Baltzer Journals

Applying the Persistent Set Approach in Temporal Reasoning

Yan Zhang[1] and Norman Y. Foo[2]
1 School of Information Technology,
Northern Territory University,
Casurania, NT 0811, Australia
E-mail: yan@it.ntu.edu.au

[2] *Basser Department of Computer Science,*
University of Sydney, NSW 2006, Australia
E-mail: norman@cs.su.oz.au

Since Hanks and McDermott raised the problem of temporal projection (e.g., the Yale shooting problem) and showed that classical nonmonotonic logics failed to solve it, many solutions have been proposed. However, as indicated by some researchers, most of them are not completely satisfactory. In [22] we presented a theory of actions called the *Persistent Set Approach* (PSA). In this paper, we extend our previous work to deal with temporal reasoning. Different from those minimality-based approaches, we propose a persistence-based formalization of actions within the situation calculus framework, and show that this gives natural and intuitive solutions to the problem of temporal projection in many cases. Explanations of some of the differences between persistence and minimality are given. We show that our approach also provides a unified framework for representing actions with disjunctive effects, while most of the current methods are inappropriate for dealing with these actions in the general case.

Keywords: disjunctive effects of actions, frame and ramification problems, persistence, reasoning about action, temporal projection problem

1 – Introduction

Since Hanks and McDermott raised the problem of *temporal projection* (e.g., the Yale shooting problem) and showed that classical nonmonotonic logics failed to solve it [6], many solutions have been proposed [10, 19, 11, 12, 8, 1, 2]. Although there are differences among these approaches, most of them employ some *minimality* principle in their formalisms.

Generally, there are two ways to deal with the problem. One is to extend the expressive power of classical nonmonotonic logics, or define some new minimality criteria within them so as to get the expected result [10, 11, 12, 8, 1, 2]. The other way is to build a new logic system to provide a general principle for temporal reasoning rather than only for some specific domain and the temporal projection problem can be solved within this new logic framework [19]. Many of the first kind of remedy have been criticized by some researchers as lacking intuitive motivation and generality [7, 17]. If so, it would seem that the second way is more attractive. However, for a theory of action, it is important to characterize what kind of problems can be solved by this theory. This is sometimes ignored by researchers.

Lin and Shoham noticed the difficulty and tried to define a precise criterion for the adequacy of theories of action. They proposed a *provably-correct* approach to formalize the effect of actions in the situation calculus and argued that their work leads to a better understanding of the formal temporal reasoning [14]. However, as will be shown latter, Lin and Shoham's approach is weaker than they expected. The main limitation is that they only represented the *direct* effect of an action, and ignored the possibility that an action may also have *implicit* effects.

In [22] we proposed a theory of actions called the *Persistent Set Approach* (PSA) that presents a method for reasoning about actions with disjunctive effects, as well as actions with conjunctive effects. In this paper, we extend our previous work to deal with temporal reasoning. We will present a new formalization of actions within the situation calculus framework [16]. The main differences between our method and other current theories are: (1) our formalization for temporal reasoning is based on the *persistence principle*, while most of current theories were proposed based on the *minimality principle*; (2) our theory provides a unified framework for representing actions with both conjunctive and disjunctive effects, while other theories are usually inappropriate for representing actions with disjunctive effects.

With the history of logical approaches to temporal projection, it is almost certain that our persistence principle too will have limitations. Indeed we know that in its current form it is not suitable for providing explanations. A deeper reason is that the validation of any particular logical approach can only be provided by separate ontological classifications of dynamic systems, but that is in fact only a very recent and tentative development. In the meantime, we believe that it is important to explore the consequences of the persistence principle and compare it to typical approaches based on minimality principles. In this way, we can accumulate experience that will help formulate the alternative ontologies that these principles are good for. In the concluding section we will outline the features of an ontology for which persistence is appropriate. A formal correspondence will have to await a consensus on how to describe ontologies in a manner acceptable to our community.

The paper is organized as follows. In the following section we present the motivation and background of the PSA. We will discuss the problem with the classical state-based theories of action PMA and PWA, which will then motivate the persistent set approach. We also characterize the difference between the PSA and the PMA. Furthermore, we discuss the satisfiability of Katsuno and Mendelzon's update postulates [9] within the PSA framework. In the third section, we propose a persistence-based formalization for temporal reasoning. This is an extension of the PSA to temporal reasoning. In the fourth section, we address the temporal projection problem, and show that our approach provides an intuitive result. In the fifth section, we extend our persistence-based formulation to represent actions with disjunctive effects. In the sixth section, we compare our work with Lin and Shoham's approach in detail. Finally, we present the conclusion and further work.

2 – Motivation and Background

2.1 – The Problem with the PWA and PMA

Winslett proposed a method for reasoning about change, which she called the *possible models approach* (PMA) [21]. The PMA was devised to remove some difficulties in Ginsberg and Smith's *possible worlds approach* (PWA) [5]. As argued by Winslett, although the PWA is an elegant, simple and powerful technique for reasoning about action, it fails to solve the *frame problem* (what facts about the world remain unchanged when an action is performed?) and the *ramification problem* (what facts about the world must change when an action is performed?) if there exists incomplete information in the world description (a logical theory).

Both the PWA and PMA were proposed based on the principle of *minimal change*, but their criteria of minimal change are different. In the PWA, the minimal change is defined on logical theories, while in the PMA, the minimal change is defined on models of logical theories. Now the question is: does the PMA always give satisfactory results for reasoning about action?

Let us consider a simple example. Suppose in an instructional computer system, all programs have three statuses called *running*, *terminated.* and *abnormally terminated.* At any time point, each program has only one of these statuses. We also assume that for any program the only occurrence of abnormal termination of the program is caused by system errors. The program executes the next instruction on receiving the command "*Continue*". If we treat the command "*Continue*" as an action, then post-condition of $Continue$ is $running \vee terminated$, that is, if there is no system error, after the program executes the next instruction, the status of this program will be *running* if the instruction is not "Stop", otherwise the status will be *terminated.*

Suppose in the initial state, program p is on the running status, and no system error occurs. After action $Continue$ occurs, how can we predict the status of p? Since $Continue$'s post-condition $running \vee terminated$ has been already satisfied in the initial state, according to the PMA or PWA's minimal change principle, nothing will be changed. That is, the program p will be running forever. From our intuition, however, after action $Continue$ occurs, there are two possibilities for program p's status, one is running, the other is terminated. In other words, the status of p should be *indefinite* with respect to action $Continue$.

Why do the PMA and PWA have this problem? The reason is clear: the minimal change principle does not represent the *indefinite effect* of an action properly. In the above example, the post-condition of $Continue$ is a disjunction which presents an indefinite effect of $Continue$. But when we use the PMA (similarly, the PWA) to reasoning about action $Continue$, only the states which satisfy $Continue$'s post-condition and have minimal difference with the initial state are considered as possible resulting states. Thus, the state in which p's status is *terminated* is excluded. From this observation, we argue that the assumption underlying the minimal change principle is sometimes too strong.

2.2 – A Persistent Set Approach (PSA): Overview

To solve the problem of the minimal change principle disclosed above, we propose a new principle called the *persistence principle* for reasoning about action and change. Generally, the persistence principle can be described as follows:

> *a fact in a state of the world will persist during a state transition if it is not logically relevant to those facts that must change or may be subject to change because of the performance of an action on this state.*

While the intuitive meanings of *logical relevance* and of *must change* and *may change* are usually clear, to formalize this principle we must give explicit representations and explanations for these concepts.

Consider a first order language with equality $\mathcal{L}$ but *without* function symbols. Let S be an interpretation of $\mathcal{L}$, and $|S|$ the universe of S. To simplify our discussion in the rest of this section, we *fix* the universe of interpretations, that is, for any two interpretations S_1, S_2 of $\mathcal{L}$, $|S_1| = |S_2|$, and each constant of $\mathcal{L}$ has the same interpretation in S_1 and S_2 respectively. For simplicity, we also assume that each constant of $\mathcal{L}$ is assigned to " itself" in the universe.

We use an interpretation S of the language $\mathcal{L}$ to represent the state of the world and assume that there is a set of closed first order formulas C to represent state constraints about the world, which should be satisfied by each state of the world, i.e., $S \models C$. An action A is specified by a pre-condition $Pre(A)$ and a post-condition $Post(A)$, where $Pre(A)$ and $Post(A)$ are closed first order formulas. We say that A is *applicable* to a state S if $S \models Pre(A)$. Informally, action A changes the state of the world from one to another, and in the resulting state $Post(A)$ is satisfied. We also assume that the pre-condition $Pre(A)$ is *complete*, in the sense that the action can definitely be performed if the pre-condition is satisfied. The post-condition, however, will not be assumed complete, since there may be many ramifications of an action in any particular situation.

Suppose that A is an action defined by pre- and post-conditions $Pre(A)$ and $Post(A)$ respectively, and S_0 the initial state to which A is applicable. First, for any literal $f \in S_0$, if f's negation is entailed by the post-condition of A (or together with C), i.e., $\{Post(A)\} \cup C \models \neg f$, then f clearly must change its truth value after A is performed. We call this kind of facts *non-persistent*.

On the other hand, suppose that a ground clause $f_1 \vee \cdots \vee f_k$, which does not include any subclause like $f \vee \neg f$, is *non-trivially* entailed by $\{Post(A)\} \cup C$ where each literal or its negation occurs in S_0, that is, $\{Post(A)\} \cup C \models \bigvee_{i \in N} f_i$ where $N = \{1, \cdots, k\}$, $C \not\models \bigvee_{i \in N} f_i$ and f_i or $\neg f_i$ in S_0, and for any proper subset M of N, $\{Post(A)\} \cup C \not\models \bigvee_{j \in M} f_j$. In this case, the truth value of each f_i is indefinite in the resulting state and we say that f_i is *indefinitely affected* by A. In other words, fact f_i may or may not be changed by the performance of A but we do not know which is the case. Conservatively, it is not safe to claim that f_i persists with respect to A under this situation, and therefore we say this kind of facts are *mutable* with respect to A.

Finally, consider a literal f in S_0 which has at least one support set in S_0. Suppose for every *minimal support set* F of f in S_0, that is, $F \neq \emptyset$, $f \notin F$, $F \cup C \models f$, and for any *proper* subset F' of F, $F' \cup C \not\models f$, there exists some $f' \in F$ which is non-persistent or mutable with respect to A. Thus, the performance of A may cause a loss of the justification of f. We also call such facts *mutable* with respect to A.

Now the persistence principle can be simply stated as follows: a fact *persists* with respect to action A if it is neither non-persistent nor mutable. The following definition formalizes these descriptions.

Definition 1
Suppose action A is applicable to S_0, i.e., $S_0 \models Pre(A)$. Let

1. $\Delta_0 = S_0 - \{f \mid f \in S_0$ *and* $\{Post(A)\} \cup C \models \neg f\}$,

2. $\Delta_1 = \Delta_0 - \{f_1, \cdots, f_k \mid \{Post(A)\} \cup C \models \bigvee_{i=1}^{k}[\neg]f_i$ *where* $\{f_1, \cdots, f_k\} \subseteq S_0$, *there does not exist subclause* $f \vee \neg f$ *in* $\bigvee_{i=1}^{k}[\neg]f_i$, $C \not\models \bigvee_{i=1}^{k}[\neg]f_i$ *and for any proper subset* M *of* $N = \{1, \cdots, k\}$ $\{Post(A)\} \cup C \not\models \bigvee_{j \in M}[\neg]f_j\}$,

3. $\Delta_i = \Delta_{i-1} - \{f \mid f \in S_0$, $\{Post(A)\} \cup C \not\models f$, f *has at least one support set in* S_0, *and for* every *minimal support set* F *of* f *in* S_0, *there exists at least one element of* F, f', *which is in* $S_0 - \Delta_{i-1}\}$.

The persistent set $\Delta(S_0, A)$ *of* S_0 *with respect to the action* A *is defined as*
$\Delta(S_0, A) = \bigcap_{i=0}^{\infty} \Delta_i$.
The notation $[\neg]$ *means that the negation sign* $\neg$ *may or may not appear.*

Note that in the above definition, we only consider the indefiniteness of disjunctions and apparently ignore the case of $\{Post(A)\} \cup C \models \exists x.p(x)$, while such implication is non-trivial, i.e., $\{Post(A)\} \cup C \not\models p(c)$ for any constant c, every instance of $\exists x.p(x)$ has a possibility to be true in the resulting state but we do not know which one. In order to reflect this semantics using the definition above, we need the domain closure assumption [18]. Thus, whenever there exists the case of $\{Post(A)\} \cup C \models \exists x.p(x)$, we assume that $\exists x.p(x) \equiv (p(c_1) \vee \cdots \vee p(c_n))$, i.e., every instance of $\exists x.p(x)$ is indefinite with respect to the action A.

Based on the definition of persistent set, we can precisely define the possible resulting states as follows.

Definition 2
We say that an interpretation S *is a* possible resulting state *of the world after action* A *is performed on the state of the world* S_0, *if*

1. $S \models C$,

2. $S \models Post(A)$, *and*

3. $\Delta(S_0, A) \subseteq S$.

The set of all such states of the world resulting from performing action A *on* S_0 *is denoted by* $Result(S_0, A)$.

Generally, given an initial state and an action, we can not predict a unique next state after performing the action, but using the approach described above we may derive a set of possible resulting states.

2.3 – Comparisons with the Minimal Change Principle

Comparison with the PMA

Consider the program termination example described previously once again. Since the post-condition of *Continue* is a disjunction $running \vee terminated$, it follows that both facts *running* and $\neg terminated$ in the initial state are *mutable*. Thus, the only persistent fact in the initial state is that "there is no system error". So, from Definition 2, we get two possible resulting states: one states that program p is still on the running status, the other says that p is terminated.

An action like *Continue* is called *indefinite*. Informally, an *indefinite action* represents an indefinite effect (e.g., its direct or indirect effect is a disjunction)[1]. From the above example, we can see that the PSA can be different from the PMA with respect to the indefinite action, and sometimes the result obtained by the PSA is more reasonable than that obtained by the PMA. In [22], we showed that even for actions with definite effects, the result obtained by the PSA can be different from that obtained by the PMA. For this case, again, sometimes, the PSA represents a *better* solution than the PMA does.

It is important to investigate why there is such difference between the PSA and PMA. Although both the PMA and PSA are state-based methods, the basic motivations within them are quite different. For the PSA, we use the persistence principle, while the PMA is constructed based on the minimal change principle. To compare them properly, we now consider the *mutable sets* obtained from them respectively.

Let S_0 be a state, C the set of state constraints about the world, A an action applicable to S_0, and $\Delta(S_0, A)$ the persistent set of S_0 with respect to A obtained by Definition 1. The *non-persistent set* (i.e., facts that *must change*) of S_0 with respect to A is obtained as follows

$$\Gamma(S_0, A) = \{f \mid f \in S_0 \text{ and } \{Post(a)\} \cup C \models \neg f\}. \tag{1}$$

Obviously, for any $S \in Result(S_0, A)$ and $S' \in Incorporate(S_0, Post(A))^2$, $\Gamma(S_0, A) \cap S = \emptyset$ and $\Gamma(S_0, A) \cap S' = \emptyset$. The mutable set obtained from the PSA is

$$\delta_{\mathrm{PSA}}(S_0, A) = S_0 - \Delta(S_0, A) - \Gamma(S_0, A). \tag{2}$$

Since in the PMA, $\bigcap Incorporate(Post(a), S_0)$ is the set of all facts that are true in every possible model, the mutable set obtained from the PMA is

$$\delta_{\mathrm{PMA}}(S_0, A) = S_0 - \bigcap Incorporate(S_0, Post(A)) - \Gamma(S_0, A). \tag{3}$$

Definition 3
Let S_0 be a state of the world, C the set of state constraints about the world, $F \subseteq S_0$ and $f \in S_0$. We say that F influences *f, if there exists a minimal support set Σ of f with respect to C in S_0, such that $F \subseteq \Sigma$.*

Directly from the above definition, we know that if F influences f, then for any non-empty subset F' of F, F' influences f. We say that F *uniquely influences* f if there exists a *unique* minimal support set Σ of f with respect to C in S_0 such that $F \subseteq \Sigma$. Suppose

[1] We noted that giving formal definitions for definite and indefinite actions is logically difficult, and so leave it as an open problem.

[2] According to the PMA, $Incorporate(S_0, Post(A))$ is defined to be the set of Models where for each model in this set, it satisfies C and $Post(A)$, and has *minimal* difference with S_0.

δ and F be two subsets of S_0, we say that δ *is influenced* by F if for every $f \in \delta$, f is influenced by F. Then we have the following proposition[3].

Proposition 1
Let A be an action applicable to S_0. Then we have:

1. *If $\delta_{\mathrm{PSA}}(S_0, A) - \delta_{\mathrm{PMA}}(S_0, A) \neq \emptyset$, then for any fact $f \in \delta_{\mathrm{PSA}}(S_0, A) - \delta_{\mathrm{PMA}}(S_0, A)$, f is influenced by some subset of $\Gamma(S_0, A)$ or indefinitely affected by A;*
2. *If $\delta_{\mathrm{PMA}}(S_0, A) - \delta_{\mathrm{PSA}}(S_0, A) \neq \emptyset$, then for any fact $f \in \delta_{\mathrm{PMA}}(S_0, A) - \delta_{\mathrm{PSA}}(S_0, A)$, f is* neither *uniquely influenced by any subset of* $\Gamma(S_0, A)$ nor *indefinitely affected by A.*

The above proposition simply says that each mutable fact obtained by the PSA is based on some *logical relevance*, i.e., influence or indefiniteness. However, in the PMA, some fact may be mutable for no reason. This is why the PMA sometimes may have unintended effects in reasoning about actions.

Comparison with Katsuno and Mendelzon's Update Postulates

To characterize the properties of the minimal change principle in propositional knowledge base updates, Katsuno and Mendelzon proposed eight postulates which they claim should be satisfied by any propositional update operators constructed on the minimal change principle [9].

Let ψ be a propositional formula which is used to represent a knowledge base, and μ a propositional formula. Let $\psi \diamond \mu$ denote the update of ψ by μ. Then, Katsuno and Mendelzon's update postulates can be described as follows.

(U1) $\psi \diamond \mu$ implies μ.
(U2) If ψ implies μ then $\psi \diamond \mu \equiv \psi$.
(U3) If both ψ and μ are satisfiable then $\psi \diamond \mu$ is also satisfiable.
(U4) If $\psi_1 \equiv \psi_2$ and $\mu_1 \equiv \mu_2$ then $\psi_1 \diamond \mu_1 \equiv \psi_2 \diamond \mu_2$.
(U5) $(\psi \diamond \mu) \wedge \phi$ implies $\psi \diamond (\mu \wedge \phi)$.
(U6) If $\psi \diamond \mu_1$ implies μ_2 and $\psi \diamond \mu_2$ implies μ_1 then $\psi \diamond \mu_1 \equiv \psi \diamond \mu_2$.
(U7) If ψ is complete then $(\psi \diamond \mu_1) \wedge (\psi \diamond \mu_2)$ implies $\psi \diamond (\mu_1 \vee \mu_2)$.
(U8) $(\psi_1 \vee \psi_2) \diamond \mu \equiv (\psi_1 \diamond \mu) \vee (\psi_2 \diamond \mu)$.

Although the PMA is constructed based on the first order logic, it is not difficult to reduce it to the propositional case. If we do so, as has been shown by Katsuno and Mendelzon, the PMA update operator satisfies all postulates (U1) – (U8).

To compare the PSA with Katsuno and Mendelzon's update postulates, we have to reduce the PSA to the propositional case. Consider a propositional language $\mathcal{L}$. We first consider updating a state S with a propositional formula μ. By Definition 1, we get the persistent set $\Delta(S, \mu)$ of S with respect to μ. Then we say that an interpretation S' of $\mathcal{L}$ is a *possible state* resulting from updating S with μ iff $S' \models C$, $S' \models \mu$, and $\Delta(S, \mu) \subseteq S'$, where C is a set of propositional formulas to represent the state constraints of the world. Denote the set of all possible states resulting from updating S with μ as $Update(S, \mu)$. Then we can define the persistent update operator $\diamond_{psa}$ as follows.

[3]All proofs of propositions that we did not present here appeared in [24].

Definition 4
Let ψ be a knowledge base, i.e., $\psi \models C$, μ a propositional formula. $\psi \diamond_{psa} \mu$ denotes the persistent update *of ψ with μ*[4]*, where*

1. *If ψ implies μ or ψ is inconsistent then $\psi \diamond_{psa} \mu \equiv \psi$, otherwise*
2. *$Models(\psi \diamond_{psa} \mu) = \bigcup_{S \in Models(\psi)} Update(S, \mu)$.*

In the above definition, condition 1 says that if ψ implies μ, then nothing is changed since the knowledge μ has been represented by knowledge base ψ[5]; or if ψ is inconsistent, then any update can not change it into a consistent knowledge base. Condition 2 says that if ψ is consistent and does not imply μ, then ψ should be changed, and this change follows the persistence principle as defined previously. We then have the following result.

Proposition 2
The persistence-based update operator $\diamond_{psa}$ satisfies Katsuno-Mendelzon's postulates (U1) – (U4), (U6) and (U8), but violates (U5) and (U7).

It is not surprising that the update operator $\diamond_{psa}$ does not satisfy (U5) and (U7). As stated by Katsuno Mendelzon, (U5) and (U7) represent a minimal change during the update, and we have shown that in the PSA the state change is not forced to be minimal generally. Instead, we have the following alternative postulates for $\diamond_{psa}$.

Proposition 3
The following results hold.

(U5') *If $\psi \diamond_{psa} \mu \models [\neg]\phi$ but $\psi \not\models \phi$, then $(\psi \diamond_{psa} \mu) \wedge \phi$ implies $\psi \diamond_{psa} (\mu \wedge \phi)$.*
(U7') *If $\psi \diamond_{psa} \mu_1 \models [\neg]\mu_2$ (or $\psi \diamond_{psa} \mu_2 \models [\neg]\mu_1$), then $(\psi \diamond_{psa} \mu_1) \wedge (\psi \diamond_{psa} \mu_2)$ implies $\psi \diamond_{psa} (\mu_1 \vee \mu_2)$.*

3 – A Persistence-based Formalization for Temporal Reasoning

In this section, we extend the PSA to deal with temporal reasoning. The PSA is based on first order logic and the state of the world is presented by an interpretation of the language. In applying the PSA in temporal reasoning we have to introduce the time concept in our formalism. The formalism is based on the situation calculus framework where the persistence principle is introduced for temporal reasoning.

[4] Here we only consider the *well-defined* update, that is, μ is consistent with C.
[5] Note that this is different from our PSA action theory, in which even the post-condition is satisfied in the initial state, the action may still cause a change of the initial state, for instance, the program termination example. We believe that this is one of the differences between action and knowledge.

3.1 – The Language

There are a number of ways to represent time in AI formalisms — continuous point-based models based on the real numberline, discrete time, intervals logics, branching time, and so on. We can omit an account of the differences among these methods because Shoham presented a detailed discussion on this [20]. Instead, we will simply adopt the discrete time-point method as used in the situation calculus [16], as it is a well established AI formalism for reasoning about action and planning.

Let $\mathcal{L}$ be a sorted first order language with equality, with three disjoint sorts for *situation, action*, and *fluent* respectively. Assume $\mathcal{L}$ has the following vocabulary:

1. Sort *situation*: with situation constants $S, S_0, S_1, \cdots$, and situation variables $s, s_1, s_2, \cdots$.
2. Sort *action*: with action constants $A, A_1, A_2, \cdots$, and action variables $a, a_1, a_2, \cdots$.
3. Sort *propositional fluent*: with fluent constants $P, P_1, P_2, \cdots$, and fluent variables $p, p_1, p_2, \cdots$.
4. Function symbol of sort *situation*: the binary function symbol *do* which takes arguments of sorts *action* and *situation* respectively.
5. Predicate symbol: the binary predicate symbol *holds* which takes arguments of *fluent* and *situation* respectively.
6. Logical constants and punctuations: as usual, including equality.

In the situation calculus, a *situation* is the complete state of the universe at an instant of time. The change of a situation is only caused by the performance of an *action* or a sequence of actions. Since the universe is too large for complete description, there is never a complete description of a situation. Instead of that, there are only facts about situations. For example, formulas $holds(P, S)$ and $holds(P, do(A, S))$ mean that fluent P is true in situation S, and fluent P is true in the situation resulting from the performance of action A in situation S respectively, and formula $\forall s.holds(P_1, s) \supset holds(P_2, s)$ means that for any situation s, if fluent P_1 is true in s, then fluent P_2 is also true in it.

In our formalism, an action A is specified by a *causal rule* that has the form:

$$\forall s.Pre(A, s) \supset Post(A, do(A, s)),$$

where there is only one situation variable s, and both $Pre(A, s)$ and $Post(A, s)$ are conjunctive normal forms: $([\neg]holds(P_1, s) \vee [\neg]holds(P_2, s) \vee \cdots) \wedge \cdots \wedge (\cdots)$. The post-condition $Post(A, s)$ presents the *direct effects* of action A.

There is also a set of formulas to represent domain constraints about the world, in which for each constraint there is no occurrence of action and the only variable is situation variable s that is universally quantified. Thus, we can write the set of domain constraints in the form:

$$\forall s.C(s)^6.$$

[6] It can be regarded as a conjunction of all constraints.

3.2 – Primitive Theories of Action

We are now ready to develop our formalization of actions. First we only consider actions with conjunctive effects and later extend our approach to actions with disjunctive effects. We assume that there are no concurrent actions in our world.

Definition 5
Let ACT be a finite set of actions that may occur in our problem domain, and FL be a finite set of fluents in which we are interested[7]*. We say that T is a (finite)* primitive theory *of actions occurring in ACT in situation S_0 iff T is consistent and*

$$T = FACT(S_0) \cup CR \cup DC \cup UN, \text{ where}$$

- *$FACT(S_0)$ is a finite set of basic facts about situation S_0. Each element in $FACT(S_0)$ has the form $[\neg]holds(P, S_0)$ where $P \in FL$.*
- *CR is a set of causal rules. For each $A \in ACT$ there is a causal rule $\forall s.Pre(A, s) \supset Post(A, do(A, s))$.*
- *DC is a finite set of domain constraints about the world. According to our previous discussion, we can write $DC = \{\forall s.C(s)\}$.*
- *UN is a set of* unique names axioms *for individual constants: $S \neq S_0 \neq S_1 \neq \cdots$, $A \neq A_1 \neq \cdots, P \neq P_1 \neq \cdots$.*

Note that given a problem domain, i.e., ACT, FL and DC are fixed, two primitive theories of ACT in situation S_0 may only be different on $FACT(S_0)$. Also, in the primitive theory it is not necessary to represent $FACT(S_0)$ completely. That is, there may exist some $P \in FL$ where neither $holds(P, S_0)$ nor $\neg holds(P, S_0)$ is in $FACT(S_0)$.

3.3 – Persistent Extensions of Primitive Theories

At the initial situation we have a primitive theory which presents the initial facts of the world, as well as the basic information about the world, such as actions, state constraints, etc. Now, the question is: how to represent world changes if actions are performed sequentially beginning from the initial situation and how to predict the resulting state about the world after all actions have occurred?

We will answer these questions using the *persistence principle* developed in earlier sections. The primitive theory of actions T represents a description of the world in the initial situation S_0. We use the persistence principle to construct the next description of the world when a situation is carried to another by an action, and we define the final description to be an *persistent extension* of the primitive theory T, so that from the persistent extension of T, we can infer facts that are true in the resulting situation. Note that here we first consider actions with *only* conjunctive effects. We will extend the following definition by combining actions with disjunctive effects in section 5.

[7] As done by Lin and Shoham [14], we fix a set of fluents.

Definition 6
Let T be a primitive theory of ACT in situation S_0, and $\{A_1, \cdots, A_n\} \subseteq ACT$. Let S_i denote $do(A_i, S_{i-1})$ for $1 \leq i \leq n$. $Persist(T; A_1, \cdots, A_n)$ is a persistent extension *of T with respect to the sequence of actions $A_1, \cdots, A_n$ in situation S_0, if:*

1. $\Sigma_0 = T$,
2. $\Sigma_i = \Sigma_{i-1} \cup \{[\neg]holds(P, S_i) \mid [\neg]holds(P, S_{i-1}) \in \Delta(S_{i-1}, A_i)\}$[8],
3. $Persist(T; A_1, \cdots, A_n) = \Sigma_n$,

where $\Delta(S_{i-1}, A_i)$, the persistent set *in situation S_{i-1} with respect to action A_i, is defined as follows:*

1. $\Delta_0 = \{[\neg]holds(P, S_{i-1}) \mid \Sigma_{i-1} \models [\neg]holds(P, S_{i-1})\}$,
2. $\Delta_1 = \Delta_0 - \{[\neg]holds(P, S_{i-1}) \mid Post(A_i, S_{i-1}) \wedge C(S_{i-1}) \models [\neg\!\!\!\!\!-]holds(P, S_{i-1})$[9]$\}$,
3. $\Delta_k = \Delta_{k-1} - \{[\neg]holds(P, S_{i-1}) \mid Post(A_i, S_{i-1}) \wedge C(S_{i-1}) \not\models [\neg]holds(P, S_{i-1})$, $[\neg]holds(P, S_{i-1})$ *has at least one support set in* Δ_0, *and for* every *minimal support set F of $[\neg]holds(P, S_{i-1})$ in Δ_0, there exists at least one element of F, f, which is in* $\Delta_0 - \Delta_{k-1}\}$,
4. $\Delta(A_i, S_{i-1}) = \bigcap_{k=0}^{\infty} \Delta_k$ *if* $\Sigma_{i-1} \models Pre(A_i, S_{i-1})$, *otherwise* $\Delta(S_{i-1}, A_i) = \Delta_0$.

Now we explain this definition in detail. We use the primitive theory T to represent the description of the world in the initial situation, say S_0. We assume that there is a sequence of actions $A_1, \cdots A_n$ occurring beginning in the situation S_0. In order to predict what facts are true once all these actions have occurred, we define the *persistent extension* of T in a constructive way. In the first part of Definition 6, condition 1 of forming $Persist(T; A_1, \cdots, A_n)$ is obvious. Condition 2 says that when action A_i occurs in situation S_{i-1} every *persistent fact* in Σ_{i-1} is carried to the next situation S_i.

Let us examine the second part of Definition 6. It is easy to see that $Persist(T; A_1, \cdots, A_n)$ depends on the *persistent set* (the second part of the definition). Suppose the current situation is S_{i-1}, and the description of the world in S_{i-1} is Σ_{i-1}[10]. Thus, for any fact in Σ_{i-1}, we say that the fact is *non-persistent* with respect to action A_i if its negation is implied by the post-condition of A_i together with the domain constraints, i.e., condition 2 of the second part of Definition 6. The fact is *implicitly affected* by A_i, if for *each* of its justifications (minimal support sets), some fact in this justification is non-persistent. Recursively, a fact is *implicitly affected* by A_i, if for each of its justification, some fact in the justification is implicitly affected by A_i, i.e., condition 3 of the second part of Definition 6. Finally, a fact in Σ_{i-1} is *persistent* with respect to A_i if it is neither non-persistent nor implicitly affected by A_i. Thus, $\Delta(S_{i-1}, A_i)$ is the set of those facts in Σ_{i-1} that *must* persist, i.e. condition 4 of the second part of Definition 6. One special

[8]This means that if $holds(P, S_{i-1})$ or $\neg holds(P, S_{i-1}) \in \Delta(S_{i-1}, A_i)$, then $holds(P, S_i)$ or $\neg holds(P, S_i) \in \Sigma_i$ respectively.

[9]$[\neg\!\!\!\!\!-]holds(P, S_{i-1})$ represents the negation of $[\neg]holds(P, S_{i-1})$.

[10]In fact, the Σ_{i-1} represents a trace of the descriptions of the world from the initial situation S_0 to the current situation S_{i-1}.

case is that if $\Sigma_{i-1} \not\models Pre(A_i, S_{i-1})$, then nothing will be changed and every fact in Σ_{i-1} persists and is then carried to the next situation S_i.

Thus we can see that $Persist(T; A_1, \cdots, A_n)$ represents a trace of the descriptions of the world from the initial situation S_0 to the final resulting situation S_n. Every fact $[\neg]holds(P, S_n)$ satisfying $Persist(T; A_1, \cdots, A_n) \models [\neg]holds(P, S_n)$ is an answer to temporal projection (see below). For any fluent P such that $Persist(T; A_1, \cdots, A_n) \not\models [\neg]holds(P, S_n)$, we can regard the set of models of $Persist(T; A_1, \cdots, A_n)$ as representing an indefiniteness of P's truth value in situation S_n.

There are some interesting properties of persistent extensions.

Proposition 4
(**Nonmonotonicity**)
Let T_1 and T_2 be two primitive theories of a set of actions ACT in situation S_0, and $\{A_1, \cdots, A_n\} \subseteq ACT$. Then $T_1 \subseteq T_2$ does not imply

$$Persist(T_1; A_1, \cdots, A_n) \subseteq Persist(T_2; A_1, \cdots, A_n).$$

Proof
We need only give one counter-example. Suppose there are domain constraints as follows:

$$\forall s. \neg holds(P_1, s) \supset \neg holds(P_2, s),$$
$$\forall s. holds(P_2, s) \supset holds(P_3, s),$$

and the only action occurring in the world is A, whose causal rule is:

$$\forall s. \neg holds(P_1, do(A, s)).$$

Now we consider two primitive theories T_1 and T_2 of A in situation S_0 such that $T_1 \subseteq T_2$, where

$$FACT(S_0)_{T_1} = \{holds(P_1, S_0),\ holds(P_3, S_0)\},$$
$$FACT(S_0)_{T_2} = \{holds(P_1, S_0),\ holds(P_2, S_0), holds(P_3, S_0)\}.$$

According to the definition of persistent extension, it is easy to show that:

$$holds(P_3, do(A, S_0)) \in Persist(T_1; A), \text{ but}$$
$$holds(P_3, do(A, S_0)) \notin Persist(T_2; A).$$

□

Proposition 5
Let $Persist(T; A_1, \cdots, A_n)$ be the persistent extension of T, φ a formula. Then

$$Persist(T; A_1, \cdots, A_n) \models \varphi$$

iff there exists some i ($1 \leq i \leq n$) such that $\Sigma_i \models \varphi$.

Proof
Directly from Definition 6. □

4 – Temporal Projection and the Yale Shooting Problem

In this section, we show how our theory presented in previous sections can be applied in temporal reasoning, especially for solving the problem of *temporal projection*.

Informally, the temporal projection problem asks, given an initial description of the world and the occurrence of some actions, what facts are true once all these actions have occurred [7]? Obviously, in order to answer the question, we have to know not only what facts will become true after the performance of the actions, but also what facts *persist* during the occurrence of the actions. Generally, the direct effect of an action is specified by the causal rules of the action, from which we can infer what facts will become true as caused by the actions. Thus, the problem becomes how to specify the persistent facts formally. This is the so called *frame problem* [16].

A typical example of the temporal projection problem is called the Yale shooting problem proposed by Hanks and McDermott [7]. Using the Yale Shooting Problem, Hanks and McDermott showed that classical nonmonotonic logics failed to solve the temporal projection problem in temporal reasoning. Since then, many solutions have been proposed to the problem [10, 19, 11, 12, 8, 1, 2]. As we mentioned earlier, although there are differences among these approaches, most of them employ some *minimality* principle in their formalisms. Now we show how our approach provides a solution for the Yale shooting problem.

In the Yale shooting problem, there are three actions called $Load$, $Wait$, and $Shoot$. In the initial situation S_0, Fred is $Alive$, and the gun becomes $Loaded$ any time a $Load$ action happens. At any time, Fred becomes dead after he is shot with the loaded gun. Then the problem is that: in situation S_0, $Load$ is performed, then $Wait$ for a while, and then $Shoot$ Fred, does Fred become dead after the three actions have occurred?

The problem is represented in our formalism as follows. There are two causal rules for actions $Load$ and $Shoot$ respectively:

$$\forall s.holds(Loaded, do(Load, s)), \tag{4}$$

$$\forall s.holds(Loaded, s) \supset \neg holds(Alive, do(Shoot, s)), \tag{5}$$

and one basic fact

$$holds(Alive, S_0). \tag{6}$$

Then the primitive theory of actions $\{Load, Wait, Shoot\}$ in situation S_0 is

$$T = \{(4), (5), (6)\} \cup UN^{11}. \tag{7}$$

Since there is no explicit causal rule for $Wait$, i.e., no pre-condition and also no direct effect for $Wait$, in this case, we take both pre- and post-conditions of $Wait$ to be tautologies.

In order to predict what fact becomes true after the sequence of actions $Load$, $Wait$, $Shoot$ has occurred, we construct the persistent extension of T, i.e., $Persist(T; Load, Wait, Shoot)$. Let $S_1 = do(Load, S_0)$, $S_2 = do(Wait, S_1)$, and $S_3 = do(Shoot, S_2)$. Then, in accordance with Definition 6, we have

[11] Note that there is no specific constraint in this problem domain.

$\Sigma_0 = T$,
$\Sigma_1 = \Sigma_0 \cup \{holds(Alive, S_1)\}$, where $\Delta(S_0, Load) = \{holds(Alive, S_0)\}$,
$\Sigma_2 = \Sigma_1 \cup \{holds(Alive, S_2),\ holds(Loaded, S_2)\}$,
where $\Delta(S_1, Wait) = \{holds(Alive, S_1),\ holds(Loaded, S_1)\}$,
$\Sigma_3 = \Sigma_2 \cup \{holds(Loaded, S_3)\}$, where $\Delta(S_2, Shoot) = \{holds(Loaded, S_2)\}$,
$Persist(T; Load, Wait, Shoot) = \Sigma_3$.

Then, it is not difficult to prove that

$$Persist(T; Load, Wait, Shoot) \models \neg holds(Alive, S_3).$$

Now we consider the ramification problem related to the Yale shooting problem presented by Baker [2]. Suppose we extend the domain of the Yale shooting problem by adding the fluent $Walking$ and a domain constraint stating that in order to be walking, one must be alive:

$$\forall s.holds(Walking, s) \supset holds(Alive, s). \tag{8}$$

Suppose also that in addition to the initial situation S_0 we know that Fred is originally walking:

$$holds(Walking, S_0). \tag{9}$$

Baker argued that after extending the problem domain, the method of causal minimization [11] can not correctly represent the *implicit effect* of actions *Load, Wait*, and *Shoot* that not only does shooting make Fred not alive, it also makes him not walking. This problem can be easily avoided in our approach. Let

$$T' = T \cup \{(8), (9)\}, \tag{10}$$

then it is easy to prove that

$Persist(T'; Load, Wait, Shoot) \models \neg holds(Alive, S_3) \wedge \neg holds(Walking, S_3)$,
and
$Persist(T'; Load, Wait, Shoot) \models holds(Walking, S_1) \wedge holds(Walking, S_2)$.

5 – Representing Actions with Disjunctive Effects

As we have discussed before, one of the difficult problems in most of current theories of action is to deal with actions with *disjunctive effects*. In section 2, we have seen that the PWA and PMA fail to solve the frame and ramification problems properly if these actions occur. Other AI formalisms of temporal reasoning, on the other hand, have difficulties in expressing actions with disjunctive effects. For example, in Shoham's chronological ignorance logic, the direct effect of action may only be expressed by an *atomic knowledge sentence* [19].

It is not difficult to extend our formulation presented in previous sections to represent actions with disjunctive effects. Consider such an action A. We say that in situation S, a fact is *indefinitely affected* by A, if it or its negation appears in a non-trivial disjunction $\varphi(S)$ where $\models \forall s.Post(A, s) \wedge C(s) \supset \varphi(s)$. A fact is *implicitly affected* by A, if for *each* of its justifications (minimal support sets), some fact in the justification is non-persistent or indefinitely affected by A. Recursively, a fact is *implicitly affected* by A, if for every one of its justifications, some fact in the justification is implicitly affected by A. We call a fact *mutable* if it is either implicitly or indefinitely affected by A. Thus, a fact is *persistent* with respect to A if it is neither non-persistent nor mutable. Now we extend Definition 6 as follows.

Definition 7
Let T be a primitive theory of ACT in situation S_0, and $\{A_1, \cdots, A_n\} \subseteq ACT$. Then the persistent extension of T with respect to $A_1, \cdots, A_n$ in situation S_0 is defined similarly to Definition 6, the only exception being the definition of the persistent set. Here we define the persistent set $\Delta(S_{i-1}, A_i)$ as follows:

1. *Δ_0 and Δ_1 are defined the same as conditions 1 and 2 respectively in the second part of Definition 6,*

2. *$\Delta_2 = \Delta_1 - \{[\neg]holds(P, S_{i-1}) \mid$ there exists fluents $P_1, \cdots, P_m$, where $[\neg]holds(P_j, S_{i-1}) \in \Sigma_{i-1}$ for each $j \in M = \{1, \cdots, m\}$, such that $Post(A_i, S_{i-1}) \wedge C(S_{i-1}) \models [\neg]holds(P, S_{i-1}) \vee \bigvee_{j \in M}[\neg]holds(P_j, S_{i-1})$ but $C(S_{i-1}) \not\models [\neg]holds(P, S_{i-1}) \vee \bigvee_{j \in M}[\neg]holds(P_j, S_{i-1})$ and $Post(A_i, S_{i-1}) \wedge C(S_{i-1}) \not\models \bigvee_{j \in M}[\neg]holds(P_j, S_{i-1})$, and for any proper subset M' of M, $Post(A_i, S_{i-1}) \wedge C(S_{i-1}) \not\models [\neg]holds(P, S_{i-1}) \vee \bigvee_{h \in M'}[\neg]holds(P_h, S_{i-1})\}$,*

3. *Δ_k is defined the same as condition 3 in the second part of Definition 6,*

4. *$\Delta(A_i, S_{i-1})$ is defined the same as condition 4 in the second part of Definition 6.*

Now let us consider an example which involves an action with an indefinite (disjunctive) effect.

Example 1
Consider a simple baby's room domain in which there are three fluents: $Ontable(Block1)$, $Ontable(Block2)$, and $Safetobaby$. We have a constraint to state that "if both $Block1$ and $Block2$ are on the table, they are safe to baby, and *vice versa*". This is represented in our formalism in the form:

$$\forall s.holds(Ontable(Block1), s) \wedge holds(Ontable(Block2), s) \equiv holds(Safetobaby, s). \quad (11)$$

Suppose in the current situation S_0, there are three basic facts:

$$\begin{aligned} FACT(S_0) = \{&\neg holds(Ontable(Block1), S_0),\ holds(Ontable(Block2), S_0), \\ &\neg holds(Safetobaby, S_0)\}. \end{aligned}$$

Now suppose the robot executes an action $Move(Block1)$ to move $Block1$ on the table, while the causal rule of the action is:

$$\forall s.holds(Ontable(Block1), do(Move(Block1), s)). \quad (12)$$

The problem is, after action $Move(Block1)$, what facts are true in situation $do(Move(Block1), S_0)$?

Let T be a primitive theory of $Move(Block)$ in situation S_0, where

$$T = \{(11), (12)\} \cup FACT(S_0) \cup UN.$$

Then the persistent extension of T with respect to $Move(Block1)$ is:

$$Persist(T; Move(Block1)) = T.$$

From $Persist(T; Move(Block1))$, we can prove:

$$Persist(T; Move(Block1)) \models holds(Ontable(Block1), do(Move(Block1), S_0)).$$

Now the question is: where is $Block2$? Since

$$\models \forall s. holds(Ontable(Block1), s) \wedge (11) \supset$$
$$\neg holds(Ontable(Block2), s) \vee holds(Safetobaby, s),$$

action $Move(Block1)$ has a disjunctive effect, and hence we only have an indefinite answer for $Ontable(Block2)$ and $Safetobaby$:

$$Persist(T; Move(Block1)) \models \neg holds(Ontable(Block2), do(Move(Block1), S_0) \vee$$
$$holds(Safetobaby, do(Move(Block1), S_0)).$$

There are two models of $Persist(T; Move(Block1))$, in one the fact $Ontable(Block2)$ is false while the fact $Safetobaby$ is true in another. Since there is no more domain information, we can explain that moving $Block1$ on the table may lead $Block2$ to fall down the floor (e.g., the table may have not enough space to place two blocks), or $Block2$ is still on the table and thus they are safe to baby.

6 – Comparison with Lin and Shoham's Approach

As mentioned previously, there are a number of formulations to deal with temporal reasoning in AI currently. It would be impossible to compare our method with *every* other method in detail. Instead, we compare our work with Lin and Shoham's work as a representative example of a recent and efficient approach to temporal reasoning.

To solve the difficulty of temporal reasoning, specifically the temporal projection problem, Lin and Shoham tried to define a precise criterion for adequacy of theories of action. They proposed a *provably-correct* approach to formalize the effects of actions within the situation calculus framework and argued that their work leads to a better understanding of the formal temporal reasoning [14]. In their theory, they identified a formal and intuitive adequacy criterion, proved a certain class of monotonic theories of action adequate relative to this criterion, and then showed an equivalent nonmonotonic counterpart for a subclass of the theories. Now we examine the class of monotonic theories of action that they called *causal theories* in detail.

A causal theory T of an action A consists of three parts: a set of state constraints, a set of causal rules of the action, and a set of frame axioms, where the domain constraints and causal rules are similar to our definitions. The frame axioms are inferred from causal rules and constraints. Formally, suppose FL a fixed set of fluents, and the state constraint and causal rules are presented as follows respectively:

$\forall s.C(s)$,
$\forall s.R_i(s) \supset [\neg]holds(P_i, do(A, s))$, where $1 \leq i \leq n$, and $R_i(s)$ is a formula with a free variable s that represents the pre-condition of A.

Let $N = \{1, \cdots, n\}$. For any $M \subseteq N$, and any $P \in FL$, if

$\not\models \forall s. \bigwedge_{i \in M} holds(P_i, s) \wedge C(s) \supset holds(P, s)$ and
$\not\models \forall s. \bigwedge_{i \in M} holds(P_i, s) \wedge C(s) \supset \neg holds(P, s)$,

then the following are frame axioms:

$$\forall s. \bigwedge_{i \in N-M} \neg holds(R_i, s) \supset (holds(P, s) \equiv holds(P, do(A, s))).$$

For a given causal theory, Lin and Shoham defined a *state* SS of situation S where for each $P \in FL$, $[\neg]holds(P, S) \in SS$ and $SS \models C(S)$. The effect of action A with respect to state SS is then described by $T \cup SS$.

A causal theory T is *epistemologically complete* about action A with respect to FL if it is consistent and for each $P \in FL$, $T \cup SS \models [\neg]holds(P, do(A, S))$ for any situation S and any state SS of S. Lin and Shoham then presented a set of conditions under which the causal theory is epistemologically complete. However, as will be shown next, even in some simple circumstance, actions can not be correctly represented by Lin and Shoham's causal theories.

Example 2
This example was proposed by Lifschitz [13]. Consider a lamp controlled by two switches. The light is on when both switches are turned on the same position. Let On, Sw_1 and Sw_2 be the propositional fluents representing the states of the lamp and of the switches.

In Lin and Shoham's formalism, we have $FL = \{On, Sw_1, Sw_2\}$, and the following state constraint:

$$\forall s.C(s) \equiv (\forall s.holds(On, s) \equiv (holds(Sw_1, s) \equiv holds(Sw_2, s))). \quad (13)$$

Consider action $Turningoff(Sw_1)$ turning off swith Sw_1, whose causal rule is

$$\forall s. \neg holds(Sw_1, do(Turningoff(Sw_1), s)). \quad (14)$$

From the above description, we infer the following frame axioms:

$$\forall s.holds(On, s) \equiv holds(On, do(Turningoff(Sw_1), s))), \text{and} \quad (15)$$

$$\forall s.holds(Sw_2, s) \equiv holds(Sw_2, do(Turningoff(Sw_1), s))). \quad (16)$$

Then the causal theory of action $Turningoff(Sw_1)$ is

$$T = \{(13), (14), (15), (16)\}. \quad (17)$$

Suppose in the initial situation S_0, a state SS of S_0 is

$$SS = \{holds(On, S_0), holds(Sw_1, S_0), holds(Sw_2, S_0)\}. \quad (18)$$

The effect of action $Turningoff(Sw_1)$ is then described by $T \cup SS$. From $\{(15),(16)\}$ $\cup SS$, however, it is easy to see that the light is still definitely on after turning off switch Sw_1. This is clearly contradictory with the state constraint (13).

Why has such inconsistency occurred in Lin and Shoham's theory? The reason is that their declaration for frame axioms embedded in causal theories is incorrect. In particular, the frame axiom defined in the causal theory represents a persistence criterion which roughly says that "if a fact and its negation are not logically entailed by the direct and implicit effect of an action, then this fact is persistent with respect to the action". However, as described above, this persistence criterion is incorrect when the action has an indefinite effect. In the above example, since

$\models \forall s.C(s) \wedge \neg holds(Sw_1, s) \supset (\neg holds(On, s) \vee \neg holds(Sw_2, s))$, and
$\models \forall s.C(s) \wedge \neg holds(Sw_1, s) \supset (holds(On, s) \vee holds(Sw_2, s))$,

action $Turningoff(Sw_1)$ has an *indefinite* effect.

Using our approach, on the other hand, we first construct the primitive theory of $Turningoff(Sw_1)$ as follows:

$$T' = \{(13),(14)\} \cup SS \cup UN. \tag{19}$$

By computing the persistent extension of T' using Definition 7, we have

$$Persist(T'; Turningoff(Sw_1)) = T',$$

from which we derive that after turning off switch Sw_1, the light will be off and the position of switch Sw_2 remains unchanged; or the light remains on while the position of switch Sw_2 changes to off. Obviously, this is an indefinite result, while the first case seems more reasonable from our intuition. However, nothing is wrong with this indefinite result since there is no more domain-dependent information. To derive the preferred solution, we need more information about the domain.

7 – Concluding Remarks

In this paper, we presented a persistence-based formalization of actions within the situation calculus framework. We have shown that our formulation presented a solution to the temporal projection problem. The most different feature between our approach and other current theories is that a conservative persistence principle is employed in our formalization, which provides a unified framework for representing actions with indefinite (disjunctive) effects.

However, as mentioned in Example 2, in order to derive the preferred result in the indefinite reasoning, our method should be extended in some way. One feasible way to extend our method, as we have shown in [23], is to divide all propositional fluents into different layers, such that the propositional fluents in different layers have different persistent priorities during the temporal reasoning. This will solve the problem of deriving the preferred solution as presented in Example 2. On the other hand, since our approach is built based on the conservative persistence principle, as we mentioned earlier, this may lead to undesired solutions in some applications. For instance, in the Yale shooting domain, if we have constraint

$$\forall s.holds(Walking, s) \supset holds(Alive, s),$$

and initially Fred is walking, then stopping walking causes the fact that Fred is alive to be *unknown* (indefinite). In this case, it seems that our approach is too conservative.

On a more foundational ground, we think that in order to handle the indefiniteness properly, we need to distinguish two kinds of domain constraints — constraints describing causal relations and constraints describing general logical invariants. For example, we can treat the above constraint

$$\forall s.holds(Walking, s) \supset holds(Alive, s)$$

as a general logical invariant, and constraint

$$\forall s.holds(Ontable(Block), s) \supset holds(Safetobaby, s)$$

as a causal relation (moving *Block* out of the table may cause that it is no longer safe to baby). Then the conservative persistence principle should be only applied to fluents that involve causal relations and disjunctive effects of actions, while the commonsense law of inertia is applied to other fluents by default.

The other limitation of our approach presented in this chapter is that our method does not "solve the problem of temporal explanation" [10, 2]. In fact, our method does not support backward reasoning. Since we use the primitive theory to present the initial description of the world in which there are no frame axioms and the persistent fact is generated by constructing the persistent extension of the primitive theory, we can not derive what fact held in the initial situation from those facts in the resulting situation. Also, in our formulation, we did not consider concurrent actions. Since our formalism is based on the situation calculus, it is difficult to represent the semantics of concurrent actions clearly within the situation calculus framework [3] although some researchers did some useful work on this [4, 15].

As we have mentioned previously, reasoning about indefinite actions and knowledge is an important issue in knowledge representation and reasoning. The PSA and its extension (presented in this paper) try to build a unified framework for this kind of reasoning. We do not think that our current work solves this problem completely. In fact, there are more interesting questions that we need to solve in the future. For example, in the original PSA formulations, we assume that the state of the world is complete. That is, we can always know if an action A is applicable to a state S by testing $S \models Pre(A)$. If the state of the world is *incomplete*, for example, we may only know that a set of formulas is satisfied in the current state, the situation will then become difficult. For instance, A is an action, $T(S)$ is a set of formulas that is satisfied in the current state S. If neither $T(S) \vdash Pre(A)$ nor $T(S) \vdash \neg Pre(A)$ holds, we will not know if A is applicable to S. Further, if after some time, the current state becomes S', where $T(S') \vdash Post(A)$, but $T(S) \vdash \neg Post(A)$, can we explain that the change of state is *probably* caused by action A? In this case, the occurrence of action A at state S is *indefinite*. In our persistence-based situation calculus formulation, we did not assume the state to be complete, but we ignored the case of indefinite occurrence of actions. This problem together with other related issues will be investigated in our future work.

Finally, the issue we raised at the beginning of a systematic methodology for validating logics of actions by appeal to ontological classes has been addressed in preliminary fashion by Lin and Shoham in their cited work on causal theories. One might ask in a similar vein, for what classes of dynamic systems the persistence principle is appropriate? While a full formal answer would take us too far afield, the informal answer is in fact very intuitive. It is that class in which if there is indefiniteness in state knowledge reflected by some clause, then for each literal in the clause there is a state in which the literal holds, and there is another in which it does not. Indeed this is why the persistence principle is conservative — in case of doubt it does not infer.

Acknowledgements

This research is supported in part by a grant from the Australian Research Council. Most of the research was done while the first author was with the Basser Department of Computer Science at University of Sydney. Our special thanks to Francisco Corella, Fangzhen Lin as well as the referees for many valuable advice and suggestions.

References

[1] A. B. Baker and M. L. Ginsberg, Temporal projection and explanation, in *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, Morgan Kaufmann Publishers, Inc., 1989, pp.906–911.

[2] A. B. Baker, Nonmonotonic reasoning in the framework of situation calculus, Artificial Intelligence, 49(1991) 5–23.

[3] M. Gelfond and V. Lifschitz, What are the limitations of the situation calculus?, in *Automated Reasoning*, Robert S. Boyer eds., Kluwer Academic Publishers Inc., 1991, pp167–176.

[4] M. P. Georgeff, Actions, processes, and causality, in *Proceedings of the 1986 Workshop on Reasoning About Actions and Plans*, M. P. Georgeff and A. L. Lansky eds., Morgan Kaufmann Publishers Inc., 1988, pp99–122.

[5] M. L. Ginsberg and D. E. Smith, Reasoning about action I: A possible worlds approach, in *Proceedings of the 1987 Workshop on the Frame Problem in Artifical Intelligence*, F. M. Brown eds., 1987, pp233-258.

[6] S. Hanks and D. McDermott, Default reasoning, nonmonotonic logics, and the frame problem, in *Proceedings of the Fifth National Conference on Artificial Intelligence*, 1986, pp328–333.

[7] S. Hanks and D. McDermott, Nonmonotonic logic and temporal projection, Artificial Intelligence, 33(1978) 379–412.

[8] B. A. Haugh, Simple causal minimizations for temporal persistence and projection, in *Proceedings of the Sixth National Conference on Artificial Intelligence*, 1987, pp218–223.

[9] H. Katsuno and A. O. Mendelzon, Propositional knowledge base revision and minimal change, Artificial Intelligence, 52(1991) 263–294.

[10] H. A. Kautz, The logic of persistence, in *Proceedings of the Fifth National Conference on Artificial Intelligence*, 1986, pp401–405.

[11] V. Lifschitz, Formal theories of action, in *Proceedings of the 1987 Workshop on the Frame Problem in Artificial Intelligence*, F. M. Brown eds., Morgan Kaufmann Publishers Inc., 1987, pp35–57.

[12] V. Lifschitz, Pointwise circumscription, in *Readings in Nonmonotonic Reasoning*, M. L. Ginsberg eds., Morgan Kaufmann Publishers Inc., 1987, pp179–193.
[13] V. Lifschitz, Frames in the space of situations, Artificial Intelligence, 46(1990) 365–376.
[14] F. Lin and Y. Shoham, Provably correct theories of action (preliminary report), in *Proceedings of the Ninth National Conference on Artificial Intelligence*, 1991, pp349–354.
[15] F. Lin and Y. Shoham, Concurrent actions in the situation calculus, in *Proceedings of the Tenth National Conference on Artificial Intelligence*, 1992, pp590–595.
[16] J. McCarthy and P. J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in *Machine Intelligence 4*, B. Meltxer and D. Mitchie eds., Edinburgh University Press, 1969, pp463–502.
[17] J. McCarthy, History of circumscription, Artificial Intelligence, 59(1993) 23–26.
[18] R. Reiter, Equality and domain closure in first-order databases, Journal of the Associate for Computing Machinery, 27(1980) 235–247.
[19] Y. Shoham, Chronological ignorance, in *Proceedings of the Fifth National Conference on Artificial Intelligence*, 1986, pp389–393.
[20] Y. Shoham, *Reasoning about Change*, MIT Press, Cambridge, Massachusetts, 1988.
[21] M. Winslett, Reasoning about action using a possible models approach, in *Proceedings of the Sixth National Conference on Artificial Intelligence*, 1988, pp89–93.
[22] Y. Zhang and N.Y. Foo, Reasoning about persistence: a theory of actions, in *Proceedings of Thirteenth International Joint Conference on Artificial Intelligence*, Morgan Kaufmann Publishers, Inc., 1993, pp718–723.
[23] Y. Zhang and N.Y. Foo, Updating knowledge bases using a persistent set approach, in *Proceedings of the 6th International Conference on Artificial Intelligence: Methodology, Systems, Applications*, World Scientific, 1994.
[24] Y. Zhang, *Reasoning about Persistence: A Unified Principle for State Change*, Ph.D. Thesis, University of Sydney, 1994.