The Generalized Theory of Model Preference (Preliminary Report)

Piotr Rychlik
Institute of Computer Science
Polish Academy of Sciences
PKiN, 00–901 Warsaw, POLAND

Abstract

In this paper we present a purely semantic view on non-monotonic reasoning. We follow the direction pointed in [16] and claim that any non-monotonic logic can be viewed as a result of transforming some base standard logic by a selection strategy defined on models. The generalized theory of model preference is shortly outlined here together with its use in modeling non-monotonic beliefs.

Introduction

One of the most serious and, at the same time, common problems encountered in implementing knowledge-based systems is that it is usually unfeasible to provide complete knowledge on which the system is supposed to operate. The ability to fill up the gaps in incomplete information is one of the factors characterizing common sense reasoning. This is common sense, which, in many circumstances, enables people to "jump to conclusions" and solve problems that could never be solved by a perfect, but purely deductively reasoning agent. Deductive reasoning is *monotonic*, because with a larger set of premises it is possible to conclude more facts. Common sense reasoning is *non-monotonic* — new facts may cause previously derived beliefs to be withdrawn.

There have been proposed many formalizations of non-monotonic reasoning: *non-monotonic logic*, *autoepistemic logic*, *default logic*, *circumscription*, and many others. The landmark papers in the discipline of non-monotonic reasoning can be found in [6].

Any non-monotonic logic can be viewed as a result of transforming some base standard logic by a selection strategy defined on models. The selection strategy is supposed to choose those models that, possibly, best explain a theory described in the base logic. In other words, it selects models that are, in some sense, more "preferred" than others. The selection strategy can be defined in terms of a binary relation $\mathcal{P}$ defined on interpretations of the base logic.

It is generally assumed that $\mathcal{P}$ is an ordering relation, either a strict partial order [16] or a quasi-order [3]. In the case of a strict partial order $\mathcal{P}$, the intuitive meaning of $\langle M_1, M_2 \rangle \in \mathcal{P}$ is that M_2 is more preferred than M_1 or it is better than M_1. If $\mathcal{P}$ is a quasi-order, then $\langle M_1, M_2 \rangle \in \mathcal{P}$ is read that M_2 is at least as good as M_1. The selection strategy simply picks up the maximal (which here means "the best") elements of $\mathcal{P}$, if such elements exist, that is, those models $M^\star$ for which there is no model M such that $\langle M^\star, M \rangle \in \mathcal{P}$.

We claim that in many circumstances $\mathcal{P}$ does not need to be an ordering relation in order to be able to capture the idea of preference. In the next section we will examine one example and show that a *preference relation* which is not assumed to be transitive and hence, which is not an ordering relation, may still make sense.

Motivations

Our example will be related to temporal reasoning. We will present a variation of the famous case of *temporal projection* discussed in [7] and commonly known as the *Yale shooting problem*. However, it seems that the problem we are going to present here is very general and does not characterize only the domains making reference to temporal information.

The temporal projection problem arises whenever, given an initial description of the world, a reasoning agent tries to determine which facts are true and which are false after some sequence of events has occurred.

To describe the problem we we will adopt *situational calculus* [12], which was chosen in [7]. There are three types of objects that we will be considering: *situations*, *facts* (also called *propositional fluents*) and *events*. A situation is meant to denote an interval of time when none of the facts changes its truth value. Events (or actions) may change the world assigning new values to fluents. We will write $Holds(f, s)$, if a fact f is true in a situation s, and $result(e, s)$ to denote the new situation that results from an event e taking place in a situation s. In order to be able to effectively represent an "inertia" of the world, and get rid of the *frame problem* [12], we will use a technique proposed in [13]. We introduce a special *abnormality predicate ab* that, for a given triple (f, e, s), will be satisfied if and only if an event e occurring in a situation s changes the value of a fact f; in this case a fact f is said to be *abnormal* with respect to an event e in a situation s. We also accept a non-monotonic inference rule which allows to conclude that a given fact is not abnormal unless its abnormality can be deduced (monotonically) from the currently available data. With these assumptions the single *frame axiom*

$$\forall f, e, s.\ Holds(f, s) \wedge \neg ab(f, e, s) \supset Holds(f, result(e, s)) \tag{1}$$

will suffice to express what otherwise would usually require a great many of axioms.

The Yale shooting scenario, only slightly modified, is as follows. First, in the initial situation S_0, someone loads the gun aimed at Fred (a *LOAD* event). This brings about the situation S_1. Subsequently, after a sequence of *WAIT* events, where nothing interesting happens, in a situation S_{n-1} the gun is fired (a *SHOOT* event), which yields the new situation S_n. Suppose that in the situation S_0 Fred is alive (*ALIVE*), loading a gun causes it to be loaded (*LOADED*), and firing a loaded gun at someone causes that person to become dead (*DEAD*). The knowledge about our domain can be characterized in a natural way by the following set of axioms [7]:

$$Holds(ALIVE, S_0), \tag{2}$$

$$\forall s.\ Holds(LOADED, result(LOAD, s)), \tag{3}$$

$$\begin{aligned}\forall s.\ & Holds(LOADED, s) \supset \\ & ab(ALIVE, SHOOT, s) \wedge \\ & Holds(DEAD, result(SHOOT, s)).\end{aligned} \tag{4}$$

The question is whether the non-monotonic mechanism described above allows us to conclude from (1)-(4) that in the situation S_n Fred is dead. Surprisingly, the answer to this question is negative. Let us recall that our non-monotonic inference rule tries to minimize the interpretation of the abnormality predicate. The answer that Fred is dead after shooting can be obtained if we chronologically minimize the abnormality predicate. The *LOAD* event causes the gun to be loaded in S_1 (3), and, since waiting has no particular effect, the gun remains loaded in the situation S_{n-1}. Hence, by (4), Fred becomes dead in S_n, which agrees with our intuition. But it is possible that first we have applied the rule of minimization in the situation S_{n-1} assuming $\neg ab(ALIVE, SHOOT, S_{n-1})$, which, together with (1), supports that Fred is alive in S_n. However, we know from (4) that the loaded gun causes the predicate *ALIVE* to be abnormal with respect to the *SHOOT* event in any situation. The only explanation of Fred being alive in S_n is that the gun somehow has become unloaded as a result of the *WAIT* event. We must, therefore, assume $ab(LOADED, WAIT, S_i)$ for some situation S_i, where $0 < i < n - 1$.

There appeared great many of solutions to the Yale shooting problem [1,2,5,8,9,10,11,14,16, and many others]. They either reformulate the domain description so that the intended prioritization of models can be captured by the properties of some well known standard non-monotonic mechanisms, or they live the description unchanged introducing new formalisms with some other preference criteria.

Our non-monotonic mechanism, which was unable to deal properly with the problem of temporal projection, imposed certain ordering on models of the domain description. We expected to find a correct answer to our example by picking up the maximal elements of this ordering, namely, the models that minimized the extension of the abnormality predicate. In all of the above solutions there are also used certain model ordering methods to fix the most preferred interpretations. However, it is not difficult to imagine a situation in which there is not enough information provided to construct a reasonable ordering of models, except, perhaps, the one that would make every model the most preferred interpretation.

Let us suppose that there is some process going on which may cause the effect of the gun being unloaded (without actually unloading it). For example, it might be continuous corrosion of the metal parts of the gun, or graceful degradation of the explosive material in the cartridge, etc. If we had no idea when, more or less, this process has started and/or how fast it is progressing, then the conclusion that Fred remains alive in S_n would seem to be equally justified as the one postulating his death after shooting. Suppose we know that besides the process which eventually might prevent killing Fred, nothing unexpected can happen between S_0 and S_n, that is, no action is performed in parallel with those that we have already mentioned. It might be the case that our restricted knowledge about the process allows us to think that if it is possible to kill Fred in a situation S_i, then it can be believed that this would also be possible in S_{i+k}, provided that the time elapsed between S_i and S_{i+k} is not too long. For example, one can safely assume that if it is known that in the situation S_8 it is possible to shoot Fred, then after five waiting events, each of which lasting a second, in the situation S_{13} shooting will also be successful.

Translating these intuitions into a preference relation on models it means that at any moment we prefer those models which allow the effects of the process to appear as late as it is only possible within certain limits. Considering the time interval corresponding to a situation S_i we may prefer the effects of the process to appear no sooner than in the time interval corresponding to a situation S_{i+k+1}. (Here, the silent, simplifying assumption has been made that the time intervals corresponding to situations are of equal length.)

In accordance with the above observations, the preference relation $\mathcal{P}$ can be defined in the following way. Let us take that $\langle M_1, M_2 \rangle \in \mathcal{P}$ if and only if $\mathcal{P}$ satisfies the following condition. The extension of the abnormality predicate for the events *WAIT* and *LOAD* in M_2 is a subset of such an extension in M_1. Otherwise, if these extensions are equal, then for every i and j such that $i < j \leq i+k$, $M_1 \in \mathcal{M}(t_i)$ and $M_2 \in \mathcal{M}(t_j)$, where $\mathcal{M}(t)$ denotes the class of models that satisfy the effects of the process in the time interval t, and t_i and t_j denote the time intervals corresponding to situations s_i and s_j, respectively. It is easy to see that such defined preference relation is not transitive, although it is locally transitive within the limits set by the constant k.

We may wonder whether the preference relation defined above can be expressed using any of the existing non-monotonic formalisms such as circumscription or default logic. It seems that the answer to this question is negative.

All versions of circumscription limit the instances which satisfy a selected predicate (or a set of predicates) in a given theory only to those that are necessary in light of this theory. The preference criterion is therefore defined by the set inclusion relation so as to minimize the extensions of the chosen set of predicates.

Inference mechanism in default logic can be explained in a similar way.[1] Suppose $\Delta = (T, D)$ is a default theory. A default rule $[\alpha : \beta/\gamma] \in D$, informally speaking, allows us to assume that the *conclusion* γ is true, if the *prerequisite* α is assumed to be true, and the *justification* β is consistent with all facts that have already been assumed.[2] Applying any default $\delta_1 = [\alpha : \beta/\gamma] \in D$ whose conclusion does not follow directly from T causes the class $\mathcal{M}_{\{\}}$ of all models of T to be narrowed to the class $\mathcal{M}_{\{\delta_1\}}$ of models of T that satisfy β and γ. If we further apply another default rule $\delta_2 \in D$, the class $\mathcal{M}_{\{\delta_1\}}$ will be narrowed to the class $\mathcal{M}_{\{\delta_1,\delta_2\}}$ which contains the models that additionally satisfy the justification and the conclusion of δ_2. And so on. The preference criterion is, again, defined in terms of the set inclusion relation. We have: $\mathcal{M}_{\{\}} \supset \mathcal{M}_{\{\delta_1\}} \supset \mathcal{M}_{\{\delta_1,\delta_2\}} \supset \ldots$. This, however, is a strict partial ordering relation on models. Since the preference relation we defined for the modified version of the Yale shooting problem is not even transitive, the existing schemes of non-monotonic reasoning cannot express it.

In [4] it is claimed that permitting a preference relation to be intransitive implies irrationality and "wreak havoc on the semantics of the resulting non-monotonic logic". It seems, however, that it is not the case. Suppose that an agent is willing to admit that M_2 is better than M_1 and M_3 is better than M_2, and considering M_3 to be the most preferred of these three models only when some new information which is not a non-monotonic conclusion is provided falsifying M_1, but still satisfying M_2 and M_3. Such behavior does not suggest irrationality of the agent, althouhg the preference criteria are intransitive.

General approach to non-monotonic reasoning[3]

Given some base standard logic $\mathcal{L}$, we can use a preference relation $\mathcal{P}$ to modify the notions of logical satisfiability, validity and entailment in $\mathcal{L}$, defining in this way a new logic $\mathcal{L}_{\mathcal{P}}$. The semantics of $\mathcal{L}_{\mathcal{P}}$ we are going to establish is very similar (actually, is inspired by) the one proposed in [16] and then generalized in [3]. The main departure from the above mentioned formalism is that a preference relation is not assumed to have any particular property and that it heavily depends on some consistent theory written in the base logic $\mathcal{L}$, and can be understood only in the context of this theory.

First-order predicate calculus is assumed as a standard logic in the following definitions. However, this choice is not crucial. It is easy to adopt these definitions to other formalisms such as modal or higher-order logics.

DEFINITION 1. Let T be a theory and $\mathcal{P}_T$ a binary relation defined over the set of models of T. We say, then, that $\mathcal{P}_T$ is a *preference relation depending on T*.

We will usually drop the index denoting a theory on which a preference relation depends when it is not confusing.

DEFINITION 2. Let $\mathcal{P}$ be a preference relation depending on a theory T. Any non-empty (possibly infinite) sequence $s = \langle M_1, M_2, \ldots \rangle$ of models of T is called a *$\mathcal{P}$-sequence* over T if and only if for every M_i and M_j,

[1] A detailed discussion on these issues can be found in [16].

[2] Actually we should talk about some instances of α, β and γ, if these formulae contain free variables. We may, however, make a simplyfying assumption that every default of D is closed, that is, it does not contain a formula with free variables. Generalization to the case where defaults are allowed to contain free variables is obvious.

[3] The proofs of all cited theorems can be found in [15].

where $i < j$, $\langle M_i, M_j \rangle \in \mathcal{P}$, and there is no model $M \notin s$ of T such that for every $M_k \in s$, $\langle M_k, M \rangle \in \mathcal{P}$. A $\mathcal{P}$-sequence s is *bounded*, if, additionally, there is a model $M^\star \in s$ such that for every $M_k \in s$, $M_k = M^\star$ or $\langle M_k, M^\star \rangle \in \mathcal{P}$. In this case we say that $M^\star$ is an *upper bound* of s. A $\mathcal{P}$-sequence with no upper bound is *unbounded.*

Definition 3. Let $\mathcal{P}$ be a preference relation depending on a theory T. A model M of T *$\mathcal{P}$-satisfies* a formula α, written $M \models_{\mathcal{P}} \alpha$, if and only if there is a $\mathcal{P}$-sequence over T such that M is its upper bound and $M \models \alpha$. In this case we say that M is a *$\mathcal{P}$-model* of α.

If M is a $\mathcal{P}$-model, then, of course, M is a model of a theory T on which the preference relation $\mathcal{P}$ depends. However, M is an upper bound of some $\mathcal{P}$-sequence over T. Hence, M is also a $\mathcal{P}$-model of T.

Definition 4. Let $\mathcal{P}$ be a preference relation depending on a theory T. We say that a formula α is *$\mathcal{P}$-satisfiable* if and only if there is a $\mathcal{P}$-model of α.

Clearly, $\mathcal{P}$-satisfiable formulae are also satisfiable. The converse might not be true. If a formula is satisfiable, it does not automatically mean that it has to be satisfied by some of the models of the theory T and, in particular, by some $\mathcal{P}$-model of T (if such a $\mathcal{P}$-model exists at all).

Our modified definition of satisfiability, unlike the Shoham's definition of *preferential satisfiability* [16], makes an explicit reference to some theory T. Shoham considers an interpretation M as a *preferred model* of some formula α if and only if there is no other model M' of α that would be strictly better than M in the sense determined by an ordering relation $\mathcal{P}$ defined over the set of all interpretations of some given logical language. Let $\mathcal{P}_{\{\alpha\}}$ be a restriction of $\mathcal{P}$ only to the models of α. With this assumption, M is a preferred model of α if and only if M is a $\mathcal{P}_{\{\alpha\}}$-model of α. In other words, a formula α is $\mathcal{P}_T$-satisfiable if and only if there is preferential model of $T \cup \{\alpha\}$. The notion of preferential satisfiability is a special case of the satisfiability introduced in Definition 3 and Definition 4.

Definition 5. Let $\mathcal{P}$ be a preference relation depending on a theory T. We say that a formula α is *$\mathcal{P}$-valid*, written $\models_{\mathcal{P}} \alpha$, if and only if T has a $\mathcal{P}$-model and α is satisfied by every $\mathcal{P}$-model of T.

Shoham defines his notion of *preferential validity* in a roundabout way. He considers a formula α to be *preferentially valid* if and only if a formula $\neg\alpha$ is not preferentially satisfiable. This makes possible situations in which a formula α is preferentially valid although it is not even preferentially satisfiable, and α and $\neg\alpha$ are both preferentially valid. Our notion of $\mathcal{P}$-validity does not suffer from this drawback, since we can talk about $\mathcal{P}$-validity only if the preference relation distinguishes at least one model of the theory on which it depends as the most preferred, that is, a $\mathcal{P}$-model of this theory. This approach has a very simple and intuitive explanation. If the relation $\mathcal{P}$ is not sufficient to express the preferences accordingly to which formulae should be assigned truth values, considering the formulae that are satisfied by all preferred models does not make much sense.

Proposition 1. If a formula α is $\mathcal{P}$-valid, then $\neg\alpha$ is not $\mathcal{P}$-satisfiable.

Unfortunately, the converse is not true. It may happen that every $\mathcal{P}$-sequence is unbounded. In this case, both α and $\neg\alpha$ are not $\mathcal{P}$-satisfiable. It is clear that a $\mathcal{P}$-valid formula need not be valid, because it has to be satisfied only by models that are upper bounds of $\mathcal{P}$-sequences. If there is no bounded $\mathcal{P}$-sequences, a valid formula is not $\mathcal{P}$-valid, although it is satisfied by all interpretations. With the additional restriction of a preference relation $\mathcal{P}$, which says that there must exist at least one bounded $\mathcal{P}$-sequence, the following propositions hold.

Proposition 2. Let $\mathcal{P}$ be a preference relation depending on a theory T, and there be a bounded $\mathcal{P}$-sequence over T. Then a formula α is $\mathcal{P}$-valid if and only if $\neg\alpha$ is not $\mathcal{P}$-satisfiable.

Proposition 3. Let $\mathcal{P}$ be a preference relation depending on a theory T, and there be a bounded $\mathcal{P}$-sequence over T. Then a formula α is $\mathcal{P}$-valid if α is valid.

Definition 6. Let $\mathcal{P}$ be a preference relation depending on a theory T. We say that $\mathcal{P}$ is *complete*, if for every formula α consistent with T, every model of α is in some bounded $\mathcal{P}$-sequence over T whose upper bound satisfies α.

Completeness is a very strong notion. Together with the emptiness of a theory on which a preference relation $\mathcal{P}$ depends, it implies that the resulting logic $\mathcal{L}_{\mathcal{P}}$ is monotonic.

Proposition 4. Let T be an empty theory, and $\mathcal{P}$ a complete preference relation depending on T. Then a formula α is satisfiable if and only if α is $\mathcal{P}$-satisfiable.

Proposition 5. Let T be an empty theory, and $\mathcal{P}$ a complete preference relation depending on T. Then a formula α is valid if and only if α is $\mathcal{P}$-valid.

Definition 7. Let $\mathcal{P}$ be a preference relation depending on a theory T. We say that α *$\mathcal{P}$-entails* β, written $\alpha \models_{\mathcal{P}} \beta$, if and only if every $\mathcal{P}$-model of α is also a $\mathcal{P}$-model of β.

The above definition corresponds to the Shoham's definition of *preferential entailment*, which says that α

preferentially entails β if and only if every preferential model of α is also a model of β. He does not require α to be preferentially satisfied by preferential models of α. We can also substitute the requirement that every $\mathcal{P}$-model of α is a $\mathcal{P}$-model of β by the requirement that every $\mathcal{P}$-model of α is a model of β. Let us note, however, that this substitution does not change the notion of $\mathcal{P}$-entailment. If β has a model M and this model is a $\mathcal{P}$-model of α, then M is a $\mathcal{P}$-model of β. In fact, the notion of preferential entailment is a special case of the $\mathcal{P}$-entailment. If $\mathcal{P}$ is an ordering relation defined over the set of all interpretations of some given logical language and $\mathcal{P}_{\{\alpha\}}$ its restriction to the models of α, then α preferentially entails β (in the sense determined by $\mathcal{P}$) if and only if α $\mathcal{P}_{\{\alpha\}}$-entails β.

Not surprisingly $\mathcal{P}$-entailment, just as preferential entailment introduced by Shoham, satisfies the following propositions:

PROPOSITION 6. Let $\mathcal{L}_{\mathcal{P}}$ be a $\mathcal{P}$-logic, and α, β and γ three formulae in it. Then, if $\alpha \wedge \beta \models_{\mathcal{P}} \gamma$, then also $\alpha \models_{\mathcal{P}} \beta \supset \gamma$.

PROPOSITION 7. If $\mathcal{L}_{\mathcal{P}}$ is a $\mathcal{P}$-logic, then $\mathcal{L}_{\mathcal{P}}$ is monotonic if and only if for all formulae $\alpha, \beta, \gamma \in \mathcal{L}_{\mathcal{P}}$, if $\alpha \models_{\mathcal{P}} \beta \supset \gamma$, then also $\alpha \wedge \beta \models_{\mathcal{P}} \gamma$.

Writing down the axioms of some theory, we usually have in mind one particular (real or imaginary) world that we want to formalize. This world is the intended model of the theory. Of course, our theory, if it is only consistent, has also other models (actually, infinitely many of them). All the models that agree with the intended model on how every formula is assigned a truth value are considered the most preferred. In fact, we will be recognizing a model as the most preferred, if it only agrees with the intended model on a fixed subset of formulae that we think is important for some reasons.

The preference relation must, therefore, be defined with respect to some partial interpretation. Any two models will be compared accordingly to how well they match the requirements about the valuation of the formulae determined by this partial interpretation. These requirements can be viewed as a function, which we will call a *preference rule*, that for a given formula and a valuation of variables states which logical value is preferred. Since it would be much easier to compare models if they had the same domains, we will be comparing them through the *corresponding Herbrand interpretations.* [4]

DEFINITION 8. Let A be a set of formulae and $As(H_T)$ a set of valuations of variables over the Herbrand universe of some theory T. A *preference rule* for A over T is any function $\pi : A \times As(H_T) \longrightarrow \{true, false\}$.

DEFINITION 9. Let M_1 and M_2 be models of a theory T, and H_1 and H_2 Herbrand interpretations of T corresponding to M_1 and M_2, respectively. M_1 is *preferred* over M_2 with respect to a preference rule $\pi : A \times As(H_T) \longrightarrow \{true, false\}$ if and only if for every formula $\alpha \in A$ and every valuation $v \in As(H_T)$, if $Val^v_{H_2}(\alpha) = \pi(\alpha, v)$, then also $Val^v_{H_1}(\alpha) = \pi(\alpha, v)$. M_1 is *strictly preferred* over M_2 with respect to π if additionally there are $\beta \in A$ and $v \in As(H_T)$ such that $Val^v_{H_1}(\beta) = \pi(\beta, v)$ and $Val^v_{H_2}(\beta) \neq \pi(\beta, v)$. $Val^v_M(\alpha)$ stands for the value of a formula α in an interpretation M and a valuation v.

DEFINITION 10. M_1 is *strongly preferred* over M_2 with respect to a set Π of preference rules if and only if for every $\pi \in \Pi$, M_1 is preferred over M_2 with respect to π.

DEFINITION 11. M_1 is *weakly preferred* over M_2 with respect to a set Π of preference rules if and only if there is a subset $X \subset \Pi$ such that M_1 is strongly preferred over M_2 with respect to X.

The notion of weak preference enables us to define a preference relation which is not transitive.

Let us return to the Yale shooting example. The preference relation for this example can be defined using the notion of weak preference. First, however, let us rewrite the axiom (4) to reflect changes in our domain description.

$$\begin{aligned} &\forall s,t.\ Holds(LOADED, s) \wedge \\ &time(s) = t \wedge Noeffects(t) \supset \\ &\quad ab(ALIVE, SHOOT, s) \wedge \\ &Holds(DEAD, result(SHOOT, s)) \end{aligned} \tag{5}$$

If *Noeffects*(t) is true, it means that the effects of the process have not emerged in a time interval t. Function *time* maps situations into corresponding time intervals. We need also following axioms to tie time intervals to situations.

$$time(S_0) = t_0, \tag{6}$$

$$\forall e, s.\ time(result(e, s)) = time(s) + 1. \tag{7}$$

Let A denote the set of axioms (1)–(3) and (5)–(7), and H_A the Herbrand universe of A. Assume that for any time interval t, Π contains a preference rule

[4] A *Herbrand interpretation* is any interpretation with a Herbrand universe as its domain. A *Herbrand universe* H_T of a theory T is the set of all well-formed expressions that can be built using the function and constant symbols that appear in the *skolemized* axioms of T. If there is no constant symbol, one such symbol is introduced. We say that two interpretations *correspond* to each other if and only if every formula is assigned the same truth value in these interpretations.

π_t whose value is true for any formula in B_t and any valuation in $As(H_A)$, where

$$B_t = \{\neg ab(f, WAIT, s), \neg ab(f, LOAD, s)\} \cup \\ \{Noeffects(t), \ldots, Noeffects(t+k)\} \cup \\ \{\neg Noeffects(t+k+1), \neg Noeffects(t+k+2), \ldots\}.$$

It is easy to see that the weak preference over Π reflects the intuition that among the models in which effects of the process have not emerged in a time interval t, those models should be preferred that allow the effects to appear after $t+k$, but not sooner. If we know nothing about the time the process has started, $\neg Noeffects(time(s))$ should be true in every situation s and, therefore, the conclusion that Fred is dead after shooting is blocked. Knowing that $Noeffects(time(S_0))$ is true, we are allowed to conclude that $\neg Noeffects(time(S_0) + k)$ is true. If k is sufficiently big, that is $k \geq n$, Fred can be believed to die in the final situation.

Conclusion

In this paper we outlined a generalized theory of model preference, following the idea that any nonmonotonic logic can be viewed as a result of transforming some base standard logic by a selection strategy defined on models. Our approach is more general than the one presented in [16], because a preference relation is parameterized by some theory written in the base logic. Moreover, we are not making any assumption about properties of preference relations. This allows us to model the situations in which our knowledge is too restricted to build a reasonable and justified ordering of models of the domain description. It seems also that the notions of $\mathcal{P}$-satisfiability and $\mathcal{P}$-validity are defined more clearly than the corresponding notions of preferential satisfiability and preferential validity introduced by Shoham.

References

[1] Baker, A.B. 1989. A simple solution to the Yale shooting problem. Proc. First International Conference on Principles of Knowledge Representation and Reasoning, pp. 11–20.

[2] Baker, A.B. & Ginsberg, M.L. 1989. Temporal projection and explanation. Proc. IJCAI, 906–911.

[3] Brown, A.L.,Jr., Shoham, Y. 1989. New results on semantical nonmonotonic reasoning, *Lecture Notes in Artificial Intelligence* **346**:19–26.

[4] Doyle, J & Wellman, M.P. 1989. Impediments to universal preference-based default theories. Proc. First International Conference on Principles of Knowledge representation and Reasoning, 94–102.

[5] Gelfond, M. 1988. Autoepistemic logic and formalization of commonsense reasoning: preliminary report, *Lecture Notes in Artificial Intelligence* **346**:176–186.

[6] Ginsberg, M.L. 1988. *Readings in Nonmonotonic Reasoning*, Los Altos, CA: Morgan Kaufmann.

[7] Hanks, S. & McDermott, D. 1987. Nonmonotonic logic and temporal projection, *Artificial Intelligence* **33**(3):379–412.

[8] Haugh, B.A. 1987. Simple causal minimizations for temporal persistence and projection. Proc. AAAI, 218–223.

[9] Kautz, H.A. 1986. The logic of persistence. Proc. AAAI, 401–405.

[10] Lifschitz, V. 1987. Pointwise circumscription, *Readings in Nonmonotonic Reasoning*, (ed M.L. Ginsberg), 179–193. Los Altos, CA: Morgan Kaufmann.

[11] Lifschitz, V. 1987. Formal theories of action (Preliminary report). Proc. IJCAI, 966–972.

[12] McCarthy, J. & Hayes, P.J. 1969. Some philosophical problems from the standpoint of artificial intelligence, *Machine Intelligence* **4**:463–502.

[13] McCarthy, J. 1986. Applications of circumscription to formalizing common sense knowledge, *Artificial Intelligence* **28**(1):89–116.

[14] Morgenstern, L. & Stein, L.A. 1988. Why things go wrong: A formal theory of causal reasoning. Proc. AAAI, 518–523.

[15] Rychlik, P. 1989. Semantic considerations on non-monotonic reasoning, Technical Report, **674**, Institute of Computer Science. Polish Academy of Sciences.

[16] Shoham, Y. 1987. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* Cambridge, MA: MIT Press.