NONMONOTONIC REASONING BY MINIMAL BELIEF REVISION

Ken Satoh

ICOT Research Center,

1–4–28, Mita, Minato-ku, Tokyo 108, Japan

csnet: ksatoh%icot.jp@relay.cs.net

uucp: {enea,inria,kddlab,mit-eddie,ukc}!icot!ksatoh

ABSTRACT

This paper presents a formalism of *nonmonotonic reasoning*. In our formalism, initial belief is represented directly. Unlike the current formalisms such as default logic or circumscription we can express the belief without any extra inference rules or special axiom. The only constraint for a belief is that it must entail knowledge. Then, we define a belief revision strategy called *minimal belief revision*. Minimal belief revision minimizes the difference between the previous belief and the new belief so that what was true in the previous belief remains true in the new belief as far as possible.

This paper discusses why belief revision must occur when belief does not entail added knowledge, presents a definition and model theory for minimal belief revision and shows that minimal belief revision performs some kind of nonmonotonic reasoning.

1 INTRODUCTION

In real life, we are sometimes forced to draw some conclusion even if there is not enough information. For one solution to those situations, we use our belief (or hypothesis) to complement unknown information. However, since the results from a belief are not logically true, they must be defeated if they are found to be false. Such reasoning is called *nonmonotonic reasoning* and has been formalized by various researchers [McCarthy80, McDermott80, and Reiter80].

Roughly speaking, current formalisms such as circumscription [McCarthy80], nonmonotonic logic [McDermott80] and default logic [Reiter80] add special axioms or define extra inference rules to incorporate the idea that an unknown fact is assumed to be false unless it is explicitly known to be true. The special axioms or extra inference rules can be regarded as producing methods of plausible belief from the current knowledge. If more knowledge is added, a different belief is produced by those axioms or extra inference rules.

This paper presents another formalism of nonmonotonic reasoning. Our approach of formalizing nonmonotonic reasoning is different from those formalisms in the following points.

(1) A belief is represented directly as a plausible hypothesis without any extra inference rules or special axioms. If no conclusion is derived from incomplete knowledge, a belief is used to complement unknown information. The only constraint for a belief is that it must entail knowledge; in other words, it is a detailed hypothetical description of knowledge to supplement a lack of knowledge.

(2) A belief is directly revised when more information is added. Since a belief has a hypothetical character, it is not always true. Therefore, if a belief does not entail added information, it must be changed to satisfy the above constraint. This process of change is called *belief revision*.

This paper investigates a special strategy of belief revision called *minimal belief revision* and shows that this strategy performs some kinds of nonmonotonic reasoning. The idea of minimal belief revision is that default rules are first defined as belief, and if any counter-example is found, the belief is changed so that the counter-example is treated as an exception to maintain consistency.

For example, suppose that belief for flying birds is expressed by the following:

$$\forall x(bird(x) \supset fly(x)).$$

The above belief expresses directly that every bird flies. Even if we only know $bird(A)$, we conclude $fly(A)$ as a consequence of the above belief. However, if we find $\neg fly(A)$ in addition to $bird(A)$, then the belief must be changed to keep consistency. However, we do not want to throw away the above belief completely, but we still want to believe that any bird other than A flies. The

minimal belief revision strategy performs such revision and changes the above belief into the following:

$$\forall x(x \neq A \equiv (bird(x) \supset fly(x))).$$

This belief revision is nonmonotonic, because from the previous belief, we can derive $fly(A)$ if we know $bird(A)$, whereas from the new belief, we can no longer derive $fly(A)$.

This paper first discusses the relationship between knowledge and belief, then shows the definition and model theory of minimal belief revision and some examples of nonmonotonic reasoning with it.

2 MONOTONIC KNOWLEDGE AND NONMONOTONIC BELIEF

This section shows why belief revision must occur when belief does not entail added knowledge. Let κ and β be a set of knowledge and a set of beliefs respectively. [Hintikka62] requires knowledge and belief to satisfy the following relation:

$$\kappa \subseteq \beta,$$

which means that if an agent knows p, he also believes p. In this paper, knowledge and beliefs are represented as logical formulas. Let knowledge and belief be formulas of K and B. The above requirement can be expressed as follows:

$$B \models K,$$

which means that B entails K [1].

We also define that knowledge is monotonic. Following Hintikka, we regard knowledge as a subset of truth, and therefore, knowledge increases monotonically. However, since we regard a belief as a set of plausible hypotheses, a belief must be nonmonotonic if we keep consistency of belief. For example, suppose that the current knowledge contains neither α nor $\neg\alpha$ and the belief contains α. Then, if $\neg\alpha$ is added to knowledge, the belief must contain it because of the above requirement. However, the simple addition of $\neg\alpha$ to the belief leads to a contradiction. Therefore, in this case, α in the belief must be retracted in order to maintain consistency. Therefore, when new knowledge is added, a belief must be revised so that it contains new knowledge and is satisfiable. This process is called *belief revision*.

A strategy on how to change belief is needed since there are many ways of changing the belief. The next section concentrates on one particular belief revision strat-

egy called *minimal belief revision*. Minimal belief revision is a strategy by which differences between the previous belief and revised belief are minimized.

3 MINIMAL BELIEF REVISION

The main idea of minimal belief revision is that we compute the differences between a model of previous belief and a model of new knowledge, and choose pairs of models where the difference between models is minimal. As a result, what was true in the previous belief remains true in the new belief if it is not contradictory to the added knowledge.

We explain the idea with the following example. Let knowledge K be identical to $\mathbf{T}$, which denotes a true proposition, and belief B be identical to $p \wedge q$. Then suppose that α, which is equivalent to $\neg p \vee \neg q$, is added to knowledge K; we must revise belief because $B \not\models K \wedge \alpha$. Let the new belief be NB. NB must entail $K \wedge \alpha$ by the above requirement, that is, a set of models of NB must be a subset of the following set:

$$\{\{\neg p, q\}, \{p, \neg q\}, \{\neg p, \neg q\}\},$$

which is a set of all models of $K \wedge \alpha$ (a model is represented as a set of the propositional constants and negation of propositional constants that are true in the model).

First, we compute the difference between each model of B and each model of $K \wedge \alpha$. In this case, we compute the differences of $(\{p, q\}, \{\neg p, q\})$, $(\{p, q\}, \{p, \neg q\})$ and $(\{p, q\}, \{\neg p, \neg q\})$ which are $\{p\}, \{q\}$ and $\{p, q\}$ respectively (the difference set is represented as propositional constants which have different truth values for each model in the pair). Then we select pairs whose differences are minimal in terms of set inclusion, that is, $(\{p, q\}, \{\neg p, q\})$ and $(\{p, q\}, \{p, \neg q\})$. Then the new belief, NB, by minimal belief revision is the disjunction of all those models of knowledge in the above pairs, that is:

$$(\neg p \wedge q) \vee (p \wedge \neg q).$$

While new knowledge has a model of $\{\neg p, \neg q\}$, NB does not have it. This is because what was true in the previous belief remains true as far as possible by minimal belief revision. Thus, either p or q (not both) remains true in the new belief.

We generalize this idea to knowledge and belief expressed in a first-order language. In the following subsections, minimal belief revision is defined in a second-order language. In a second-order language, we can use predicate variables and function variables in addition to object variables. Predicate variables vary over predicates

---

[1]Note that circumscription and default logic satisfy this requirement, because both produce a belief including an initial axiom.

and function variables vary over functions. In addition, we use predicate constants such as **T** for true, **F** for false and = for equality, and logical connectives such as $\oplus$ for exclusive-or and $\equiv$ for equivalence.

We also give the model theory of minimal belief revision for a second-order language. A *structure*, M, for a second-order language consists of a domain D, which is a non-empty set, and an interpretation function such that every n-ary function constant, F_n, is mapped onto a function from D^n to D (written $M[F_n]$), and every n-ary predicate constant, P_n, is mapped into a subset of D^n (written $M[P_n]$). N-ary function variables range over any function from D^n to D, and n-ary predicate variables range over any subset of D^n. $< t_1, ..., t_n >_M$ denotes an interpreted tuple where $t_1, ..., t_n$ are terms. If $P_n(t_1, ..., t_n)$ is true in M, this fact is expressed as $< t_1, ..., t_n >_M \in M[P_n]$. A *model* of a second-order sentence is any structure, M, such that every formula in the set is true in M.

3.1 Definition of Minimal Belief Revision

Let $B(P)$, $K(P)$ and $\alpha(P)$ be first-order sentences whose predicate constants are among those of $P = p_1, ..., p_n$, and $B(P) \models K(P)$. $B(P)$ is the current belief and $K(P)$ is the current knowledge and $\alpha(P)$ is the added knowledge.

We define a minimal revised belief, $NB(P)$, with respect to $(K(P), B(P))$ and $\alpha(P)$ as follows.
$NB(P) \overset{\text{def}}{=}$

$$\exists P_B(B(P_B) \wedge K(P) \wedge \alpha(P) \wedge$$
$$\neg \exists P_B' \exists P'(B(P_B') \wedge K(P') \wedge \alpha(P') \wedge$$
$$(P_B', P') \prec (P_B, P))),$$

where

(1) P_B is a tuple of predicate variables $p_{B1}, ..., p_{Bn}$ which have the same arities of $p_1, ..., p_n$ respectively, and P_B' is also a tuple of predicate variables $p_{B1}', ..., p_{Bn}'$ which satisfy the same condition as P_B, P' is a tuple of predicate variables $p_1', ..., p_n'$ which also satisfy the same condition as P_B,

(2) and $B(P_B)$ is a sentence obtained by substituting predicate variables of P_B for any occurrence of corresponding predicate constants in $B(P)$, and $B(P_B')$ and $K(P')$ are sentences obtained in a similar way,

(3) and $(P_B', P') \prec (P_B, P)$ is an abbreviation of:

$$(P_B', P') \preceq (P_B, P) \wedge \neg((P_B, P) \preceq (P_B', P')),$$

where $(P_B', P') \preceq (P_B, P)$ is an abbreviation of:

$$\forall x((p_{B1}'(x) \oplus p_1'(x)) \supset (p_{B1}(x) \oplus p_1(x))) \wedge ... \wedge$$
$$\forall x((p_{Bn}'(x) \oplus p_n'(x)) \supset (p_{Bn}(x) \oplus p_n(x))).$$

$(P_B', P') \prec (P_B, P)$ means informally that the difference of extensions for P_B' and P' is less than the difference of extensions for P_B and P. $NB(P)$ expresses informally that there is some tuple of extensions for P_B which changes minimally into a tuple of extensions for P.

3.2 Model Theory

Let $B(P), NB(P), K(P)$ and $\alpha(P)$ be the same sentences in the above definition, and let M_B and M_B' be models of $B(P)$, and let M_{NB} and M_{NB}' be models of $K(P) \wedge \alpha(P)$. We define a partial order relation, $\preceq$, over pairs of models. $(M_B', M_{NB}') \preceq (M_B, M_{NB})$ is defined as the following.

(1) M_B, M_B', M_{NB} and M_{NB}' have the same domain.

(2) Every constant and function receives the same interpretation in M_B, M_B', M_{NB} and M_{NB}'.

(3) The following statement is true. (We write $< x >$ as an interpreted term in M_B, M_B', M_{NB} and M_{NB}', because it receives the same interpretation in all of those models.)

$$\forall x((< x > \in M_B'[p_1] \oplus < x > \in M_{NB}'[p_1]) \supset$$
$$(< x > \in M_B[p_1] \oplus < x > \in M_{NB}[p_1])) \wedge ... \wedge$$
$$\forall x((< x > \in M_B'[p_n] \oplus < x > \in M_{NB}'[p_n]) \supset$$
$$(< x > \in M_B[p_n] \oplus < x > \in M_{NB}[p_n])).$$

This ordering means that the difference of extensions of each p_i in M_B' and M_{NB}' is not more than the difference of extensions of each p_i in M_B and M_{NB}.

A *minimally different pair*, (M_B, M_{NB}), with respect to $B(P)$ and $K(P) \wedge \alpha(P)$ is defined as the pair of models for $B(P)$ and $K(P) \wedge \alpha(P)$ respectively such that there is no pair, (M_B', M_{NB}'), such that $(M_B', M_{NB}') \preceq (M_B, M_{NB})$ and not $(M_B, M_{NB}) \preceq (M_B', M_{NB}')$.

A *minimal revised model*, M_{NB}, with respect to $B(P)$ and $K(P) \wedge \alpha(P)$ is defined as the model such that there exists M_B such that (M_B, M_{NB}) is a minimally different pair with respect to $B(P)$ and $K(P) \wedge \alpha(P)$.

The relation between the definition of minimal revised belief and the model theory is as follows.

Proposition 1. *Let $B(P), NB(P), K(P)$ and $\alpha(P)$ be the same sentences in the definition of minimal revised belief. M_{NB} is a model of $NB(P)$ iff M_{NB} is a minimal revised model with respect to $B(P)$ and $K(P) \wedge \alpha(P)$.*

This proposition means that any result derived from $NB(P)$ is true in all minimal revised models.

As shown by the following proposition, if added knowledge is consistent with the belief, a set of models of the new belief is a maximum restricted set of models of belief so that any model in the new belief entails added knowledge.

Proposition 2. *Let $B(P)$, $NB(P)$, $K(P)$ and $\alpha(P)$ be the same sentences in the definition of minimal revised belief. If $\alpha(P)$ is consistent with $B(P)$, $NB(P) \equiv B(P) \wedge \alpha(P)$.*

The proofs of the above propositions are found in the appendix.

4 EXAMPLES

The previous formula of a minimal revised belief, $NB(P)$, can be translated into the following form:

$$\exists P_B(B(P_B) \wedge K(P) \wedge \alpha(P) \wedge$$
$$\forall P'_B \forall P'((B(P'_B) \wedge K(P') \wedge \alpha(P') \wedge$$
$$(P'_B, P') \preceq (P_B, P)) \supset (P'_B, P') \equiv (P_B, P))),$$

where $(P'_B, P') \equiv (P_B, P)$ is an abbreviation of:

$$\forall x((p'_{B1}(x) \oplus p'_1(x)) \equiv (p_{B1}(x) \oplus p_1(x))) \wedge ... \wedge$$
$$\forall x((p'_{Bn}(x) \oplus p'_n(x)) \equiv (p_{Bn}(x) \oplus p_n(x))).$$

The above formula is used in the following examples.

The first example shows a propositional case.

Example 1:

$P = p, q$
$B(P) \equiv p \wedge q,$
$K(P) \equiv T,$
$\alpha(P) \equiv \neg p \vee \neg q$

$B(P)$ and $K(P)$ are the current belief and the current knowledge respectively and $\alpha(P)$ is the added knowledge. Then the minimal revised belief, $NB(P)$, is defined as follows.

$$NB((p, q)) \equiv$$
$$\exists p_B \exists q_B($$
$$\quad p_B \wedge q_B \wedge (\neg p \vee \neg q) \wedge$$
$$\quad \forall p'_B \forall q'_B \forall p' \forall q'((\ $$
$$\quad\quad p'_B \wedge q'_B \wedge (\neg p' \vee \neg q') \wedge$$
$$\quad\quad ((p'_B \oplus p') \supset (p_B \oplus p)) \wedge$$
$$\quad\quad ((q'_B \oplus q') \supset (q_B \oplus q))) \supset$$
$$\quad\quad ((p'_B \oplus p') \equiv (p_B \oplus p)) \wedge$$
$$\quad\quad ((q'_B \oplus q') \equiv (q_B \oplus q)))).$$

In this example, each propositional variable varies over F and T. Each tuple of the truth-value assignment for (p'_B, q'_B, p', q') makes the conditional part of the

last conjunct false except (T,T,T,F), (T,T,F,T) and (T,T,F,F).

(1) If (T,T,T,F) is assigned to (p'_B, q'_B, p', q'), then we can derive $(q_B \oplus q) \supset (p_B \equiv p)$ from the last conjunct.

(2) If (T,T,F,T) is assigned to (p'_B, q'_B, p', q'), then we can derive $(p_B \oplus p) \supset (q_B \equiv q)$ from the last conjunct.

(3) If (T,T,F,F) is assigned to (p'_B, q'_B, p', q'), then we can derive T from the last conjunct.

Then the above formula is reduced to:

$$\exists p_B \exists q_B($$
$$\quad p_B \wedge q_B \wedge (\neg p \vee \neg q) \wedge$$
$$\quad ((q_B \oplus q) \supset (p_B \equiv p)) \wedge$$
$$\quad ((p_B \oplus p) \supset (q_B \equiv q))).$$

Then, each tuple of the truth-value assignment for (p_B, q_B) makes the above formula false, except (T,T). Then the above formula is reduced to:

$$(\neg p \vee \neg q) \wedge (p \vee q),$$

which is equivalent to the result in the previous section.

The second example shows inference from the belief.

Example 2:

$P = b, f$
$B(P) \equiv \forall x(b(x) \supset f(x)),$
$K(P) \equiv T,$
$\alpha(P) \equiv b(A).$

$B(P)$ expresses that every bird flies, and $\alpha(P)$ expresses that A is a bird. In this case, $\alpha(P)$ is consistent with $B(P)$, therefore, the new belief, $NB(P)$ is $B(P) \wedge \alpha(P)$ by proposition 2. Belief, that A flies, can be derived from the new belief. This example shows that if the added knowledge is consistent with the current belief, we can infer normal results from the current belief and the added knowledge.

The next example shows treatment of the counterexample to the belief which was discussed in the introduction.

Example 3:

$P = b, f$
$B(P) \equiv \forall x(b(x) \supset f(x)) \wedge b(A),$
$K(P) \equiv b(A),$
$\alpha(P) \equiv \neg f(A).$

$B(P)$ expresses that every bird flies and A is a bird, and $\alpha(P)$ expresses that A does not fly. In fact, $B(P)$

is the new belief of example 2. Then the new belief is defined as follows.

$NB((b, f)) \equiv$
$\exists b_B \exists f_B ($
$\quad \forall x(b_B(x) \supset f_B(x)) \wedge b_B(A) \wedge b(A) \wedge \neg f(A) \wedge$
$\quad \forall b'_B \forall f'_B \forall b' \forall f'((\,$
$\qquad \forall x(b'_B(x) \supset f'_B(x)) \wedge b'_B(A) \wedge b'(A) \wedge \neg f'(A) \wedge$
$\qquad \forall x((b'_B(x) \oplus b'(x)) \supset (b_B(x) \oplus b(x))) \wedge$
$\qquad \forall x((f'_B(x) \oplus f'(x)) \supset (f_B(x) \oplus f(x)))) \supset$
$\qquad (\forall x((b'_B(x) \oplus b'(x)) \equiv (b_B(x) \oplus b(x))) \wedge$
$\qquad \forall x((f'_B(x) \oplus f'(x)) \equiv (f_B(x) \oplus f(x)))))).$

Let $b'_B(x)$ and $f'_B(x)$ and $b'(x)$ be identical to $\mathbf{T}$ and let $f'(x)$ be identical to $x \neq A$.

Then the left-handside of the last conjunct of $NB((b, f))$ becomes as follows:

$\mathbf{T} \supset \mathbf{T} \wedge \mathbf{T} \wedge \mathbf{T} \wedge \neg(A \neq A) \wedge$
$\forall x((\mathbf{T} \oplus \mathbf{T}) \supset (b_B(x) \oplus b(x))) \wedge$
$\forall x((\mathbf{T} \oplus x \neq A) \supset (f_B(x) \oplus f(x)))$

We can easily see that all conjuncts are true, assuming $\forall x(b_B(x) \supset f_B(x))$ and $b_B(A)$ and $\neg f(A)$. Therefore, the left-handside of the last conjunct of $NB((b, f))$ is true, assuming the other conjuncts of $NB((b, f))$. Thus, we can derive the following from $NB((b, f))$.

$\exists b_B \exists f_B ($
$\quad \forall x(b_B(x) \supset f_B(x)) \wedge b_B(A) \wedge b(A) \wedge \neg f(A) \wedge$
$\quad \forall x((\mathbf{T} \oplus \mathbf{T}) \equiv (b_B(x) \oplus b(x))) \wedge$
$\quad \forall x((\mathbf{T} \oplus x \neq A) \equiv (f_B(x) \oplus f(x)))).$

The second conjunct from the last is equivalent to:

$\forall x(b_B(x) \equiv b(x)),$

and the last conjunct is equivalent to:

$(f_B(A) \oplus f(A)) \wedge \forall x(x \neq A \supset (f_B(x) \equiv f(x))).$

Therefore, the above formula is reduced to:

$\exists b_B \exists f_B (\forall x(b(x) \supset f_B(x)) \wedge b(A) \wedge \neg f(A) \wedge$
$\quad \forall x(x \neq A \supset (f_B(x) \equiv f(x)))).$

Since we can derive $\forall x(f_B(x) \supset (x \neq A \supset f(x)))$ from the last conjunct, we can derive the following from the above formula:

$\forall x(b(x) \supset (x \neq A \supset f(x))) \wedge b(A) \wedge \neg f(A),$

which is equivalent to:

$\forall x(x \neq A \equiv (b(x) \supset f(x))).$

While $f(A)$ was true in the previous belief, we can no longer derive $f(A)$ from this new belief. Thus, this example shows nonmonotonicity of minimal belief revision.

And, from the new belief, we can still show that every bird except A flies. This is an effect of minimal belief revision.

5 RELATED RESEARCH

5.1 Formalisms of Nonmonotonic Reasoning

The current formalisms of nonmonotonic reasoning try to define extra inference rules or axioms to produce belief. Default logic [Reiter80] uses special inference rules called *defaults* and circumscription [McCarthy80] adds special axioms to the knowledge. However, if we wish to represent a belief that every bird flies, then we cannot express this belief directly but must modify it to match special mechanisms of the above formalisms. In default logic, we must present the above belief by using extra inference such as:

$$\frac{bird(x) : \mathbf{M} fly(x)}{fly(x)}.$$

In circumscription, we must introduce special predicate ab to express the above belief as:

$$\forall x((bird(x) \wedge \neg ab(x)) \supset fly(x)),$$

and minimize ab. However, in our formalism, belief can be represented directly as:

$$\forall x(bird(x) \supset fly(x)).$$

We use a kind of minimization technique adopted in circumscription to formalize minimal belief revision. While circumscription minimizes predicates to produce a belief, minimal belief revision minimizes the difference between the previous belief and the new belief. However, minimal belief revision inherits the problems of circumscription regarding computability and satisfiability. Since minimal belief revision is defined in second-order language, it is not computable in general. Also, since minimality is based on set-inclusion of the difference between a model of the current belief and a model of the new belief, it can become unsatisfiable if there is an infinite decreasing chain of set difference of models. Therefore, we must investigate useful subclasses of first-order sentences to make minimal belief revision computable and satisfiable as the researchers of circumscription have been studying.

5.2 Truth Maintenance System

In a sense, the formalism in this paper can be regarded as a generalization of the truth maintenance system [Doyle79], because the TMS uses hypothetical con-

texts which correspond to models of belief in our formalism and performs belief revision. However, while the current TMS can only manipulate propositions (or ground sentences), our formalism can manipulate arbitrary sentences. Moreover, TMS uses only one context at one time, whereas we can use multiple contexts at one time because a sentence for belief expresses a set of models.

5.3 Database Updates

In the database community, there have been several reports about research on semantics of updates ([Winslett88] gives a survey for them). For example, [Fagin83 and Kuper84] define minimal updates of syntactic formulas in databases. However, they do not give a model theoretical analysis. And the revised databases may be logically different from the logically equivalent databases when their representations are different.

5.4 Logic of Theory Change

[Alchourrón85] gives formal aspects of the process of contraction and revision of a theory. One of the authors of that paper, Gärdenfors, gives desirable postulates for a revision of a theory. $A \dot{+} x$ is a revised theory if a formula x is added to a theory A (A set of formula A is a *theory* iff $A = Cn(A)$ where Cn is a consequence operation).

(1) $A \dot{+} x$ is a theory.

(2) $x \in A \dot{+} x$.

(3) If $\neg x \notin Cn(A)$, then $A \dot{+} x = Cn(A \cup \{x\})$.

(4) If $\neg x \notin Cn(\phi)$, then $A \dot{+} x$ is consistent under Cn.

(5) If $Cn(x) = Cn(y)$, then $A \dot{+} x = A \dot{+} y$.

And he also gives "supplementary postulates".

(6) $A \dot{+} (x \wedge y) \subseteq Cn((A \dot{+} x) \cup \{y\})$.

(7) If $\neg y \notin A \dot{+} x$, then $Cn((A \dot{+} x) \cup \{y\}) \subseteq A \dot{+} (x \wedge y)$.

To relate our definition of revision with their postulates, we translate them into model-theoretic terms(We do not distinguish A and $Cn(A)$ because they are model-theoretically identical).

(2)' $A \dot{+} x \models x$.

(3)' If $A \not\models \neg x$, then $A \dot{+} x = A \cup \{x\}$.

(4)' If $\not\models \neg x$, then $A \dot{+} x$ is satisfiable.

(5)' If $\models x \equiv y$, then $\models A \dot{+} x \equiv A \dot{+} y$.

(6)' $(A \dot{+} x) \cup \{y\} \models A \dot{+} (x \wedge y)$.

(7)' If $A \dot{+} x \not\models \neg y$, then $A \dot{+} (x \wedge y) \models (A \dot{+} x) \cup \{y\}$.

Knowledge and belief are distinguished in our formalism, but not in theirs. However, when current knowledge K is T and added knowledge α is added to current belief B, then we can compare our definition with theirs. Then postulates (2)', (3)', (5)', (6)' are satisfied, but in general (4)' and (7)' are not satisfied. However if we consider propositional logic, (4)' is satisfied. (7)' is not satisfied because our minimality is based on set-inclusion of the difference between the current belief and the new belief. However, set inclusion of the difference is essential for making our formalism applicable to first-order sentences.

5.5 A Theory of Knowledge Base Revision

[Dalal88] have taken a similar approach to ours because they define minimal change of belief in the model-theoretic terms. The differences between our definition and theirs are the following.

(1) Their definition is for propositional logic; ours is for first-order logic.

(2) In their formalization all belief can be changed; in ours some parts of belief (knowledge) cannot be changed.

(3) While their minimality is based on a number of propositions which receive different interpretations from the current belief and the new belief, our minimality is based on set-inclusion of the difference.

Concerning (3), if the current knowledge is T and the current belief is $p \wedge q \wedge r$ and the added knowledge is $\neg(p \vee q) \vee \neg r$ then, our definition gives $(\neg p \wedge \neg q \wedge r) \vee (p \wedge q \wedge \neg r)$ as a new belief, but their definition gives $p \wedge q \wedge \neg r$. So in a sense, their formalization is more conservative than ours. However, it seems difficult to extend their definition to first-order logic because it is difficult to define the difference of models by number in an infinite domain.

6 CONCLUSION

This paper presents a formalism of nonmonotonic reasoning by direct representation of belief and belief revision. Belief is defined as a detailed description of knowledge so that it entails knowledge. Belief revision occurs when belief does not entail added knowledge. This paper concentrates on a particular belief revision strategy called *minimal belief revision*. Minimal belief revision treats the counter-example for the previous belief as an exception in order to maintain consistency. It also keeps

what was true in the previous belief as far as possible. This paper presents the definition and the model theory for minimal belief revision. We plan the following future work.

(1) As mentioned in 5.1, We need to find useful sub-classes of first-order sentences to make minimal belief revision computable and satisfiable.

(2) If the initial beliefs are conflicting (for example, rules of flying birds and non-flying penguins), minimal belief revision cannot decide which belief should be changed when contradiction occurs. To solve this problem, we should introduce a hierarchical structure into belief.

(3) It should be better if we have a mechanism to produce initial belief from known facts so that we need not find initial belief by ourselves. The mechanism should have a learning ability.

(4) If a lot of exceptions are found, then belief should be changed drastically so that exceptions are converted to defaults. In that case, we need another criterion to perform such revision.

ACKNOWLEDGMENTS

REFERENCES

[Alchourrón85] Alchourrón, C., Gärdenfors, P., Makinson, D., "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions", *The Journal of Symbolic Logic*, Vol. 50, No.2, (1985), pp510–530

[Dalal88] Dalal, M., "Investigation into a Theory of Knowledge Base Revision: Preliminary Report", *Proc. of AAAI-88* (1988), pp475–479

[Doyle79] Doyle, J., "A Truth Maintenance System", *Artificial Intelligence*, Vol. 12 (1979), pp231–272

[Fagin83] Fagin, R., Ullman, J. D. and Vardi, M. Y., "On the Semantics of Updates in Databases", *Proc. of 2nd ACM Symp. on the Principles of Database Systems* (1983), pp352–356

[Hintikka62] Hintikka, J., "Knowledge and Belief", Cornell University Press (1962)

[Kuper84] Kuper, G. M., Ullman, J. D. and Vardi, M. Y., "On the Equivalence of Logical Databases", *Proc. of 3rd ACM Symp. on the Principles of Database Systems* (1984), pp221–228

[McCarthy80] McCarthy, J., "Circumscription - a Form of Non-monotonic Reasoning", *Artificial Intelligence*, Vol. 13 (1980), pp27–39

[McDermott80] McDermott, D. and Doyle, J., "Non-monotonic Logic I", *Artificial Intelligence*, Vol. 13 (1980), pp41–72

[Reiter80] Reiter, R., "A Logic for Default Reasoning", *Artificial Intelligence*, Vol. 13 (1980), pp81–132

[Winslett88] Winslett, M., "A Framework for Comparison of Update Semantics (Extended Abstract)", *Proc. of 7th ACM Symp. on the Principles of Database Systems* (1988), pp315–324

APPENDIX: PROOFS OF PROPOSITIONS

Proposition 1. *Let $B(P)$, $NB(P)$, $K(P)$ and $\alpha(P)$ be the same sentence in the definition of minimal revised belief. M_{NB} is a model of $NB(P)$ iff M_{NB} is a minimal revised model with respect to $B(P)$ and $K(P) \wedge \alpha(P)$.*

Proof:

($\Rightarrow$) Suppose that M_{NB} is a model of $NB(P)$, but for every model, M_B, of $B(P)$, (M_B, M_{NB}) is not a minimally different pair with respect to $B(P)$ and $K(P) \wedge \alpha(P)$.

In other words, for every pair (M_B, M_{NB}), there is a pair of models (M'_B, M'_{NB}) for $B(P)$ and $K(P) \wedge \alpha(P)$ such that $(M'_B, M'_{NB}) \preceq (M_B, M_{NB})$ and not $(M_B, M_{NB}) \preceq (M'_B, M'_{NB})$.

Therefore

$$\forall x((< x > \in M'_B[p_1] \oplus < x > \in M'_{NB}[p_1]) \supset$$
$$(< x > \in M_B[p_1] \oplus < x > \in M_{NB}[p_1])) \wedge ... \wedge$$
$$\forall x((< x > \in M'_B[p_n] \oplus < x > \in M'_{NB}[p_n]) \supset$$
$$(< x > \in M_B[p_n] \oplus < x > \in M_{NB}[p_n])),$$

and

$$\neg ($$
$$\forall x((< x > \in M_B[p_1] \oplus < x > \in M_{NB}[p_1]) \supset$$
$$(< x > \in M'_B[p_1] \oplus < x > \in M'_{NB}[p_1])) \wedge ... \wedge$$
$$\forall x((< x > \in M_B[p_n] \oplus < x > \in M_{NB}[p_n]) \supset$$
$$(< x > \in M'_B[p_n] \oplus < x > \in M'_{NB}[p_n]))).$$

Let $M_{NB}[p_{Bi}] = M_B[p_i](1 \leq i \leq n)$, where p_{Bi} is a predicate constant which is not in P. Then since $M_B \models B(P)$, $M_{NB} \models B(P_B)$, where $P_B = (p_{B1}, ..., p_{Bn})$. Similarly, let $M_{NB}[p'_{Bi}] = M'_B[p_i]$ and $M_{NB}[p'_i] = M'_{NB}[p_i]$, then $M_{NB} \models B(P'_B)$ where $P'_B = (p'_{B1}, ..., p'_{Bn})$, and

$M_{NB} \models K(P') \wedge \alpha(P')$ where $P' = (p'_1, ..., p'_n)$.

And since

$\forall x((< x > \in M_{NB}[p'_{B1}] \oplus < x > \in M_{NB}[p'_1]) \supset$
$(< x > \in M_{NB}[p_{B1}] \oplus < x > \in M_{NB}[p_1])) \wedge ... \wedge$
$\forall x((< x > \in M_{NB}[p'_{Bn}] \oplus < x > \in M_{NB}[p'_n]) \supset$
$(< x > \in M_{NB}[p_{Bn}] \oplus < x > \in M_{NB}[p_n])),$

and

$\neg($
$\forall x((< x > \in M_{NB}[p_{B1}] \oplus < x > \in M_{NB}[p_1]) \supset$
$(< x > \in M_{NB}[p'_{B1}] \oplus < x > \in M_{NB}[p'_1])) \wedge ... \wedge$
$\forall x((< x > \in M_{NB}[p_{Bn}] \oplus < x > \in M_{NB}[p_n]) \supset$
$(< x > \in M_{NB}[p'_{Bn}] \oplus < x > \in M_{NB}[p'_n]))),$

by substituting $M_{NB}[p_{Bi}]$ for $M_B[p_i](1 \leq i \leq n)$, $M_{NB}[p'_{Bi}]$ for $M'_B[p_i]$, and $M_{NB}[p'_i]$ for $M'_{NB}[p_i]$ respectively in the above statement, for any tuples of extensions, P_B, satisfying $B(P_B)$, there exist P'_B and P' satisfying $B(P'_B)$ and $K(P') \wedge \alpha(P')$ respectively such that $M_{NB} \models (P'_B, P') \prec (P_B, P)$.

It contradicts the fact that there exists P_B satisfying $B(P_B)$, such that

$M_{NB} \models \neg \exists P'_B \exists P'(B(P'_B) \wedge K(P') \wedge \alpha(P') \wedge (P'_B, P') \prec (P_B, P)))$

($\Leftarrow$) Suppose that M_{NB} is a minimal revised model and $M_{NB} \models \forall P_B((B(P_B) \wedge K(P) \wedge \alpha(P)) \supset \exists P'_B \exists P'(B(P'_B) \wedge K(P') \wedge \alpha(P') \wedge (P'_B, P') \prec (P_B, P)))$.

In other words, for any P_B satisfying $B(P_B)$, we can take P'_B and P' such that $B(P'_B) \wedge K(P') \wedge \alpha(P') \wedge (P'_B, P') \prec (P_B, P)$.

Then since $M_{NB} \models (P'_B, P') \prec (P_B, P)$,

$\forall x((< x > \in M_{NB}[p'_{B1}] \oplus < x > \in M_{NB}[p'_1]) \supset$
$(< x > \in M_{NB}[p_{B1}] \oplus < x > \in M_{NB}[p_1])) \wedge ... \wedge$
$\forall x((< x > \in M_{NB}[p'_{Bn}] \oplus < x > \in M_{NB}[p'_n]) \supset$
$(< x > \in M_{NB}[p_{Bn}] \oplus < x > \in M_{NB}[p_n])),$

and
$\neg($
$\forall x((< x > \in M_{NB}[p_{B1}] \oplus < x > \in M_{NB}[p_1]) \supset$
$(< x > \in M_{NB}[p'_{B1}] \oplus < x > \in M_{NB}[p'_1])) \wedge ... \wedge$
$\forall x((< x > \in M_{NB}[p_{Bn}] \oplus < x > \in M_{NB}[p_n]) \supset$
$(< x > \in M_{NB}[p'_{Bn}] \oplus < x > \in M_{NB}[p'_n]))).$

We take M_B, M'_B and M'_{NB} such that they have the same domain as M_{NB}, and every constant and function receives the same interpretation in M_B, M'_B and M'_{NB} as in M_{NB}, and the interpretations of predicates of P in M_B, M'_B and M'_{NB} receive the interpretations of predicates of P_B, P'_B and P' in M_{NB} respectively.
Since $M_{NB} \models B(P_B) \wedge B(P'_B) \wedge K(P') \wedge \alpha(P')$, $M_B \models B(P)$, $M'_B \models B(P)$ and $M'_{NB} \models K(P) \wedge \alpha(P)$.

And since

$\forall x((< x > \in M'_B[p_1] \oplus < x > \in M'_{NB}[p_1]) \supset$
$(< x > \in M_B[p_1] \oplus < x > \in M_{NB}[p_1])) \wedge ... \wedge$
$\forall x((< x > \in M'_B[p_n] \oplus < x > \in M'_{NB}[p_n]) \supset$
$(< x > \in M_B[p_n] \oplus < x > \in M_{NB}[p_n])),$

and

$\neg($
$\forall x((< x > \in M_B[p_1] \oplus < x > \in M_{NB}[p_1]) \supset$
$(< x > \in M'_B[p_1] \oplus < x > \in M'_{NB}[p_1])) \wedge ... \wedge$
$\forall x((< x > \in M_B[p_n] \oplus < x > \in M_{NB}[p_n]) \supset$
$(< x > \in M'_B[p_n] \oplus < x > \in M'_{NB}[p_n]))),$

by substituting $M_B[p_i](1 \leq i \leq n)$ for $M_{NB}[p_{Bi}]$, $M'_B[p_i]$ for $M_{NB}[p'_{Bi}]$, and $M'_{NB}[p_i]$ for $M_{NB}[p'_i]$ respectively in the above statement, for M_{NB} and any models M_B satisfying $P(B)$, there exists a pair of models (M'_B, M'_{NB}), satisfying $P(B)$ and $K(P) \wedge \alpha(P)$ respectively such that $(M'_B, M'_{NB}) \preceq (M_B, M_{NB})$ and not $(M_B, M_{NB}) \preceq (M'_B, M'_{NB})$.
It contradicts the fact that M_{NB} is a minimal revised model. **QED**

Proposition 2. *Let $B(P)$, $NB(P)$, $K(P)$ and $\alpha(P)$ be the same sentences in the definition of minimal belief revision. If $\alpha(P)$ is consistent with $B(P)$, $NB(P) \equiv B(P) \wedge \alpha(P)$.*

Proof:

Let M be any model of $B(P) \wedge \alpha(P)$. Since $B(P) \models K(P)$, M is a model of $K(P) \wedge \alpha(P)$. Let us consider a pair of models for $B(P)$ and $K(P) \wedge \alpha(P)$, (M, M). It is a minimally different pair because for any pair of models of $B(P)$ and $K(P) \wedge \alpha(P)$, (M', M''), $(M, M) \preceq (M', M'')$, that is there is no pair of models (M', M'') such that $(M', M'') \preceq (M, M)$ and not $(M, M) \preceq (M', M'')$.

If M is a model of $K(P) \wedge \alpha(P)$, but not $B(P) \wedge \alpha(P)$, we can show that it is not a minimal revised model. Suppose that it is a minimal revised model. Then there is a model of $B(P)$, M', such that (M, M') is a minimally different pair. Since M is not a model of $B(P)$, M and M' are not identical. Let M'' be a model of $B(P) \wedge \alpha(P)$. Then $(M'', M'') \preceq (M, M')$ and not $(M, M') \preceq (M'', M'')$ because M and M' are not identical. It contradicts the fact that (M, M') is a minimally different pair. Thus, M is not a minimal revised model.

Therefore, a set of all models of $B(P) \wedge \alpha(P)$ is equivalent to a set of all minimal revised models, that is, a set of all models of $NB(P)$. In other words, $NB(P) \equiv B(P) \wedge \alpha(P)$. **QED**