Deriving Properties of Belief Update from Theories of Action (II)

Alvaro del Val and Yoav Shoham
Robotics Lab
Computer Science Department
Stanford University
Stanford, CA 94305
{delval,sliohain}@cs.stanford.edu

Abstract

In [del Val and Shoham, 1992] we showed that the postulates for belief update recently proposed by Katsuno and Mendelzon [I99I] can be analytically derived using the formal theory of action proposed by Lin and Shoham [I99I]. The contribution of this paper is twofold;

- Whereas in [del Val and Shoham, 1992] we only showed that our encoding of the update problem satisfied the KM postulates, here we use an independently motivated generalization of the theory of action used in that paper to provide a one-to-one correspondence between our construction and KM update semantics
- We show how the KM semantics can be generalized by relaxing our construction in a number of ways, each justified in certain intuitive circumstances and each corresponding to one specific postulate. It follows that there are reasonable update operators outside the KM family.

1 Introduction

Katsuno and Mendelzon [I99I] have recently proposed a characterization of belief update in terms of a set of postulates that every update operator should satisfy. Though these postulates are similar in spirit to those proposed by Gardenfors and colleagues [Alchourron et *al.*, 1985; Gardenfors, 1988] to characterize belief revision, one of the most novel aspects of the KM proposal is the suggestion that update and revision should be seen as two *distinct* types of belief change. Loosely speaking, the latter says that the beliefs may have been wrong and in need of revision, whereas the former says that the beliefs were correct, but the world has in the meanwhile evolved and the beliefs must be updated. According to this proposal, therefore, the problem of update is fundamentally one of *reasoning about change,* a problem which has received substantial attention over the years in research on non-monotonic temporal reasoning

The question we set out to answer in [del Val and Shoham, 1992] is the following: *Why* should the KM postulates be accepted? Our answer was based on the idea that since the basic intuition underlying update was one about change, it should be possible to reduce the KM proposal to a theory of action and change. Specifically, the logic behind our answer was as follows:

- We began with an independently motivated theory of action, namely the one proposed in [Lin and Shoham, 1991].
- We provided a natural encoding of the update problem in this theory of action.
- Finally, we showed that this encoding allowed us to analytically derive the KM postulates.

This argument can be subject to a number of critic isms, some of which have been articulated by Goldszmid! and Pearl [1992].

- The construction of [del Val and Shoham, 1992] *sat-* ----- the postulates, but so do others, *e.g.* [Goldszmidt and Pearl, 1992]. The KM proposal, however, seems to be more general.
- It can be argued that the representation theorem for KM operators ([Katsuno and Mendelzon, 1991], reviewed in section 2) provides sufficient validation for the KM proposal.

These are quite reasonable concerns, and our goal in this paper is to meet them. We retain the intuition that theories for reasoning about action and change should be at the basis of a solution to the problem of update, but try to answer a more general question than in [del Val and Shoham, 1992], namely: *Under what circumstances* should the KM postulates be accepted? Our answer has two parts:

- Using an independently motivated generalization of our construction (in a nutshell, allowing the set of facts that "persist by default" to depend on the current state of the world), we strengthen our previous results so as to identify a class of theories of action which stands in a one-to-one correspondence with the family of KM operators. We thus address the first potential concern.
- We show that this correspondence is only achieved by imposing certain arbitrary restrictions on the theory of action. Each of the restrictions we will consider corresponds to exactly one postulate, and thus the KM proposal can be seen as embodying

these same restrictive assumptions. We will show how relaxing these assumptions leads to reasonable update operators outside the KM family. We thus address the second concern.

In addition to answering these concerns, our construction will also allow us to shed some light on the predicate *frame,* which plays a key technical role in a number of papers in reasoning about action [Lifschitz, 1990; Lin and Shoham, 1991], but whose intuitive meaning was somewhat unclear.

One final question concerns our specific choice of a given formal framework to encode update. Admittedly, our formalization of update is comparatively complicated, but we believe the added complication is worth the price in that we can thereby make explicit the temporal evolution of the database, provide greater flexibility for update operators, and explicate the impact on update of some fundamental problems in reasoning about change. As regards to language, we use the situation calculus, which allows us to make explicit the temporal evolution of the database in a relatively simple way, and with the convenience of using FOL. More generally, the idea of encoding update using techniques from reasoning about action and non-monotonic temporal reasoning is justified because these techniques have proven fruitful in formulating, and then proposing solutions to, some key problems in reasoning about change. To illustrate the connection between non-monotonic temporal reasoning and update, suppose we have a database of facts about university life, and that we want to update the database with the fact that Smith is enrolled in CS205 this quarter. Clearly, this new fact should not affect many other facts in the KB, such as the composition of the faculty or the color of the university buildings. On the other hand. it. *should* have an effect on, say, her schedule and the total number of units she is taking, and these changes can in turn have other indirect effects, e.g. on whether she is a full-time student and the amount of tuition she has to pay, which in turn may depend on other factors such as whether she is an undergraduate, etc. These are clear instances of the frame and ramification problems. In particular, it would be unreasonable to require the database user to specify in advance all what, should or should not change as a result of the update

We believe, therefore, that any proposed update semantics must be defended in terms of its abilits to provide solutions to these problems, as lias also been argued by Reiter [l992a]. The theory of action of [bin and Shoham, 1991] has some well established properties in this regard in the case of actions with deterministic (direct and indirect.) effects, and our generalization of it appears to provide enough flexibility to deal with these problems in the context of indeterministic actions[1]. In principle, any other satisfactory solution to these problems could provide a foundation for update, but the generality of the results presented in this paper provides strong evidence for the reasonableness of our encoding

The structure of this paper is as follows Section '1 reviews Katsuno and Mendelzon's approach to update. Section 3 encodes the update problem in situation calculus. Sections 4 and 5 provide the one-to-one correspondence between KM semantics and a class of theories of action. Section 6 considers the consequences of relaxing some of the restrictive assumptions on the theory of action, thus generalizing the KM proposal. Related work is discussed in the concluding section.

[1]See section (> for examples of how indeteniiinisiie actions can complicate the frame problem

2 Update in propositional languages: Review

Katsuno and Mendelzon proposed eight postulates that should be satisfied by update operators. Let o be an update operator for a propositional language *C* with a finite number of propositional variables *Vc* The KM postulates are the following:

(U1) $\psi \diamond \mu$ implies μ.

(U2) If ψ implies μ then $\psi \diamond \mu$ is equivalent to ψ.

(U3) If ψ and μ are satisfiable then $\psi \diamond \mu$ is also satisfiable.

(U4) If $\models \psi_1 \equiv \psi_2$ and $\models \mu_1 \equiv \mu_2$ then $\psi_1 \diamond \mu_1$ is equivalent to $\psi_2 \diamond \mu_2$.

(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 $\psi \diamond \mu_1$ is equivalent to $\psi \diamond \mu_2$.

(U7) If ψ is complete then $(\psi \diamond \mu_1) \wedge (\psi \diamond \mu_2)$ implies $\psi \diamond (\mu_1 \vee \mu_2)$.

(U8) $(\psi_1 \vee \psi_2) \diamond \mu$ is equivalent to $(\psi_1 \diamond \mu) \vee (\psi_2 \diamond \mu)$

Update operators satisfying these postulates can be characterized as follows. An *update assignment* is a function which assigns to each interpretation I a relation $\leq_I$ over the set of interpretations of the language. We say that this assignment is *faithful* iff for any J, if $I \neq J$ then $I \leq_I J$ and $J \not\leq_I I$. Let $Mod(\phi)$ stand for the models of the formula ϕ, and $Min(S, \leq)$, for any set S and relation $\leq$ over S, denote the set of elements of S that are minimal under $\leq$.

Theorem 1 *[Katsuno and Mendelzon, 1991] An update operator $\diamond$ satisfies conditions (U1)-(U8) iff there exists a faithful assignment that maps each interpretation I to a partial order $\leq_I$ such that:*

$$Mod(\psi \diamond \mu) = \bigcup_{I \in Mod(\psi)} Min(Mod(\mu), \leq_I).$$

3 The update problem in situation calculus

The basic idea of our encoding of update is discussed in more detail in [del Val and Shoham, 1992]. Although update is supposed to reflect changes that have taken place in the world over time, the update problem (like t hat of belief revision) is almost always formulated using a language incorporating no model of time or change. Our approach consists in translating the update problem into the richer language of situation calculus, so as to make explicit this temporal information. The initial database is taken to describe a particular situation, and the update formula is taken to describe the occurrence of a special action, denoted by A^S_μ, whose intuitive reading is 'that action which, when taken in S, causes μ'. The updated database is taken to describe the situation $result(A^S_\mu, S)$. A formal theory of action is then used to infer facts about the result of taking the particular action in the particular situation. Finally, anything inferred about the resulting situation can be backtranslated to the timeless framework of belief update.

We use the standard situation calculus formalism, a three sorted predicate calculus with action, situation and fluent sorts, the binary predicate *holds*, with arguments of sort fluent and situation, respectively, and the binary function *result*, with arguments of sorts action and situation, respectively. The constant action terms have the form A^S_μ, with the intuitive meaning just described, where S is any situation term and μ is any *satisfiable* formula of the original propositional language $\mathcal{L}$ over the variables $\mathcal{P}_\mathcal{L}$. Situation terms include S_0, intuitively denoting the initial situation, together with all terms of the form $result(A, S)$ for A an action term and S a situation term. The set $\mathcal{P}$ of fluent terms includes all variables in $\mathcal{P}_\mathcal{L}$, together with the "non-primitive" fluent terms of the form $not(\psi)$ and $or(\psi)$, for ψ a fluent term. Non-primitive fluent terms are required to satisfy the following axioms:

$$\forall p, s.\ holds(not(p), s) \equiv \neg holds(p, s);$$
$$\forall p, q, s.\ holds(or(p, q), s) \equiv holds(p, s) \vee holds(q, s)$$

We can now translate the formulas of $\mathcal{L}$ into situation calculus. We first map any formula $\psi \in \mathcal{L}$ into a fluent term ψ^t as follows:

$$p^t = p \text{ if } p \in \mathcal{P}_\mathcal{L}$$
$$(\neg\psi)^t = not(\psi^t)$$
$$(\psi \vee \phi)^t = or(\psi^t, \phi^t).$$

(Other propositional connectives are assumed to be defined in terms of $\neg$ and $\vee$.) Propositional formulas are then translated as holding at a particular situation. We denote the translation of ψ at situation S by ψ^S, defined simply as $\psi^S = holds(\psi^t, S)$.

We now present the theory of action that we use to encode the update problem. The "causal theory" T for the actions we have introduced is given by the schema:

$$holds(\mu^t, result(A^S_\mu, S)),$$

where μ^t is the fluent term corresponding to a satisfiable propositional formula μ, and A^S_μ and S are closed terms of the appropriate sorts.

We will now define a circumscription policy to apply to this theory. Intuitively, our goal is to minimize changes ("abnormalities") between successive states of the world, one situation at a time. Following [Lin and Shoham, 1991], in [del Val and Shoham, 1992] we used $ab(p, a, s)$ as an abbreviation for:

$$frame(p) \wedge (holds(p, s) \equiv \neg holds(p, result(a, s))).$$

As it is clear from this definition, the role of the *frame* predicate (first introduced in [Lifschitz, 1990]) is to select a set of fluents for which changes must be minimized, or in other words, a set of fluents whose value 'persists by default'. In [Lin and Shoham, 1991; del Val and Shoham, 1992], the set of changes that should be minimized was assumed to be fixed once and for all, by means of some axiom F uniquely determining some set $\mathcal{F}$ of "frame fluents". This assumption is unnecessary, and is lifted in this paper. The set of facts that we think likely to persist in any given situation might well depend on the state of the world at the time. To allow for this dependence, we need to add a situation argument to the predicate *frame*. For greater clarity, we will in fact replace *frame* by a new binary predicate *persistent* with first argument of sort fluent and second argument of sort situation, redefining the abnormality predicate $ab(p, a, s)$ to be an abbreviation for

$$persistent(p, s) \wedge (holds(p, s) \equiv \neg holds(p, result(a, s))).$$

We assume that knowledge about which facts are to be treated as persistent in each state is encoded in our theory of action by means of some *persistence axiom* (P), which we leave as a parameter in the theory of action. This axiom, might, as a very simple example, declare every primitive fluent to be persistent at any situation. We want maximum generality, but some assumptions are needed in order to ensure that the persistence axiom fulfills its intended role. In particular, we require that the addition of the persistence axiom results in a conservative extension of the theory of action with respect to the language obtained by removing the predicate *persistent*; some further requirements are needed in order to ensure that the extension of *persistent* at any state depends only on the current and possibly past state of the database, not on future states or states along alternative "time branches". The formal statement of these requirements is given in the full paper and in [del Val, 1993].

Our circumscription policy is then as follows. Let $N1$ and $N2$ be unique names for fluents and situations, respectively, let *holds'* and *persistent'* be two new predicates with arguments of the same sort as *holds* and *persistent*, and let $W(s)$ be the set of formulas:

$$T \cup \{P, N1, N2, \forall p.\ holds(p, s) \equiv holds'(p, s),$$
$$\forall p.\ persistent(p, s) \equiv persistent'(p, s)\}.$$

Our final theory of action, which we denote $Comp(T)$, stands for the union over all closed terms A and S of the appropriate sorts of

$$Circum(W(S); ab(p, a/A, s/S); holds, persistent),$$

i.e. the circumscription of $ab(p, a/A, s/S)$ in $W(S)$ with *holds* and *persistent* allowed to vary. Intuitively, what this policy does is to minimize changes one situation at a time. For any situation S, the minimization will allow *holds* to vary at any other point except at S, since *holds'* is kept fixed.

Suppose now that we are given an initial propositional database ψ. Let ψ^{S_0} be the translation of ψ into situation calculus as holding at S_0. We can then take the result of updating ψ with μ as the set of consequences about the situation $result(A^{S_0}_\mu, S_0)$ entailed by $Comp(T) \cup \{\psi^{S_0}\}$. To capture this, let

$$U(\psi, \mu) \quad = \quad \{\varphi \mid Comp(T) \models \psi^{S_0} \supset \varphi, \text{ and } \varphi \text{ contains } result(A^{S_0}_\mu, S_0) \text{ as only situation term}\}$$

Define now the update operator $\circ$, for any database ψ and satisfiable formula μ, as follows:

Definition 1 $\psi \circ \mu \models \phi$ *iff* $holds(\phi^t, result(A_\mu^{S_0}, S_0)) \in U(\psi, \mu)$.

The result of sequences of updates can be similarly defined.

4 From KM semantics to theories of action

The operator $\circ$ just defined depends on the persistence axiom (P), which constrains the set of persistent fluents at each particular state of the world. Definition 1 does not therefore characterize any specific operator, but a family of them. Interesting subfamilies can be obtained by making further restrictions on the persistence axiom.

We will consider three such restrictions. The first one (SDP, for "state determines persistent fluents") requires that complete knowledge of the state of the world in any *given* situation be sufficient to uniquely determine the set of persistent facts *in* that situation. The second one (PDS, for "persistent fluents determine state") ensures that the values of the persistent fluents (at a situation S, in a model of $W(S)$) are sufficient to completely characterize a state. Finally, the "time independence" condition (TI) requires the set of persistent fluents to be identical for identical states at different times.

Before giving these restrictions, we need first to be more precise about the notion of "state". For any situation term S we say that R_S is a *state* of situation S iff there is some set of closed fluent terms F such that

$$R_S = \{holds(\theta, S) \mid \theta \in F\} \cup \{\neg holds(\theta, S) \mid \theta \in \mathcal{P} \setminus F\}^2,$$

and R_S is consistent with the axioms for non primitive fluents. Intuitively, a state of a situation is a complete specification of the values of all fluents at that situation. By analogy with the notation $Mod(\theta)$ for the models of θ, for any set of formulas $\Gamma(S)$ with *holds* as only predicate and S as only situation term, we use $States(\Gamma(S))$ to denote the set of states R of S such that $R \models \Gamma(S)$. Intuitively, the set of states of $\Gamma(S)$ corresponds to the models of the translation of $\Gamma(S)$ into $\mathcal{L}$.

The optional restrictions on the persistence axiom are then as follows:

Definition 2 (SDP condition) *A theory of action satisfies the SDP condition iff the persistence axiom is such that for any situation term S, any state R of S consistent with $W(S)$, and any fluent term θ, either $W(S), R \models persistent(\theta, S)$ or $W(S), R \models \neg persistent(\theta, S)$.*

Definition 3 (PDS condition) [3] *A theory of action satisfies the PDS condition iff for any situation term S, any two states R_1 and R_2 of S, and any set of fluent terms P such that $\{persistent(\theta, S) \mid \theta \in P\} \cup \{\neg persistent(\theta, S) \mid \theta \in \mathcal{P} \setminus P\}$ is consistent with $W(S) \cup R_1$: if R_1 and R_2 agree on the value of every $p \in P$ then $R_1 = R_2$.*

Definition 4 (TI condition) *A theory of action satisfies the TI condition if the persistence axiom has the form $\forall s.\ P(s)$, where $P(s)$ contains no situation term other than s.*

We can now prove the following key result:

Theorem 2 *For any update operator $\circ$ satisfying (U1)-(U8) there exists an operator $\circ'$ based on definition 1 and satisfying the PDS, SDP, and TI conditions such that:*

$$Mod(\psi \circ \mu) = Mod(\psi \circ' \mu).$$

The key part of the proof is the choice of persistence axiom; the construction is given in the appendix. We remark that this theorem holds for arbitrary sequences of updates as well, and thus we can capture the *temporal evolution* of the database under any KM operator.

5 From theories of action to KM semantics

Though we have shown that every KM operator can be captured in our generalized framework, a natural question is whether this generalization is "the right one". Does this construction still satisfy the KM postulates, as the more restricted one of [del Val and Shoham, 1992] did? The answer is, under certain conditions, positive.

The circumscription of ab for each situation and action results in a set of strict partial orderings $<_{ab, A_\mu^S, S'}$ over situation calculus interpretations, such that $I <_{ab, A_\mu^S, S'} J$ iff I and J have the same domains and agree on everything except *holds*, *persistent* and ab, and the extension of $ab(p, a/A_\mu^S, s/S')$ in I is a proper subset of its extension in J. As in [del Val and Shoham, 1992], models of $Comp(T)$ can also be characterized in terms of orderings over states rather than over interpretations. The intuition here is that the results of update should only depend on the immediately preceding state and the action corresponding to the update formula, modulo some set of persistent fluents associated to the preceding state. As the next lemma shows, the circumscription selects states of $result(A_\mu^S, S')$ which are "closest" to a given state of S', in the sense of differing in set-inclusion fewer persistent fluents than any other state. The ordering we use is defined as follows:

Definition 5 *Let R and T be states of some situation S', and M_S a state of some situation S, and let F be a set of fluent terms. We say that $R \leq_{M_S}^F T$ iff*

$$\{\theta \in F \mid R \models \neg holds(\theta, S') \text{ iff } M_S \models holds(\theta, S)\} \subseteq$$
$$\{\theta \in F \mid T \models \neg holds(\theta, S') \text{ iff } M_S \models holds(\theta, S)\}$$

We can use these orderings to characterize $Comp(T)$. For example, for a single update, we have:

Lemma 3 *Suppose the SDP condition holds, and for any state R of S_0, let F_R be the unique set of fluent terms $\{\theta \mid W(S_0) \cup R \models persistent(\theta, S_0)\}$. Then $States(U(\psi, \mu)) =$*

$$\bigcup_{R \in States(\psi^{S_0})} Min(States(holds(\mu^t, result(A_\mu^{S_0}, S_0))), \leq_R^{F_R})$$

---

[2] Recall that $\mathcal{P}$ is the set of closed fluent terms of the language

[3] This condition generalizes the "frame completeness condition" of [del Val and Shoham, 1992].

As in the representation theorem for propositional update, this can be seen as selecting for each state of the original theory (for each model, in the propositional case) the set of closest states (models) satisfying the update formula. Similar characterizations can be obtained without the SDP condition (in which case several partial orders will be associated to each state, corresponding to the various specifications of *persistent* at that state consistent with the persistence axiom) and for sequences of updates. Using these characterizations, we can obtain the converse of theorem 2:

Theorem 4 *Suppose the PDS, SDP, and TI conditions are satisfied. Then the update operator* o *of definition 1 satisfies postulates (Ul)-(US).*

6 Beyond KM semantics

So far we have achieved half the goal of this paper, with theorems 2 and 4 providing a very tight correspondence between theories of action and KM semantics As said, though, since our intuition is that theories of ad ion should form the basis for update, we believe that further insight can be achieved by relaxing some of the assumptions embodied by the theories of action we have considered. The KM postulates identify an interesting class of operators, but there are important operators outside this class. Specifically, it can be shown that postulates (UI), (U3), (U5), *(U6)* and *(US)* still hold if we lift the SDP, PDS and TI assumptions from theorem 4. but the remaining postulates are lost.

How reasonable is (U2)? Consider the following example, due to Goldszmidt and Pearl [1992]. Suppose we order a robot to paint a wall in blue or white. If the wall is initially white, then (U2) entails that it will remain white after the action. If the robot has no way of knowing the original color, however, there is no reason wh> the color of the wall should persist after the action For another example, suppose we know that Fred decided today whether to leave his current job to accept another offer, but we do not know his specific decision; according to (U2), the result of updating the database *current-job* with *current-job* V *new-job* should be that Fred rejects the offer! In both cases, the problem arises because *(V'2)* requires the database to remain unchanged when a disjunctive update arrives and the disjunction is already satisfied by the database.

Theorem 5- *In the presence of TI, the operator o satisfies (U2) iff the PDS condition holds* 4.

Both Goldszmidt and Pearl, and Katsuno and Mendelzon, have a possible solution to this type of problem (the latter through the operation of "erasure"). In our view. the key issue is whether certain facts should persist or not when disjunctive updates arrive; our construction allows for an explicit axiomatization of default persistence. and thus we suggest that it will provide the greatest flexibility in handling this problem.

4Throughout this section, o is assumed to he based on definition 1. Without TI, PDS is still sufficient for *(U2).* but is only necessary if restricted to states of the database that are "reachable" as a result of some sequence of updates

The TI condition can also result overrestrictive in forcing the set of persistent fluents at any state to be independent of past states. This assumption is connected to postulate (U4), according to which the update of equivalent databases with equivalent formulas should produce equivalent results. In the timeless framework in which the update problem is formulated in the KM proposal, the postulate appears to encode a principle of syntax independence. If we consider the evolution of the database, however, it is clear that it implies more than that. Suppose the database is represented by ψ 1at some time 1, is updated with *ji* to yield a database ψ'2) and after a series of updates the database ends up at time *t* in a state in which it can be represented by some formula *ψ1t* which is equivalent to ψ1 according to (U4), updating *ψt* with /i should result in a database equivalent to ψ2-

Clearly, this is not always reasonable. Myers and Smith [1988] present a number of examples in which the reasonableness of treating a fact as persistent by default depends on the way in which we came to know that fact. Similarly, it is easy to find examples in which some fact should be treated as persistent or not in virtue of what causes it. For example, the practice of mountain climbing results in a high risk of injury, but the risk will disappear as soon as the practice is given up; whereas the professional practice of tennis might result in a high risk of (elbow) injury long after the practice is quit. Thus, whether "high injury risk" needs to be treated as persistent depends on the circumstances that cause the risk.

Since the results of our circumscriptive policy at any given situation depend only on the set of persistent fluents at that situation, it is easy to drop this assumption as well, while preserving all postulates except (U4). This does *not* mean that updates depend on the syntactic form of the database, only that they depend on the set of persistent facts, which might differ for identical states at different times.

As the next theorem states, there is a very close connection between (U4) and the TI condition. In order to establish this connection, we need a way to compare theories of action with different persistence axioms, so as to filter out "inconsequential" violations of TI. Let us first explicitly include the persistence axiom in our notation for the update function, writing $\diamond_P$ to indicate that the underlying theory of action contains the persistence axiom (P). We say that two theories of action with persistence axioms (P) and (P*), respectively, are *update-equivalent* iff they are equivalent with respect to update, *i.e.* iff for any propositional ψ, satisfiable propositional $\mu_1, \ldots, \mu_n$, we have: $(\ldots((\psi \diamond_P \mu_1) \diamond_P \mu_2) \ldots) \diamond_P \mu_n \equiv (\ldots((\psi \diamond_{P^*} \mu_1) \diamond_{P^*} \mu_2) \ldots) \diamond_{P^*} \mu_n$.

Theorem G *The update operator o satisfies (U4) iff it is based on a theory of action which is update equivalent 1o a theory of action satisfying the TI condition.*

Finally, (U7) is connected to SDP. Intuitively, a (nontrivial) violation of SDP corresponds to situations in which we do not regard our knowledge of the domain to be sufficient to uniquely determine which facts are likely to persist, or in which we want to consider the effects of alternative assumptions about persistence. This is a perhaps rare but certainly conceivable circumstance. which is not covered by KM. We have:

Theorem 7

1. If the SDP condition holds then o satisfies (U7).
2. If o satisfies (U7), and the PDS condition holds, then o is based on a theory of action which is update equivalent to a theory of action satisfying SDP.

There appears to be some subtle interaction between (U2) and (U7), which is why the PDS condition appears to be needed in order to derive SDP from (U7) in the second part of the theorem.

7 Discussion

In this paper, we have shown that there is a very tight correspondence between a certain class of theories of action and the class of update operators satisfying Katsuno and Mendelzon's update postulates. In addition to providing a one-to-one correspondence between these two classes, we have shown that there are reasonable operators that do not satisfy all the postulates, and thus that the KM semantics can sometimes be too restrictive. The failure of each of the postulates than can be violated in our construction, furthermore, can be directly traced to an specific assumption on the specification of the set of facts that should "persist by default". We hope to have demonstrated that the greater complication of our construction in comparison with the KM approach is worth its price in the expanded expressivity and flexibility to be gained from it.

Reiter [1992b; 1992a] has proposed an account of database update in terms of recent proposals for solving the frame problem. The relation between his approach and ours is still unclear, as it is based on a slightly different theory of action. The theory that Reiter uses appears vulnerable to the ramification problem, though it is too early to say whether this is a fundamental limitation; in contrast, we can handle ramifications in exactly the same way as in [del Val and Shoham, 1992], without a need to explicitly list all the conceivable circumstances in which a fluent can change its value as a result of an action or update. Goldszinidt and Pearl [1992] have also provided an alternative, in this case probabilistically motivated, account of database update satisfying the KM postulates, under some special handling for disjunctive updates. Time is only implicit in their proposal, and. as any other KM operator, theirs is covered by our construction if disjunctive updates are handled as required for KM compliance.

Our results show that important insights can be obtained by encoding the problem of database update in a theory of action. We are currently investigating whether similar insights can also be obtained for database n-vision, an area in which we expect to report positive results in the near future.

References

[Alchourron et al., 1985] Carlos E. AlchournSn. Peter Gardenfors, and David Makinson. On the logic of theory change: Partial meet functions for contraction and revision. Journal of Symbolic Logic, 50:510 530. 1985.

[del Val and Shoham, 1992] Alvaro del Val and Yoav Shoham. Deriving properties of belief update from theories of action. In Proceedings of the Tenth Conference of the AAAI, 1992.

[del Val, 1993] Alvaro del Val. Belief Revision and Update. PhD thesis, Stanford University, 1993.

[Cardcnfors. 1988] Peter Gardenfors. Knowledge in Flux. The MIT Press, 1988.

[Oioldszmidt and Pearl, 1992] Moises Goldszmidt and Judea Pearl. Rank-based systems. In Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning, 1992.

[Katsuno and Mendelzon, 199l] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge database and revising it. In Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning, 1991.

[Lifschitz. 1990] Vladimir Lifschitz. Frames in the space of situations. .4rtificial Intelligence, 46:365-376, 1990.

[Lin and Shoham, 199l] Fanghzen Lin and Yoav Shoham. Provably correct theories of action (preliminary report). In Proceedings of the Ninth Conference of the AAAI, 1991.

[Myers and Smith, 1988] Karen L. Myers and David E. Smith. The persistence of derived information. In Proreedings of the Seventh Conference of the AAAI, 1988.

[Reiter. 1992a] Raymond Reiter. On specifying database updates Technical Report KRR-TR-92-5, Computer Science Department. University of Toronto, 1992.

[Reiter, 1992b] Raymond R. Reiter. On formalizing database updates: Preliminary report. In Proceedings of the Third International Conference on Extending Database Technology, 1992.

Appendix

Proof of theorem 2. (Construction.) Since $\circ$ satisfies (U1)-(U8), there exists a faithful update assignment of a partial order $\leq_M$ to every $M \in \mathcal{W}$ such that $\circ$ can be defined in terms of the representation theorem 1, where $\mathcal{W}$ is the set of all (propositional) interpretations of $\mathcal{L}$. The crucial part on the proof is providing an adequate persistence axiom. For any $W \subseteq \mathcal{W}$, choose one formula θ_W such that $Mod(\theta_W) = W$. For any $M \in \mathcal{W}$, let $\Sigma_M = \{\theta_W \mid W = \{I \mid I \leq_M J\}$ for some $J \in \mathcal{W}\}$, and let $M(s)$ be a finite axiomatization of the state of situation s such that $M \models \theta$ iff $M(s) \models holds(\theta^t, s)$. Let $\mathcal{M}_s$ be the (finite) set of all such finitely axiomatized states, and for each $M(s) \in \mathcal{M}_s$, let $P_M(s)$ be a finite first order axiomatization of '$persistent(\theta^t, s)$ iff $\theta \in \Sigma_M$'. (The finite vocabulary of $\mathcal{L}$ guarantees finitariness in all these cases.) Then $\circ'$ is defined as in definition 1 in terms of a theory of action whose persistence axiom is:

$$\forall s. \bigwedge_{M(s) \in \mathcal{M}_s} M(s) \supset P_M(s).$$

It is now easy to show that the TI, SDP, and PDS conditions are satisfied by this persistence axiom, the latter because of faithfulness. The rest of the proof uses the results of section 5 to map operators definable in our construction to "update assignments", and will be provided in the full paper (see also [del Val, 1993]). □