5.23

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 5.24

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 5.21 and 5.23 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 5.25

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 5.15 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 5.7

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:

If $S \in AP$, then there is R such that $(R,S) \in MA$. By Lemma 5.15 $R \in M$. Therefore, by Lemma 5.24 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 above.

Theorem 5.8

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

Theorem 5.9

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.

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 by conventional logical concepts and therefore does not represent some sort of extended mode of inference.

5.3.2 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 presentation of a recursive function which we may also call f. Thus, the above rule as it would be impelemented 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" and $f(\alpha,B)$ = "no" for each instance of α and A such that f would return "yes" and 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 simlilar to the two-level system defined for the first category.

5.3.3 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. Let us suppose we have a computer reasoning system employing such assertions and relying on an algorithm which handles these assertions in the manner described in section 4. We cannot treat such a system in exactly the same terms as the previous two cases because in this case it is not reasonable to say that all formulas in memory at any time are accepted as true by the system. The system's initial axiom set is itself inconsistent, and it is not reasonable to suppose that the system (or a human) somehow accepts simultaneously assertions which obviously contradict each other. Instead 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 seperate theory.

6. Nonmonotonicity and Other Properties

It is often assumed that any system for default reasoning will be nonmonotonic. This means that the system will be such that if we add an assumption to a given example of the system, thereby creating a second example, it may be the case that a formula which could be inferred in the first example cannot be inferred in the second example. Default theories and nonmonotonic theories are both claimed to be nonmonotonic as are the systems using the heuristic rules discussed by Winograd. We will argue here that the claimed nonmonotonicity of these systems is the results of a misconception about what constitutes the systems' axiom sets. We will also consider the problem of the recursive enumerability of default reasoning systems.

6.1 Nonmonotonicity

If we are to consider whether a given reasoning system is nonmonotonic, we must first decide what constitutes the assumptions or axioms of the system. A two-level system consists of a collection of formal theories each with its own set of axioms. The meta-axioms represent a set of assumptions which remain in force during the reasoning process while each possible axiom set represents assumptions which may be accepted by the reasoner at some point during the process. A possible axiom set consists of the union of a set of initial axioms and a set of default assumptions. Since the meta-axioms and the initial object level axioms remain in force throughout the reasoning process they must certainly be considered among the axioms of the system. However, the default assumptions in force at any point during the reasoning process are also axioms of the system at that point. They are, after all, formulas representing assertions which the reasoner accepts as true at that point but which have not been proved from other axioms. We cannot say that the axioms of a two-level system consist of the meta-axioms, the initial object axioms, and all default assumptions which could be introduced because then we would have an inconsistent system. In fact, it does not make sense to speak of "the axioms" of a two-level system. A two-level system is a collection of axiom sets and the property of nonmonotonicity must be considered seperately for each of these sets.

Of course, the theories of a two-level system are monotonic by definition so the question of nonmonotonicity is trivially answered. However, it is useful to examine some examples of two-level systems more closely in order to better understand the apparent nonmonotonicity of some of the systems mentioned above. As we have already seen, the general default theories and nonmonotonic theories are not very interesting systems. We therefore examine two-level systems for a closed normal default theory and our example of Winograd's "memory contents" principle.

6.1.1

In the case of Reiter's approach 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. This possibility also exists for closed normal default theories. Both D and W are viewed as representing the axioms of (D,W). Thus, the default theories (D,W) and (C,V) appear to be nonmonotonically related. We will argue that the interpretations intended for (D,W) and (C,V) involve additonal axioms which are not made explicit.

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 $\alpha:M\beta/\beta$.

Theorem 6.1

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 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. This axiom also represents an assumption of the reasoner, one that he must be making if he is going to apply rules of the form: If α is consistent with what I know... 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.

The second part of the result states that it cannot be the case that every object theory of Σ is a subtheory of an object theory of Σ'. If this were the case, every object-level theorem of Σ would be an object-level theorem of Σ'.

We argue that an assumption about what is initially known is implicitly present in the intuitive interpretation of a default as stated by Reiter. If $\alpha{:}M\beta/\beta$ is to mean: Assume 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 making some assumption about what it is that he knows. We have shown in section 5 that for a closed normal default theory "what is known" can be identified initially with the inital 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 assumptions concerning what is known. Although these assumptions 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.

There is a second sense in which the assumptions of (D,W) differ from those of (C,V). Our analysis of closed normal default theories in section 5 shows that (D,W) and (C,V) each represent a collection of sets of assumptions just as a two-level system does. The second part of the above result tells us that at least one of these sets for (D,W) is not a subset of any such set for (C,V) since these sets correspond to the possible axiom sets of Σ and Σ'.

Thus, we argue that the nonmonotonic relation of (D,W) and (C,V) is fictitious for two reasons. First, because (D,W) involves an implicit assumption about what is known, an assumption which is not in force in (C,V). Second, a default theory actually represents a collection of sets of assumptions. The assumption set of (D,W) which entails the wff α that cannot be derived in (C,V) is in fact not a subset of any assumption set of (C,V).

6.1.2

In section 5 we considered a hypothetical computer system employing a heuristic default reasoning rule based on Winograd's "memory contents" principle. Our definition of a memory-state sequence is such that we can obviously define the provable formulas of the system to be just those which 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 6.1

Theorem 6.2

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 6.1.[]

6.2 Recursive Enumerability for Default Reasoning Systems

Obviously, for any default theory (D,W) and wff α for which there is an extension of (D,W) containing α, we would like to have an algorithm which, given α, determines that there is an extension for it. It is easy to see that there are default theories for which no algorithm exists. However, by considering our analysis of closed normal default theories in section 5 we can give conditions under which an algorithm exists.

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 one criterion for the existence of an algorithm for closed normal (D,W).

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 as one of the conditions for applying a meta-axiom corrseponding to the default 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 6.3

Suppose (D,W) and S satisfy the above conditions. Then there is a procedure for enumerating the members of all extensions of (D,W).

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 enumereated 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^1_{A_1} on L_1
for k = 1 to ∞ do
  for each (α_j,β_j) , j ≤ k do
    for each i such that A_i is on L do
      if α_j is on L_i then
        if β_j is consistent with A_i then
          let A_m be the last element of L
          A_{m+1} = A_i ⊔ {β_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.[]

Since a formula is a member of an extension if and only if it is provable from some possible axiom set the above procedure enumerates all members of all extensions. Hence there is a procedure which, given a member of some extension, will determine that the formula is indeed a member of an extension. However, it is easy to see that there are default theories for which no good algorithm is known.

Theorem 6.4

There is a closed normal default theory for which the problem of deciding whether a given wff is a member of some extension is NP complete.

Proof:

We define a 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 proposistion 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 extensions of (D,W) is thus the set of all satisfiable propositions. Hence, the problem of deciding whether α is a member of an extension of (D,W) is just the problem of deciding whether α is satisfiable which is an NP complete problem.[]

7. Conclusion

7.1 Deleting Assumptions

Throughout this paper we have only discussed the introduction of assumptions. What about 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, 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. 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.

7.2 Summary

Default reasoning has been considered to be a process which cannot be understood in terms of the idea of inference by conventional rules of inference in an ordinary formal theory. Several approaches to default reasoning have been put forward which do not appear to be explicable in terms of the concepts of conventional logic. In particular these approaches appear to be non-monotonic.

We have introduced in this paper a definition for a reasoning system, called a two-level system, based on conventional inference and provided evidence that two-level systems subsume several apparently nonmonotonic approaches to default reasoning. Using our characterizations of these systems in terms of two-level systems we have argued that their claimed nonmonotonicity was fictitious. Finally, we have shown the possibility of mechanizable versions of certain of these systems.

References

1. Barwise, J., (ed.), Handbook of Mathematical Logic, North-Holland Publishing Co., Amsterdam, 1977.

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

3. Hayes, P. J., "The Logic of Frames", unpublished paper, 1977.

4. Kowalski, R., Logic For Problem Solving, Elsevier North Holland, New York, N.Y., 1979.

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

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

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

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