TR-336

A Model Theory for Hypothetical Reasoning

by

K. Inoue

January, 1988

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

A Model Theory for Hypothetical Reasoning

Katsumi Inoue

ICOT Research Center,
1-4-28, Mita, Minato-ku, Tokyo 108, Japan
phone: +81-3-456-3192
telex: ICOT J 32964
csnet: inoue%icot.jp@relay.cs.net
uucp: {enea,inria,kddlab,mit-eddie,ukc}!icot!inoue

ICOT Technical Report TR-336
January 1988

Abstract

This paper describes a logical framework for hypothetical reasoning. Hypothetical reasoning is a kind of non-monotonic reasoning, and it is desirable when dealing with incomplete knowledge in problem solving, making hypotheses for this type of knowledge, with which we can infer the kinds of formula that hold based on the different kinds of hypothesis. The paper also shows that the semantics for default logic can be partly incorporated in the model theory. In this simple and clear logical framework for hypothetical reasoning, truth maintenance systems that have been widely used but lack model-theoretic semantics, can be analyzed theoretically. The result shows that hypothetical reasoning is an important subcase of default logic for which efficient theorem proving techniques exist.

Paper Length: 5000 words.

Topic: Automated Reasoning.

Keywords: Hypothetical Reasoning, Truth Maintenance, Default Logic, Non-monotonic Reasoning.

1. Introduction

In the real world, we are often forced to make some decision even if complete information is not available in problem solving. For example, in order to solve synthesis problems (such as design and planning) or in conjectural reasoning (such as inductive reasoning and analogy), we need draw some conclusion from incomplete knowledge, making reasoning non-monotonic. An approach to non-monotonic reasoning that has been pursued to use in practice is seen in the field of *truth maintenance systems* (TMSs) such as [Doyle 79] and [de Kleer 86]. The main task of TMSs is to maintain consistency of dynamic knowledge bases. In TMSs, when alternatives are found in solving problems, assumptions are used for them or a choice is selected nondeterministically and *dependency-directed search* is done to avoid redundant computing and rediscovering failures. Although TMSs have widely used as non-monotonic reasoning systems, there has been little formal research on them. Because they lack clear semantics, TMSs have not been well defined or understood.

The motivation of this research was to formalize reasoning systems with incomplete knowledge, that is, to clarify model-theoretic semantics of such reasoning systems. One example of this research can be seen in the *Clause Management System* (CMS) [Reiter & de Kleer 87], which simply formalizes de Kleer's *assumption-based TMS* (ATMS). We propose an alternative approach for reasoning with incomplete knowledge, in terms of *hypothetical reasoning*, which is also simple and more general than the CMS. Hypothetical reasoning is desirable when dealing with alternative knowledge or incomplete knowledge in problem solving, and it makes hypotheses for these types of knowledge, with which we can infer what formulas hold based on what hypotheses. In this paper, firstly, two different aspects of hypothetical reasoning are shown, and are then represented in a simple logical framework. Secondly, Reiter's *default logic* [Reiter 80] is shown to be incorporated in the framework, where the hypothesis corresponds to the simplest normal default and a default theory is defined in terms of extensions. Lastly, the model theory for the ATMS and Doyle's TMS are presented, based on this logical framework for hypothetical reasoning.

2. Two Aspects of Hypothetical Reasoning

In hypothetical reasoning, *hypotheses* (or *assumptions*) are not guaranteed to be true or not always true. The hypotheses are first given by a problem solver, which may be domain-dependent, and they are used as unknown information in the problem solver's decision. Typical uses of hypothetical reasoning are broadly classified as follows.

1. Abductive reasoning
In model-based reasoning, when an observation, q, of a system's behavior is found, we want to know hypotheses that explain observation q with knowledge of descriptions of the system and devices. Consider the following reasoning, called the *fallacy of affirming the consequent* in traditional logic:

$$q, \quad \text{and} \quad p \supset q, \quad \text{therefore} \quad p.$$

As this is an invalid rule, we cannot conclude deductively the truth of p. We can only say that p is a *hypothesis* that explains q. This type of reasoning is called ***abductive reasoning***, and in engineering, this can be applied directly to diagnostic reasoning such as the theory formation in [Poole 86]. When the specifications are given as observations, it can be also applied to design such as the resolution residue in [Finger & Genesereth 85].

2. Consistency maintenance in knowledge bases
When we must select from alternatives but information is not enough complete to select one deterministically, we must choose one, say p, and process reasoning further from there. Consider the following reasoning, called *hypothetical syllogism* or *modus ponens* in traditional logic:

$$p, \quad \text{and} \quad p \supset q, \quad \text{therefore} \quad q.$$

This is a valid rule but if we cannot conclude that p is true, we cannot conclude the truth of q either. In such a case, we can only say that q holds under *assumption p*. This type of reasoning is sometimes called *assumption-based reasoning*. Because the consequents of these assumptions and formulas derived from them are not guaranteed to be true, if a contradiction occurs in the reasoning process, we must remove the original assumptions and select other assumptions instead. Typical AI systems with this mechanism are implemented as TMSs, which maintain the consistency of knowledge bases, so that they can deal with contradictions. In engineering, constraint satisfaction, which is needed in parametric design, can be solved with this type of reasoning (e.g., [Inoue 87]).

Formal theories are studied mainly in abductive reasoning, such as [Poole 86] and [Reiter & de Kleer 87]. However, consistency maintenance mechanisms are pursued algorithmically rather than theoretically. We suggest that both types of hypothetical reasoning should be understood in the same model theory, that is, the latter case can be analyzed in the same way as the former. These two types of hypothetical reasoning just correspond to different aspects of the same logical framework for hypothetical reasoning and can be explained as the unified formalism given in the next section.

3. General Formalism

This section considers a model theory for hypothetical reasoning. First, reasoning from knowledge is formalized. Then we show a logical framework for supporting hypotheses which sanction a given propositional well-formed formulas. The framework corresponds to a model theory for abductive reasoning. Next, reasoning from beliefs is formalized within the framework. This formalism reflects the important property of consistency maintenance. Finally, the formalism is compared with Reiter's default logic. All proofs of theorems in this paper are shown in the full paper.

3.1 Supporting Hypotheses

Let U be a set of finite propositional symbols, and L be a propositional language with U and logical connectives $\neg$ and $\supset$ and the set of *well-formed formulas* (*wffs*) defined in the usual way. Logical connectives $\vee$ and $\wedge$ are defined in terms of $\neg$ and $\supset$, that is, $\alpha \vee \beta = \neg\alpha \supset \beta$ and $\alpha \wedge \beta = \neg(\alpha \supset \neg\beta)$. A *literal* is a symbol in U or the negation of a symbol in U. A *clause* is a wff such that it is a finite disjunction of zero or more literals. An *interpretation*, I, of U is defined as a set of literals such that for each $\alpha \in U$, either $\alpha \in I$, or $\neg\alpha \in I$, corresponding to be assigned to true or false, respectively. A set of interpretations is denoted as Ω. The relation, $\models$, is defined as follows: let $I \in \Omega$ and $\alpha, \beta \in L$, (1) for $\alpha \in U$, $I \models \alpha$ iff $\alpha \in I$, (2) $I \models \neg\alpha$ iff not $I \models \alpha$, and (3) $I \models \alpha \supset \beta$ iff $I \models \neg\alpha$ or $I \models \beta$. And let $M \in \Omega$ be an interpretation and W be a set of wffs in L, then the fact that M *satisfies* W is defined as $M \models \alpha$ for all $\alpha \in W$. In this case, we say that M is a *model* of W. W is said to be *satisfiable* if it has at least one model. The set of all models of W is denoted as $MOD(W)$. We say that for a wff, $\alpha \in L$, W *entails* α (written $W \models \alpha$), if for each model, $M \in MOD(W)$, $M \models \alpha$ holds. In this case, we say that α is a *logical consequence* of W.

Definition 1. A set of *premises*, Σ, is a satisfiable set of wffs. A *set of knowledge of* Σ is a set of all logical consequences of Σ. A wff, $w \in L$, is a *knowledge of* Σ iff $\Sigma \models w$. □

Definition 2. Suppose that a set of premises, Σ, is given. A wff, $d \in L$, is *indefinite with respect to* Σ iff $\Sigma \not\models d$ and $\Sigma \not\models \neg d$, that is, there exist models $\sigma_1, \sigma_2 \in MOD(\Sigma)$ such that $\sigma_1 \neq \sigma_2$, $\sigma_1 \models \neg d$ and $\sigma_2 \models d$. A set of wffs, $D_{(\Sigma)}$, is a *set of hypotheses with respect to* Σ iff each wff, $d \in D_{(\Sigma)}$ (called *hypothesis*), is indefinite with respect to Σ. A set of wffs, E, is an *environment of* $D_{(\Sigma)}$ iff $E \in 2^{D_{(\Sigma)}}$. The set of all indefinite wffs with respect to Σ is denoted as $DMAX_{(\Sigma)}$. In the subsequent discussion, we omit the subscript "$_{(\Sigma)}$" in D or $DMAX$ when they are clear in the situation, and we simply say "indefinite wff" or "set of hypotheses", omitting "with respect to Σ". □

Definition 3. Let Σ, D and w be a set of premises, a set of hypotheses and a wff in L, respectively. A wff, $d \in D$, is a *supporting hypothesis for* w *with respect to* Σ *and* D iff

$$\Sigma \models d \supset w$$

holds. The set of all supporting hypotheses for w with respect to Σ and D is denoted as $SMAX(\Sigma, D, w)$. □

Proposition 1. Let d be a supporting hypothesis for w with respect to Σ and D. Then,
(1) $\Sigma \cup \{d\}$ is satisfiable.
(2) $\Sigma \cup \{d\} \models w$.
(3) $\Sigma \cup \{\neg w\} \models \neg d$. □

Proposition 2. (1) If w is a knowledge of Σ, that is, $\Sigma \models w$, then for any wff set, $D \in 2^{DMAX}$, $SMAX(\Sigma, D, w) = D$.
(2) $w \in D_{(\Sigma)}$ iff $w \in SMAX(\Sigma, D_{(\Sigma)}, w)$.
(3) $\Sigma \models \neg w$ iff $SMAX(\Sigma, DMAX, w) = \phi$. □

Intuitively, a supporting hypothesis, d, for w with respect to Σ and D is a supplementary wff with which the truth of w is assured by the entailment of Σ, keeping consistency with Σ. Given w, the computation of a set of supporting hypotheses for w corresponds to abductive reasoning. We should pay attention to *the principle of parsimony*, that is, such supporting hypotheses are to be minimal as follows.

Definition 4. Let Σ, D and w be the same as Definition 3. A wff, $d \in D$, is a *minimal supporting hypothesis for* w *with respect to* Σ *and* D iff $d \in SMAX(\Sigma, D, w)$ and there exists no wff, $d' \in SMAX(\Sigma, D, w)$, such that $d \not\equiv d'$ and $d \supset d'$. The set of all minimal supporting hypotheses for w with respect to Σ and D is denoted as $SMIN(\Sigma, D, w)$. If $\Sigma \models w$, then $SMIN(\Sigma, D, w)$ is denoted as Φ. □

To analyze the properties of $SMIN(\Sigma, D, w)$, we define the set of all models of Σ that satisfy each indefinite wff, e, in an environment, E, of a set of hypotheses, D, as follows.

$$M(\Sigma, E) \stackrel{\text{def}}{=} \bigcup_{e \in E} MOD(\Sigma \cup \{e\}) .$$

Lemma 3. Let d and d' be two hypotheses in D, w be a wff in L. If $\Sigma \models d \supset w$ and $d' \supset d$, then $\Sigma \models d' \supset w$. □

Theorem 4. $M(\Sigma, SMIN(\Sigma, D, w)) = M(\Sigma, SMAX(\Sigma, D, w))$. □

Theorem 4 shows that only from $SMIN(\Sigma, D, w)$, all models of Σ that satisfy each hypothesis in $SMAX(\Sigma, D, w)$ can be computed.

Corollary 5 (monotonicity of supporting hypotheses). Let Σ be a set of premises, and D and D' be two sets of hypotheses. If $D \subseteq D'$, then $M(\Sigma, SMIN(\Sigma, D, w)) \subseteq M(\Sigma, SMIN(\Sigma, D', w))$. □

Example 1. Suppose that Σ is a set of premises, and that D_1 and D_2 are two sets of hypotheses, as follows.

$\Sigma = \{a \supset c,\ b \supset c,\ a \wedge b \supset g,\ \neg g\}$
$D_1 = \{a,\ b\}$
$D_2 = \{a,\ b,\ a \vee b,\ c,\ b \wedge c\}$

From this, the following sets of supporting hypotheses can be obtained.

$SMAX(\Sigma, D_1, c) = SMIN(\Sigma, D_1, c) = \{a,\ b\}$
$SMAX(\Sigma, D_2, c) = \{a,\ b,\ a \vee b,\ c,\ b \wedge c\}$
$SMIN(\Sigma, D_2, c) = \{a \vee b,\ c\}$
$SMAX(\Sigma, D_2, b \wedge c) = \{b,\ b \wedge c\}$
$SMIN(\Sigma, D_2, b \wedge c) = \{b\}$ □

The above definition of *supporting hypotheses* can be compared with the notion of 'support' of the CMS in [Reiter & de Kleer 87], where a support, $\neg S$, of the CMS roughly corresponds to our supporting hypothesis, S. However, while in the CMS all wffs have to be translated to the clausal normal form and some or all 'prime implicants' of Σ of propositional clauses are considered so that $SMIN(\Sigma, DMAX, C)$ for any clause C is computed, we do not require this translation. We focus only on model-theoretic semantics, and our formalism with any wffs is more general than one with the clausal normal form. For example, our formalism naturally provides minimal supporting hypotheses for conjunctive observations such as $SMIN(\Sigma, D_2, b \wedge c)$ in Example 1. Moreover, we deal with indefinite wffs not by $DMAX$ but by the *set of hypotheses* supplied by the problem solver, which are all that is needed to construct a set of supporting hypotheses for any wff. The separation between hypotheses and other wffs makes knowledge bases comprehensive and helps the system to compute clearly and efficiently. In our formalism, as the set of hypotheses, D, is explicitly given in the model theory, the sets of *beliefs* can be constructed from Σ and environments of D in Section 3.2, and they can be characterized in terms of *extensions* in default logic in Section 3.3.

3.2 Beliefs

One technique for *efficient* computation of $SMIN(\Sigma, D, w)$ for a wff, $w \in L$, can be considered to be a mechanism to keep dependencies or the original assumptions with each wffs in TMSs, as stated in Section 4. Therefore, the above formalism is a model theory for consistency maintenance mechanisms as well as for abductive reasoning. Now, we define reasoning from beliefs of a set of premises and an environment, which is a basis for consistency maintenance mechanisms, as follows.

Definition 5. Suppose that Σ is a set of premises and that D is a set of hypotheses. Let E be an environment of D such that $\Sigma \cup E$ is satisfiable. A *set of beliefs of* Σ *and* E (denoted $B(\Sigma, E)$) is a satisfiable set of all logical consequences of $\Sigma \cup E$. A wff, $w \in L$, is a *belief of* Σ *and* E iff $\Sigma \cup E \models w$. □

Lemma 6. $MOD(B(\Sigma, E)) = MOD(\Sigma \cup E) \subseteq M(\Sigma, E)$. □

Theorem 7. Suppose that Σ, D and E are the same as Definition 5. Let w be a wff in L, then, (1) if w is a belief of Σ and E, then E is a set of supporting hypotheses for w with respect to Σ and D, and (2) if an indefinite wff, $d \in D$, is a supporting hypothesis for w with respect to Σ and D, then w is a belief of Σ and E, where E is an environment of D and there exists a hypothesis, $d' \in E$, such that $d' \supset d$. □

Theorem 7 gives a bridge between consistency maintenance and abductive reasoning. The next two theorems show that the set of models of $B(\Sigma, E)$ decrease monotonically when either a hypothesis is added to E, or a premise is added to Σ.

Theorem 8 (model monotonicity of hypotheses)**.** Suppose that Σ, D and E are the same as Definition 5. Let d be a hypothesis in D. And let E and E' be two environments of D. Then, the following properties hold.
(1) If $d \notin E$, then $MOD(B(\Sigma, E \cup \{d\})) = MOD(B(\Sigma, E)) - \{\gamma \mid \gamma \models \neg d,\ \gamma \in \Omega\}$.
(2) If $E \subseteq E'$, then $MOD(B(\Sigma, E')) \subseteq MOD(B(\Sigma, E))$. □

Theorem 9 (model monotonicity of premises)**.** Suppose that Σ and Σ' are two sets of premises, and that E and E' are environments of D. If $\Sigma \subseteq \Sigma'$ and $E' = E - \{w \mid \Sigma' \models w$ or $\Sigma' \models \neg w\}$, then $MOD(B(\Sigma', E')) \subseteq MOD(B(\Sigma, E))$. □

3.3 Extensions

A set of hypotheses, D, is very closely related to a restricted case of a set of normal default in Reiter's default logic [Reiter 80]. Here, we investigate the relationship between a set of beliefs and an extension of a normal default theory. In the following discussion, a normal default with the form, $:\mathrm{M}\, d \,/\, d$, is only necessary for our logic.

Definition 6. Let Σ be a set of premises. A *default with respect to* Σ is any expression of the form $:\mathrm{M}\, d \,/\, d$, where d is an indefinite wff with respect to Σ. And let $DR_{(\Sigma)}$ be a set of defaults with respect to Σ. The *set of consequents of defaults of* $DR_{(\Sigma)}$ (denoted $CONS(DR_{(\Sigma)})$) is $\{d \mid :\mathrm{M}\, d\, /\, d \in DR_{(\Sigma)}\}$. □

Given a set of defaults, $DR_{(\Sigma)}$, we can characterize a *set of hypotheses*, D, as $D_{(\Sigma)} = CONS(DR_{(\Sigma)})$. We now give a definition of an *extension* by a fixed point construction of sets of beliefs of Σ and an environment of D, which can be shown to be equivalent to the original definition in [Reiter 80].

Definition 7. Let Σ be a set of premises, and D be a set of hypotheses. A fixed point operator for Σ and D is defined as:

$$NM_{(\Sigma,D)}(S) \stackrel{\text{def}}{=} B(\Sigma, \{d \mid d \in D, \neg d \notin S\}).$$

An *extension of* Σ *and* D is the fixed point of this operator, that is, S is an extension of Σ and D iff $NM_{(\Sigma,D)}(S) = S$. □

All concepts and results of normal default theories in default logic are now applicable to our model theory. In particular, by the property of *semi-monotonicity*, $MOD(\Sigma)$ is restrictedly selected by repeatedly applying each hypothesis toward a set of all models of an extension of Σ and D. We have already seen in Theorem 8 that $MOD(B(\Sigma, E))$ decreases monotonically as environment E increases. A set of beliefs of Σ and such a maximal environment E' in 2^D corresponds to an extension of Σ and D, where $MOD(B(\Sigma, E'))$ is such a minimal set of models in $2^{MOD(\Sigma)}$. This result corresponds to the semantics for normal default theories in [Etherington 87].

Last, in this section, it should be noted that although we state supporting hypotheses or default theories in the propositional language to make the discussion clear, all definitions of them can be extended to have the first order predicate calculus, where closed formulas correspond to wffs and a model, M, of a set of function-free formulas is represented by enumerating the ground atomic formulas true for M.

4. Truth Maintenance Systems

The logical framework for hypothetical reasoning subsumes various TMSs. This section shows how the general formalism is related to TMSs, especially to de Kleer's *assumption-based TMS* (ATMS) [de Kleer 86] and Doyle's non-monotonic *justification-based TMS* (JTMS) [Doyle 79], which are two different kinds of representative TMSs.

4.1 de Kleer's ATMS

The ATMS maintains a global, concurrent representation of all sets of beliefs by labeling each atomic formulas with all minimal supporting hypotheses. In other words, the ATMS keeps all models of all extensions of Σ and D. More formally, in the ATMS, a set of premises, Σ, is given as a set of *justifications* submitted by the problem solver, each of which is either a Horn clause (i.e., rule clause, unit clause, or negative clause (called *nogood*)), or a positive clause (called *choose*) in the extended version. In the ATMS, a set of *assumptions*, AS, is given as a set of indefinite atomic formulas with respect to Σ in U. Then, in our general model theory, the set of hypotheses, D, is as follows.

$$D \stackrel{\text{def}}{=} \{ d \mid d = \bigwedge_{a \in A} a, \ A \in 2^{AS} \}.$$

Each positive literal, w, in the set of all extensions of Σ and D forms a *node*, within which $SMIN(\Sigma, D, w)$ (called the *label*) is maintained. Each hypothesis, d, in the label $SMIN(\Sigma, D, w)$ is called an 'environment' in the ATMS. A set of all atomic clauses in a set of beliefs of Σ and A ($A \in 2^{AS}$) corresponds to a *context* in the ATMS.

We can generalize the above characterization of the ATMS with our formalism. In particular, the generalization of the set of premises, Σ, from Horn clauses to general clauses is equivalent to one with the CMS [Reiter & de Kleer 87]. In this case, a node can be extended to express a supported wff of not only a negative clause but of any clause. The following results give a characterization of the generalized ATMS.

Lemma 10. Suppose that Σ, AS and D are the same as the above definition, except that Σ is given as a set of (general) clauses. Let E be an environment of D such that $\Sigma \cup E$ is satisfiable. Then, $B(\Sigma, E) = B(\Sigma, \{ a \mid a \in AS, \ \bigwedge a \in E \})$. □

Theorem 11. Suppose that Σ, AS and D are the same as Lemma 10. The set of all extensions of Σ and D is equivalent to the set of all extensions of Σ and AS. □

The next theorem is an immediate consequence of Theorem 7.

Theorem 12. Suppose that Σ, AS and D are the same as Lemma 10. Let w be a wff in L. $SMIN(\Sigma, D, w) \neq \phi$ iff w is a belief of some extension of Σ and AS. $SMIN(\Sigma, D, w) = \Phi$ iff w is a knowledge of Σ. □

How do we consider constructing the deductive system that preserves the semantics of the ATMS? The computation of all extensions of Σ and AS is called *interpretation construction*. By the property of *semi-monotonicity* of normal default theories [Reiter 80], each hypothesis (or even assumption in the ATMS) can be treated independently, so that interpretation construction can be done simply by applying one assumption after another and ignoring some assumptions. However, in the real world, a major problem arises. Since not all premises as well as all assumptions are explicitly available, the ultimate set of all premises, Σ, is implicitly defined. Indeed, Theorem 9 shows the global property of Σ, that is, if Σ grows, then the environment taken into account decreases. Let us consider the following example.

Example 2. Suppose that Σ_1 and Σ_2 are two sets of premises, and that AS_1 and AS_2 are two sets of assumptions. Let D_1 and D_2 be two sets of hypotheses for AS_1 and AS_2, respectively, given by the previous definition. From

$\Sigma_1 = \{ a \vee b, \ a \wedge c \supset \perp, \ c \supset g \}$ and
$AS_1 = \{ a, \ b, \ c \}$,

we can obtain $SMIN(\Sigma_1, D_1, g) = \{c\}$. In this case, we have two extensions of Σ_1 and AS_1, i.e., $B(\Sigma_1, \{a, b\})$ and $B(\Sigma_1, \{b, c\})$. Then, the following premise is added to Σ_1, and Σ_2 is created as

$\Sigma_2 = \Sigma_1 \cup \{ b \wedge c \supset \perp \}$.

Since $\Sigma_2 \models \neg c$, c is no longer indefinite. Therefore, from

$$AS_2 = AS_1 - \{ c \},$$

we now obtain $SMIN(\Sigma_2, D_2, g) = \phi$. There is only one extension of Σ_2 and AS_2, i.e., $B(\Sigma_2, \{a, b\})$. □

To avoid redundant computing during interpretation construction, *dependency-directed search* (DDS) is utilized in the ATMS. DDS is also required if only part of the search space should be explored for the purpose of the characteristics of tasks, such that not all solutions are required at once, or more efficiency is required. In [Inoue 87], an efficient algorithm is given for DDS, where reasoning is controlled by an AND/OR tree search mechanism, and assumptions and premises can be added to a set of beliefs incrementally constructing all extensions.

4.2 Doyle's TMS

The JTMS focuses only one current model of a set of beliefs at a time. In our model theory, it is implicitly assumed that there are all models of every possible set of beliefs. Therefore, the JTMS is interpreted as an extended reasoning module for the logical framework for hypothetical reasoning, where a problem solver will somehow choose a single model of an extension within which to reason about the world. DDS is provided if a contradiction occurs. Switching the set of beliefs is called *belief revision.* This mechanism is the most popular method to provide default reasoning systems. The JTMS tries to establish a maximal satisfiable set of *nodes* supported by the *justifications.*

We give a very simple formalism of the JTMS. Note that this is not an exact model theory for the JTMS, but one explaining the above notion of the extended reasoning module. An interpretation, γ, of U is called a *labeling.* A *node* is defined to be associated with each atomic formula in U. Let γ_c be the *current* labeling. Then, (1) a node w is 'believed' (or is labeled *in*) iff $w \in \gamma_c$ (i.e., $\gamma_c \models w$), and (2) a node w is not 'believed' (or is labeled *out*) iff $\neg w \in \gamma_c$ (i.e., $\gamma_c \not\models w$). A *justification* for w has the form, $J(w) \supset w$, where $J(w)$ consists of the following two disjoint parts, $I(w)$ (called the *inlist*) and $O(w)$ (called the *outlist*) such that

$$J(w) = I(w) \wedge O(w), \text{ where } J(w) \neq \phi, \text{ and } I(w) = \bigwedge_i \alpha_i,\ O(w) = \bigwedge_j \neg\beta_j,\ \alpha_i, \beta_j \in U.$$

A set of premises, Σ, is given as a set of justifications and unit clauses. Given Σ, we say that a labeling, γ, is *valid* iff $\gamma \in MOD(\Sigma)$. And we say that a labeling, γ, is *well-founded* iff γ is valid and for each 'believed' node, w, all literals of $J(w)$ are in γ, that is, all nodes on $I(w)$ are *in* and all nodes on $O(w)$ are *out.*

Theorem 13. Let Σ be a set of premises, and γ be a labeling. If $\gamma \in MOD(\,\Sigma \cup \{ w \supset J(w) \mid J(w) \supset w \in \Sigma \})$, then γ is well-founded. □

The truth maintenance process involves finding a well-founded labeling. The JTMS selects only one model from the set of all well-founded models. Whenever new facts or justifications are added to Σ, the process is repeated. From Theorem 9, when a new justification is added, the set of models of beliefs will decrease. We must choose some model from the restricted models, $MOD(B(\Sigma', E'))$ in Theorem 9, and this corresponds to belief revision. We need a criterion for belief revision; intuitively, this is done by exchanging the elements between sets of positive and negative literals in γ *as minimally as possible*. This kind of criterion is addressed in [Satoh 87].

4.3 Discussion

The construction of the ATMS is more straightforward than the JTMS in the model theory for hypothetical reasoning. The 'brave' character of default logic, where each extension is treated as an acceptable set of beliefs is very close to the notion of keeping all models of beliefs in the ATMS. Our formalism gives one natural way to interpret the JTMS as an extended reasoning module for the ATMS, so that the *current model* is selected from all models of sets of beliefs. *Minimal belief revision* should be involved by comparing it with other models maintained by the ATMS.

A major problem of the JTMS is that the algorithm and the data structure are too complex, and as a result, the formal analysis is very difficult. Our formalism for the JTMS is not enough complete but much simpler and clearer. In [Doyle 83], the exact specification for the JTMS is tried, so that the correctness of his algorithm is proven. Conversely, we first formalize hypothetical reasoning and interpret the JTMS as an implementation of the extra reasoning module for it. In [Morris 87], the relationship between the JTMS and Reiter's non-normal defaults is shown not semantically but syntactically. However, the motivation of their research is to solve a multiple extensions problem in default logic, which can be solved by the *unidirectional* property of the justifications of the JTMS (shown in Theorem 13, in our model theory). Again, we first construct the formalism for hypothetical reasoning, then we compare the model theory only with the normal default theory, which is more tractable.

5. Conclusion

This paper presented a logical framework for hypothetical reasoning, which formalizes reasoning systems with incomplete knowledge. The model theory was shown to be the formalism for both abductive reasoning and consistency maintenance in knowledge bases. The key idea is that hypotheses are treated as indefinite wffs, making it possible to relate them to the restricted case of normal default. The paper also described the formalism for TMSs in our model theory for hypothetical reasoning. These results show that hypothetical reasoning is an important subcase of default logic for which efficient theorem proving techniques exist.

Acknowledgments

I would like to thank Dr. Koichi Furukawa, Ken Satoh and Jun Arima of ICOT for their helpful discussions and useful comments for this research.

References

[de Kleer 86] de Kleer, J., "An Assumption-based TMS", *Artificial Intelligence* **28** (1986), pp. 127–162.

[Doyle 79] Doyle, J., "A Truth Maintenance System", *Artificial Intelligence* **12** (1986), pp. 231–272.

[Doyle 83] Doyle, J., "The Ins and Outs of Reason Maintenance", *Proc. IJCAI-83* (1983), pp. 349–351.

[Etherington 87] Etherington, D. W., "A Semantics for Default Logic", *Proc. IJCAI-87* (1987), pp. 495–498.

[Finger & Genesereth 85] Finger, J.J. and Genesereth, M.R., *RESIDUE: A Deductive Approach to Design Synthesis*, Technical Report STAN-CS-85-1035, Department of Computer Science, Stanford University, 1985.

[Inoue 87] Inoue, K., *Pruning Search Trees in Assumption-based Reasoning*, ICOT Technical Report TR-333, ICOT, 1987.

[Morris 87] Morris, P., "Curing Anomalous Extensions", *Proc. AAAI-87* (1987), pp. 437–442.

[Poole 86] Poole, D. L., *Default Reasoning and Diagnosis as Theory Formation*, Technical Report CS-86-08, Department of Computer Science, University of Waterloo, 1986.

[Reiter 80] Reiter, R., "A Logic for Default Reasoning", *Artificial Intelligence* **13** (1980), pp. 81–132.

[Reiter & de Kleer 87] Reiter, R. and de Kleer, J., "Foundations of Assumption-based Truth Maintenance Systems: Preliminary Report", *Proc. AAAI-87* (1987), pp. 183–188.

[Satoh 87] Satoh, K., *A Minimal Change of Belief—A Criterion of Belief Revision*, ICOT Technical Report TR-297, ICOT, 1987.