An Event-Based Abductive Model of Update

Craig Boutilier
Department of Computer Science
University of British Columbia
Vancouver, British Columbia
CANADA, V6T 1Z4
email: cebly@cs.ubc.ca

Abstract

The Katsuno and Mendelzon theory of belief update has been proposed as a reasonable model for revising beliefs about a changing world. However, the semantics of update relies on information which is not readily available. We describe an alternative semantical view of update in which observations are incorporated into a belief set by: a) explaining the observation in terms of a set of plausible events that might have caused that observation; and b) predicting further consequences of those explanations. We also allow the possibility of *conditional explanations*. We show that this picture naturally induces an update operator under certain assumptions. However, we argue that these assumptions are not always reasonable, and they restrict our ability to integrate update with other forms of revision when reasoning about action.

1 Introduction

Reasoning about action and change has been a central focus of research in AI for many years, dating back at least to the origins of the situation calculus (McCarthy and Hayes 1969). For example, a planning agent must be able to predict the effects of its actions on the world in order to verify whether a potential plan achieves a desired goal. Actions can be viewed as effecting changes in the world, and agents must be able to change their beliefs about the world to reflect such considerations.

One of the most influential theories of belief change has been the *AGM theory* proposed by Alchourrón, Gärdenfors and Makinson (1985). Imagine an agent possesses a belief set or knowledge base *KB*. The AGM theory provides a set of postulates constraining the possible ways in which the agent can change *KB* in order to accommodate a new belief *A*. Notice that this *revision* of *KB* need not be straightforward, for the new belief *A* may conflict with beliefs in *KB*. It was pointed out by Winslett (1988) that the AGM theory is inappropriate for reasoning about changes in belief due to the evolution of a changing world. A new form of belief change dubbed *update* was proposed in full generality by Katsuno and Mendelzon (1991), who provided a set of postulates, distinct from the AGM postulates, that characterize this type of belief change.

Semantically, Katsuno and Mendelzon have shown that belief update can be viewed by positing a family of orderings over possible worlds, with each ordering being indexed by some world. The ordering associated with a specific world can be viewed intuitively as describing the most plausible ways in which that world can change. To update a knowledge base *KB* with some proposition *A*, the worlds admitted by *KB* are each updated by finding the most plausible change associated with that world satisfying *A* (we describe this formally below).

In this paper, we present an abductive view of update that breaks the Katsuno-Mendelzon semantics into smaller, more primitive parts. We argue that such a model provides a more natural perspective on belief update in response to changes in the world, and exploits information that is more readily available. In general, we take update to be a two stage process of explanation followed by prediction: first, an agent *explains* an observation by postulating some *plausible event* or events that could have caused that observation to hold, relative to its initial state of knowledge; second, an agent *predicts* the (further) consequences of these events, relative to this initial state. We formalize this notion in an abstract manner obtaining a class of *explanation-change* operators. We show that explanation-change satisfies some of the properties of update operators determined by the Katsuno-Mendelzon (KM) theory. Furthermore, if we make two additional assumptions our model determines a KM update operator. However, we will argue that these additional assumptions are not always appropriate. In particular, should we intend to use update to reason about action, and have the results of actions provide information about the state of the world, the general form of update has to be modified. This modification is pursued in (Boutilier 1993; Boutilier 1994b).

We also briefly describe and characterize a special class of update operators. Finally, we compare our construction to the model of update proposed by del Val and Shoham (1992). Proofs of the results can be found in (Boutilier 1994b).

2 The Semantics of Update

Katsuno and Mendelzon (1991) have proposed a general characterization of belief update. Update is distinguished from belief *revision* conceptually by viewing update as reflecting belief change in response to changes in the world, whereas revision is thought to be more appropriate for changing (possibly erroneous) beliefs about a static world. Update is described by Katsuno and Mendelzon with a set of postulates constraining acceptable update operators and a possible worlds semantics, which we review here.

We assume the existence of some knowledge base KB, perhaps the set of beliefs held by an agent about the current state of the world. We take our underlying logic to be propositional, based on a finitely generated language $\mathbf{L}_{CPL}$. We use W to denote the set of *possible worlds* (or models) suitable for this language.

If some new fact A is observed in response to some (unspecified) change in the world (i.e., some action or event occurrence), then the formula $KB \diamond A$ denotes the new belief set incorporating this change. The *KM postulates* (Katsuno and Mendelzon 1991) governing admissible update operators are

(U1) $KB \diamond A \models A$

(U2) If $KB \models A$ then $KB \diamond A$ is equivalent to KB

(U3) If KB and A are satisfiable, then $KB \diamond A$ is satisfiable

(U4) If $\models A \equiv B$ then $KB \diamond A \equiv KB \diamond B$

(U5) $(KB \diamond A) \wedge B \models KB \diamond (A \wedge B)$

(U6) If $KB \diamond A \models B$ and $KB \diamond B \models A$ then $KB \diamond A \equiv KB \diamond B$

(U7) If KB is complete then $(KB \diamond A) \wedge (KB \diamond B) \models KB \diamond (A \vee B)$

(U8) $(KB_1 \vee KB_2) \diamond A \equiv (KB_1 \diamond A) \vee (KB_2 \diamond A)$

A better understanding of the mechanism underlying update can be achieved by considering the possible worlds semantics described by Katsuno and Mendelzon, which they show to be equivalent to the postulates. For any proposition A, let $\|A\|$ denote the set of worlds satisfying A. Clearly, $\|KB\|$ represents the set of possibilities we are prepared to accept as the actual state of affairs. Since observation O is the result of some change in the actual world, we ought to consider, for each possibility $w \in \|KB\|$, the most plausible way (or ways) in which w might have changed in order to make O true. To capture this intuition, Katsuno and Mendelzon postulate a family of preorders

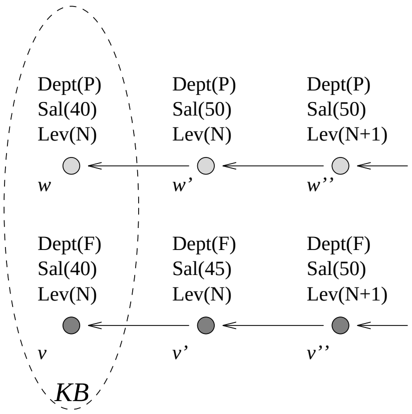

Figure 1: An Update Model

$$\{\leq_w : w \in W\}$$

where each $\leq_w$ is a reflexive, transitive relation over W. We interpret each such relation as follows: if $u \leq_w v$ then u is at least as plausible a change relative to w as is v. Finally, a *faithfulness condition* is imposed: for every world w, the preorder $\leq_w$ has w as a minimum element; that is, $w <_w v$ for all $v \neq w$.

Naturally, the most plausible candidate changes in w that result in O are those worlds v satisfying O that are minimal in the relation $\leq_w$. The set of such minimal O-worlds for each relation $\leq_w$, and each $w \in \|KB\|$, intuitively capture the situations we ought to accept as possible when updating KB with O. In other words,

$$\|KB \diamond O\| = \bigcup_{w \in \|KB\|} \{\min_{\leq_w}\{v : v \models O\}\}$$

where $\min_{\leq_w} X$ is the set of minimal elements (w.r.t. $\leq_w$) within X. Katsuno and Mendelzon show that such a formulation of update captures exactly the same class of change operators as the postulates; thus, we can treat this as an appropriate semantics for update.

As an example, consider the following scenario illustrating the application of the KM update semantics to database update. We know certain facts about an employee Fred: his salary is \$40,000, his job classification is level N, and so on. But, we are unsure whether he works for the Purchasing department or the Finance department. Thus, our KB admits two possibilities, w and v, reflecting this uncertainty (see Figure 1). If the orderings $\leq_w$ and $\leq_v$ are as indicated in the figure, then KB updated with the fact that Fred's salary is \$50,000 contains, among other things, the facts `Dept(P)` $\vee$ `Dept(F)`, `Dept(P)` $\supset$ `Level(N)` and `Dept(F)` $\supset$ `Level(N+1)`. This is due to the fact that the closest world to w with the new salary is w', while the closest to v is v''; hence, *KB* is determined by the set of worlds $\{w', v''\}$. This may reflect the fact that such a raise comes only with a promotion in Finance, whereas promotions are rare and raises more frequent in Purchasing.

3 Update as Explanation

3.1 Plausible Causes of Observations

The orderings upon which update semantics are based are interpreted as describing the most plausible manner in which that world might change. Given the role of update, this interpretation seems correct: worlds closer to w in the ordering $\leq_w$ are somehow more plausible states into which w might evolve. It seems reasonable then to update a *KB* by considering those most plausible changes. In our example above, if Fred is in Purchasing (world w), then a change of salary of this type is more likely to come without a change in rank (w') than with a change in rank (w'').

While reasonable, it begs the question: why would one change be judged more plausible than another? Intuitively, it seems that there are certain *events* or *actions* that would *cause* a change in w, and that those leading to w' are more plausible than those leading to w''. For example, the event `RAISE` might be more probable than the event `PROMOTION` (at least, in Purchasing).

Given an observation `Sal(50000)` — in this case an update transaction — an agent might come to believe `Dept(P)` $\supset$ `Level(N)` (as we have in our example) as follows. Assuming `Dept(P)`, the most plausible event that might *cause* such a change in salary is `RAISE` (rather than `PROMOTION`). Thus `RAISE` is the best *explanation* for the observation. Adopting this explanation has, as a further consequence, that job rank (and department) stays the same; thus, belief in `Level(N)` remains. In contrast, `RAISE` (to $50,000) is less likely than `PROMOTION` in the Finance department.[1] Thus, `PROMOTION` is the most plausible explanation for the observation, which has the additional consequence `Level(N+1)`. Thus, the two beliefs `Dept(P)` $\supset$ `Level(N)` and `Dept(F)` $\supset$ `Level(N+1)` hold in the updated belief state.

This leads to a very different view of update. When confronted with an observation or update O, an agent seeks an *explanation* of O, in terms of some external event that would have caused O had it occurred.[2] While many events might explain O in this way, some will be more plausible than others, and it will be those the agent adopts. Given such an explanation, one may then proceed to *predict* further consequences of these events, and produce the set of beliefs arising from the observation. With this point of view, the essence of update is captured by a two-step process: a) *explanation* of the observation in terms of some event(s); and b) *prediction* of the (additional) consequences of that event.

Before formalizing this idea, it is important to realize that this perspective is very natural. It is reasonable to suppose that an agent (or builder of a *KB*) has ready access to some description of the preconditions and effects of the possible events in a given domain. This assumption underlies all work in classical planning and reasoning about action, ranging from STRIPS (Fikes and Nilsson 1971) to the situation calculus (McCarthy and Hayes 1969; Reiter 1991) to more sophisticated probabilistic representations (Kushmerick, Hanks and Weld 1993; Dean et al. 1993). With such information, the predictions associated with explanations (event occurrences) can be easily determined. Furthermore, an ordering over the relative likelihood of possible events also seems something which an agent or system designer or user might easily postulate. This should certainly be easier to construct than a direct ordering over worlds according to their likelihood of "occurring." Indeed, we will show that such an ordering over worlds is *derivable* from this more readily available information.

This provides a possible interpretation of the update process, and in our view, a very natural one.[3] Furthermore, as we describe in the concluding section (and in detail in (Boutilier 1993)), by breaking update into two components, we will be able to extend the type of reasoning about action one can perform in this setting.

Using explanation for reasoning about action has been proposed by a number of people, especially within the framework of the situation calculus. Work on temporal projection and prediction failures often exploits the notion of explanation. For instance, Morgenstern and Stein (1988) propose a model where an observation that conflicts with the predicted effects of an agent's actions causes the agent to infer the existence of some external event occurrence. Shanahan (1993) proposes a model with a similar motivation, but adopts a truly abductive model (where candidate events are hypothesized rather than deduced from an observation). Our model will be rather different in several ways. First, explanations will be *conditional* (i.e., explaining events are conditioned on certain propositions). Second, the criteria used for adopting explaining events will be based on the relative plausibility of events. Third, we will not limit attention to any particular model of action (such as the situation calculus). Finally, our goal is to show how explanation

[1] In our example, we assume that a raise to $45,000 is most likely (world v'), but that a higher raise is unlikely without a promotion.

[2] In this paper we will usually think of (external) *events* as the impetus for change, rather than *actions* over which the agent has direct control (or of which the agent has direct knowledge).

[3] This should not be taken as a criticism of update for requiring that a reasoning agent have an explicitly specified family of preorders at its disposal. One can reason about update with syntactic constraints or by any other means. The point is that, from a semantic point of view, the preorders and syntactic constraints seem to be *induced* by considerations about action effects and plausible event occurrences.

can account for the *update* of a knowledge base. We should point out that Reiter (1992, and personal communication) has informally suggested that update can be viewed as explanation to events causing an observation. We will now proceed to show that this is, in fact, the case.

3.2 A Formalization

To capture update in terms of explanation, we require two ingredients missing from the Katsuno-Mendelzon account: a set of *events* that cause changes, and an *event ordering* that reflects the relative plausibility of different event occurrences.

We assume a finitely generated propositional language with an associated set of worlds W. Let E be a finite *event set*, the elements of which are primitive events. In general, $e \in E$ is a mapping $e : W \to 2^W$. For $w \in W$ and $e \in E$, we use $e(w)$ to denote the *result* of event e occurring in world w. This is a set of worlds, each of which is a possible *outcome* of e occurring at w. An event with more than one possible outcome is *nondeterministic*. A *deterministic* event is any $e \in E$ such that $e(w)$ is a singleton set for each $w \in W$. A *deterministic event set* is an event set all of whose events are deterministic. We assume that events are total functions on the domain W, so that every event can be applied to each world.[4]

Typically, events are not specified as mappings of this type. Rather, for each event (or action), a list of conditions are provided that influence the outcome of the event. For each such condition, a set of effects is specified. An example of this is the classical situation calculus representation of actions (in the deterministic case). Another is the modified STRIPS representation presented in (Kushmerick, Hanks and Weld 1993). The key feature of these, and other representations, is that each action/event induces a function between worlds (or worlds and sets of worlds).[5] Thus, most action representations will fit within this abstract model.

As a further generalization, if events are nondeterministic, we might suppose that the possible outcomes are ranked by probability or plausibility. We set aside this complication (but see (Boutilier 1993)).

In order to explain certain observations by appeal to plausible event occurrences, we need some metric for ranking such explanations. We assume that the events in the set E are ranked by plausibility; hence, we postulate an indexed family of *event orderings*

$$\{\preceq_w : w \in W\}$$

over E. We take $e \preceq_w f$ to mean that event e is at least as plausible (or likely to occur) as event f in world w. We require that $\preceq_w$ be a preorder for each w, and will occasionally assume that $\preceq_w$ is a total preorder.

Putting these ingredients together, we have the following definitions:

Definition An *event model* is a triple $\langle W, E, \preceq \rangle$, where W is a set of worlds, E is a set of events (mappings $e : W \to 2^W$) and $\preceq$ is an indexed family of events orderings $\{\preceq_w : w \in W\}$ (where each $\preceq_w$ is a preorder over E).

Definition A *deterministic event model* is an event model where every $e \in E$ is deterministic (i.e., for all $w \in W$, $e(w) = \{v\}$ for some $v \in W$). A *total order event model* is an event model where each event ordering $\preceq_w$ is a total preorder over E.

Given an event model, an agent is able to incorporate a new piece of information through a process of explanation and prediction as discussed above. An explanation of an observation is some event e that, when applied to the world under investigation, possibly causes O. However, the agent should be interested only in the most plausible such events.

Definition Let O be some proposition and $w \in W$. The set of *weak explanations* of O relative to w is

$$\mathit{Expl}(O, w) = \min_{\preceq_w}\{e \in E : e(w) \cap \|O\| \neq \emptyset\}$$

An event e is a *weak explanation* of O relative to w iff $e \in \mathit{Expl}(O, w)$. If $\mathit{Expl}(O, w) = \emptyset$, we say that O is *unexplainable* relative to w.

In other words, e explains O in a world w just when there is some possible result of e that satisfies O, and no more plausible event e' has this feature. Such explanations are called weak explanations because, before the observation O is made, an agent would not, in general, be able to *predict* that O would result from e. The agent merely knows that O is true in *some* possible outcome. A strong explanation is similar, but is predictive: *each* outcome of e satisfies O.

Definition The set of *strong explanations* of O relative to w is

$$\min_{\preceq_w}\{e \in E : e(w) \subseteq \|O\|\}$$

The distinction between weak and strong explanations is very similar to that made between *consistency-based* diagnosis (Reiter 1987) and *predictive* (or *abductive*) diagnosis (Poole 1988). This distinction is illustrated in Figure 2. Both e and f are nondeterministic events.

[4] It is best to think of events as analogous to "action attempts." If the preconditions for the "successful" occurrence of the event are not true at a given world, then the effects can be null, or unpredictable or something like that. Allowing preconditions is a trivial and uninteresting addition for our purposes here.

[5] In the case of the situation calculus, dynamic logic or other temporal formalisms, one would require some solution to the frame problem. For example, the solution of Reiter (1991) induces just such a mapping.

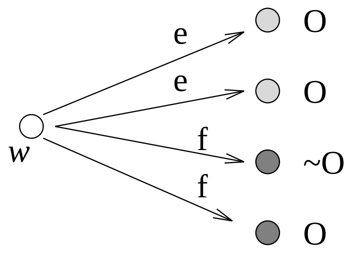

Figure 2: Weak and Strong Explanations

Event e strongly explains O, while f weakly explains O but does not strongly explain O. We are interested here in weak explanations, for these seem most appropriate when dealing with nondeterministic events. However, we note the following:

Proposition 1 *If e is a deterministic event, then e weakly explains O iff e strongly explains O.*

For a particular world w, $\mathit{Expl}(O, w)$ denotes those most plausible events that would cause O to be true. The possibilities admitted by such a set of explanations are the possible results of each of these events; that is:

Definition The *result* of O relative to w is the set of worlds

$$\mathit{Res}(O, w) = \bigcup\{e(w) \cap \|O\| : e \in \mathit{Expl}(O, w)\}$$

Note that if O is unexplainable relative to w, then $\mathit{Res}(O, w) = \emptyset$. Thus, that w might have evolved into a world satisfying O is impossible.

Taking a cue from the Katsuno-Mendelzon update semantics, the result of an observation with respect to a knowledge base KB is obtained by considering all plausible evolutions of each world $w \in \|\mathit{KB}\|$. However, if O is unexplainable for some $w \in \|\mathit{KB}\|$, we take O to be unexplainable relative to KB as a whole.

Definition The *result* of O relative to KB is the set of worlds

$$\mathit{Res}(O, \mathit{KB}) = \bigcup\{\mathit{Res}(O, w) : w \in \|\mathit{KB}\|\}$$

If $\mathit{Res}(O, w) = \emptyset$ for some $w \in \|\mathit{KB}\|$, we let $\mathit{Res}(O, \mathit{KB}) = \emptyset$.

The motivation for this last condition, that O must be explainable relative to every $w \in \|\mathit{KB}\|$, comes from update semantics itself. In update, the updated KB is constructed by considering the possible evolution of *every* possibility admitted by KB. We might have allowed the result of O to be nontrivial even if some worlds could not evolve so as to satisfy O, and define $\mathit{Res}(O, \mathit{KB})$ without this last condition. However, we adopt the current approach for two reasons: first, our goal is to pursue the analogy with update semantics; and second, when we drop this restriction, we intend to make this definition even weaker than we can by simply dropping the last condition. In (Boutilier 1993), we consider how to exclude both impossible and *implausible* evolutions. We elaborate on this in the concluding section.

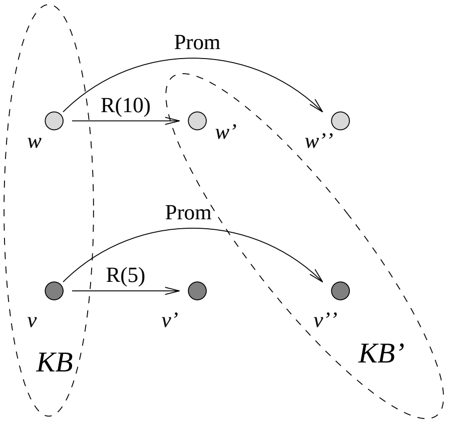

Figure 3: An Event Ordering

With such a result function, we can now define the *explanation-change operator* relative to a given event model, which determines the consequences of adopting an observation.

Definition The *explanation-change operator* induced by an event model EM is $\diamond_{EM}$:

$$\mathit{KB} \diamond_{EM} O = \{A \in \mathbf{L}_{CPL} : \mathit{Res}(O, \mathit{KB}) \models A\}$$

In our example, we have two event types, Promotion and Raise. A `PROM` event (promotion of one level) ensures an employee's rank is increased and his salary is raised \$10,000. Events `RS(5)` and `RS(10)` raise salary \$5000 and \$10,000, respectively. We assume the following event orderings for each department:

Purchasing: `RS(10)` $\prec$ `PROM` $\prec$ `RS(5)`
Finance: `RS(5)` $\prec$ `PROM` $\prec$ `RS(10)`

This is illustrated in Figure 3, where shorter event arcs depict more plausible occurrences. The explanation relative to purchasing is a raise, while for finance it is a promotion. The updated KB' is determined by w' and v'' and induces the beliefs described earlier.

As another example, imagine that a warehouse control agent expects a series of trucks to pickup and deliver certain shipments, but at time $t1$ an expected truck A has not arrived. Assume that this might be explained by snow on Route 1 or a breakdown. If snow is the most plausible of the two events, the agent might reach further conclusions by predicting the consequences of that event; for example, trucks B and D will also be delayed since they use the same route. The proper explanation and subsequent predictions are crucial, for they will impact on the agent's decision regarding staffing, scheduling and so on. Notice also that such explanations are defeasible, which is reflected in the defeasibility of update: if A is late but B is on time, then snow is no longer plausible (therefore, e.g., D will not be delayed).

We should remark at this point that the intent of this model is to provide an abductive semantic model for update, not a computational model. Just as we do not expect actions or events to be represented as abstract functions between worlds, explanations will not typically be generated on a world by world basis. Usually, the same event will explain an observation for a large subset of the worlds with $\|KB\|$. In particular, we expect that $\|KB\|$ to be partitioned according to some small number of propositions (or conditions) for which a certain event is deemed to be a reasonable explanation. Indeed, these can naturally be viewed as *conditional explanations*, for example, "If Fred is in Finance, a `PROMOTION` must have occurred; but otherwise a `RAISE` must have occurred." How such conditional explanations should be generated will be intimately tied to the action or event representation chosen, and is beyond the scope of this paper.

3.3 Relationship to Update

We are interested in the question of whether the explanation-change operator satisfies the update postulates. As presented above, this is not the case.

Proposition 2 *Let $\diamond_{EM}$ be the explanation-change operator induced by some event model. Then $\diamond_{EM}$ satisfies postulates (U1), (U4), (U6) and (U7).*

There are two reasons why the remainder of the postulates are not satisfied in general, hence two assumptions that can be made to ensure that $\diamond_{EM}$ is an update operator.

The first difference in the explanation-change operator is reflected in the failure of (U2), which asserts that $KB \diamond A$ is equivalent to KB whenever KB entails A. A simple example illustrates why this cannot be the case is general. Consider a KB satisfied by a single world w where $w \models A$. Postulate (U2) requires that the observation of A induce no change in KB. However, it may be that the most plausible event in the ordering $\preceq_w$ is e, where $e(w) = \{v\}$ for some distinct world v. But assuming $v \models A$, then $KB \diamond_{EM} A$ is captured by v and is thus distinct from w. In order to conform to postulate (U2), we must make the assumption that no change in w is more plausible than change induced by some event. Formally, we postulate *null events* and make these most plausible.

Definition The *null event* is an event n, where $n(w) = \{w\}$ for all $w \in W$.

Definition Let $EM = \langle W, E, \preceq \rangle$ be an event model. EM is *centered* iff the null event $n \in E$ and, for each $w \in W$ and $e \in E$ $(e \neq n)$ we have $n \prec_w e$.

Thus, a centered event model is one in which the null event is the most plausible event that could occur at any world. This seems to be the crucial assumption underlying postulate (U2).

Proposition 3 *Let $\diamond_{EM}$ be the explanation-change operator induced by some centered event model. Then $\diamond_{EM}$ satisfies postulates (U1), (U2), (U4), (U6) and (U7).*

This assumption of persistence of the truth of KB seems to be reasonable in many domains, but should probably be called into question as a general principle. It may be the case in a domain where change is the norm that, despite the fact that an observation is already believed, some change in KB should be forthcoming. In this sense, the more general nature of the explanation-change operator may be desirable.

Postulate (U3) is also violated by our model, and for a similar reason, so too are (U5) and (U8). For a given KB, we may have that $Res(O, w) = \emptyset$ for each $w \in \|KB\|$. In other words, there are no possible events that would cause an observation O to become true. The potential for such unexplainable observations clearly contradicts (U3), which asserts that $KB \diamond O$ must be consistent for any consistent O. The assumption underlying (U3) in update semantics seems to be the following: every consistent proposition is explainable, no matter how unlikely. In order to capture this assumption, we propose a class of event models called *complete*.

Definition Let $EM = \langle W, E, \preceq \rangle$ be an event model. EM is *complete* iff for each consistent proposition O and $w \in W$, O is explainable relative to w.

Proposition 4 *If EM is a complete event model then $Res(O, KB) \neq \emptyset$ for any consistent O and KB.*

Of course, this condition is sufficient to ensure (U5) and (U8) are satisfied as well.

Proposition 5 *Let $\diamond_{EM}$ be the explanation-change operator induced by some complete event model. Then $\diamond_{EM}$ satisfies postulates (U1), (U3), (U4), (U5), (U6), (U7) and (U8).*

The completeness of an event model refers, in fact, to the completeness of its event set E. If this set is rich enough to ensure that, for every world and observation, some event can make that observation hold, then the event model will be complete. Typically, domains will not be so well-behaved. However, the simple addition of a *miracle* event to an event set will ensure completeness. Intuitively, a miracle is some event which is less plausible than all others and whose consequences are entirely unknown.

Definition Let $EM = \langle W, E, \preceq \rangle$ be an event model. A *miracle* is an event m such that $m(w) = W$ for all $w \in W$, and $e \prec_w m$ for all $w \in W$ and $e \in E$ $(e \neq m)$.

Proposition 6 *Let $EM = \langle W, E, \preceq \rangle$ be an event model. If E contains a miracle event, then EM is complete.*

If all observations must be explainable, and no observation is permitted to force an agent into inconsistency, then miracles are one embodiment of the required assumptions. The reasonableness of such a requirement can be called into question, however. Having unexplainable observations is, in general, a natural state of affairs. Rather than relying on miraculous explanations, the threat of an inconsistency can force an agent to reconsider the observation, its theory of the world, or both. As we will see in the concluding section, it is just this type of inconsistency that can force an agent to revise its beliefs about the world prior to the observation. Update postulate (U3) makes it difficult to combine update with revision in this way.

If we put together Propositions 3 and 5, we obtain the main representation result for explanation-change.

Theorem 7 *Let $\diamond_{EM}$ be the explanation-change operator induced by some complete, centered event model. Then $\diamond_{EM}$ satisfies update postulates (U1) through (U8).*

A useful perspective on the relationship between explanation change and update comes to light when one considers that the plausibility ordering on events quite naturally induces an indexed family of preorders of the type required in the Katsuno-Mendelzon update semantics.

Definition Let $EM = \langle W, E, \preceq \rangle$ be an event model. The plausibility ordering *induced by EM*, for each $w \in W$, is defined as follows: $v \leq_w u$ iff for any event e_u such that $u \in e_u(w)$, there is some event e_v (where $v \in e_v(w)$) such that $e_v \preceq_w e_u$.

Theorem 8 *Let $\{\leq_w \colon w \in W\}$ be the family plausibility orderings induced by some complete, centered event model EM. Then*

(a) Each relation $\leq_w$ is a faithful preorder over W.

(b) The change operation determined by $\{\leq_w \colon w \in W\}$ is an update operator.

(c) The update operator determined by $\{\leq_w \colon w \in W\}$ is equivalent to the explanation-change operator $\diamond_{EM}$.

If we have an event model where each event ordering is a total preorder, then the induced plausibility orderings over worlds are also preorders.

Proposition 9 *Let $EM = \langle W, E, \preceq \rangle$ be an event model such that $\preceq_w$ is a total preorder for each $w \in W$. Then each plausibility ordering $\leq_w$ induced by EM is a total preorder.*

Since such a circumstance may arise rather frequently, the properties of such *total update operators* are of interest. We can extend the Katsuno-Mendelzon representation theorem to deal with update operators of this type. The required postulate embodies a variant of the principle of rational monotonicity, cited widely in connection with nonmonotonic systems of inference and conditional logics (see, e.g., (Boutilier 1994a)).

(U9) If KB is complete, $(KB \diamond A) \not\models \neg B$ and $(KB \diamond A) \models C$ then $(KB \diamond (A \wedge B)) \models C$ then

Theorem 10 *An update operator $\diamond$ satisfies postulates (U1) through (U9) iff there exists an appropriate family of faithful* total *preorders $\{\leq_w \colon w \in W\}$ that induces $\diamond$ (in the usual way).*

As a final remark, we note that the converses of Theorems 7 and 8 are trivially and uninterestingly true. For any update operator $\diamond$, one can construct an appropriate set of events (and orderings) that will induce that operator. This not of interest, since the point of explanation-change is to provide a natural view of update, characterizable in terms of the events of an existing domain. The ability to construct such events to capture a particular update operator provides little insight into update. The appropriate perspective is to reject any update operator (in a given domain) that cannot be induced by the existing set of events (or event model).

4 Concluding Remarks

We have provided an abductive model for incorporating into an existing belief set observations that arise through the evolution of the world. While our model allows more general forms of change than KM-update, we can impose restrictions on our model to recover precisely the KM theory. However, these restrictions are inappropriate in many cases, calling into question the suitability of some of the update postulates.

Of particular concern, as emphasized earlier, is postulate (U3). This embodies the assumption that all observations are explainable in terms of some event. This is not always reasonable. For instance, in our database example we might have a transaction to update Fred's salary to \$90,000 when there is a salary cap of \$80,000 in Finance. Thus, no event could have caused such an occurrence if Fred is indeed in Finance. Far from being a miraculous occurrence, it suggests that Fred in actually in Purchasing. Thus the observation not only forces KB to be updated (reflecting a change in the world), but also revised (reflecting additional knowledge about the world.

Note that this is not an artifact of out definition of update, where we insist that an observation be explainable for *every* $w \in \|KB\|$. One might argue that we should simply update those worlds for which explanations exist and ignore the others. This is reasonable, but it is no longer simply update; rather it is a combination of update and revision. Furthermore observations may often be unexplainable for every world in $\|KB\|$. For instance, suppose a solution is believed to be an acid and a litmus strip is dipped into it, which promptly turns blue. This is not explainable for any *KB*-world (it should turn red) in terms of event effects. Instead, the intuitive explanation (the solution is a base) requires that *KB* be revised before adopting the update observation. Finally notice that an observation need not be strictly unexplainable to force revision. Often an implausible explanation will suffice. For instance, a raise to $90,000 might not be impossible in Finance, but just so implausible that the database is willing to accept the fact that Fred is in Purchasing.

Issues of this sort make postulate (U3) (and certain aspects of (U5) and (U8)) somewhat questionable, and provides further motivation for adopting an abductive view of update. This perspective is especially fruitful when combining the process of update (changing knowledge) with belief revision (gaining knowledge). A model that puts both components together in a broader abductive framework is described in (Boutilier 1993; Boutilier 1994b). Roughly, the logic for belief revision set forth in (Boutilier 1994c) is used to capture the revision process, but is combined with elements of dynamic logic (Harel 1984) to capture the evolution of the world due to action occurrences.

Other have presented models of update that, like ours and unlike the KM-model, have their basis in reasoning about action. del Val and Shoham (1992; 1993), using the situation calculus, show how one can determine an update operator by reasoning about the changes induced by a given action. Very roughly, when some *KB* is to be updated by an observation O, they postulate the existence of some action A_O^{KB} whose predicted effects, when applied to the "situation" embodied by *KB*, determine the form of the update operator. Most critically, the effect axiom for such an action states that O holds when A_O^{KB} is applied to *KB*, and other effects are inferred via persistence mechanisms.

This model differs from ours in a number of rather important ways. First, del Val and Shoham assume that the update formula O describes the occurrence of some action or event. This severely restricts the scope of update, which in general can accept arbitrary propositions. They provide no mechanism for explaining an observation using the specification of *existing* actions. In order to deal with arbitrary observations an action is "invented" for the purpose of causing any observation in any situation. Naturally, the effects of such new actions are not specified *a priori* in the domain theory. So they propose that the effect of invented actions is to induce minimal change in the knowledge base according to some persistence mechanism. However, the plausible cause of an observation O may carry with it, in general, other drastic (rather than minimal) changes in *KB*. This can only be accounted for by explaining an observation in terms of existing actions. A persistence mechanism is required primarily because existing action or event specifications are not employed.

Another drawback of this model is its failure to account for the possibility that any of a number of actions might have caused O, and that update should reflect the most plausible of these causes. Finally, there is an assumption that the update of *KB* is due to the occurrence of a (known) single action. As we have described above, this will usually not be the case. Conditional explanations, explanations that use different actions for different "segments" of *KB*, will be very common.

A related mechanism is proposed by Goldszmidt and Pearl (1992), who use qualitative causal networks to represent an action theory. Again, update formula are implicitly assumed to be propositions asserting the occurrence of some action or event. An observation O is incorporated by assuming some proposition $do(O)$ has become true, and using a forced-action semantics to propagate its effects. Explanations are not given in terms of existing actions.

We should point out that both theories adopt a theory of action that provides a representation mechanism for actions and effects, as well as incorporating a solution to the frame problem (implicitly in the case of Goldszmidt and Pearl). We have side-stepped such issues by focusing on the semantics of update. We are currently investigating various action representations, such as STRIPS and the situation calculus, and the means they provide for generating conditional explanations. This is partially developed in (Boutilier 1993; Boutilier 1994b), where we provide a representation for actions using a conditional default logic to capture the defeasibility and nondeterminism of action effects, and use elements of dynamic logic to capture the evolution of the world. Action theories such as those exploited in (del Val and Shoham 1992; Goldszmidt and Pearl 1992) might also be used to greater advantage.

References

Alchourrón, C., Gärdenfors, P., and Makinson, D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530.

Boutilier, C. 1993. Explaining observations in reasoning about action. (manuscript).

Boutilier, C. 1994a. Conditional logics of normality: A modal approach. *Artificial Intelligence*. (in press).

Boutilier, C. 1994b. Two types of explanation in reasoning about action. Technical report, University of British Columbia, Vancouver. (Forthcoming).

Boutilier, C. 1994c. Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence*. (in press).

Dean, T., Kaelbling, L. P., Kirman, J., and Nicholson, A. 1993. Planning with deadlines in stochastic domains. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 574–579, Washington, D.C.

del Val, A. and Shoham, Y. 1992. Deriving properties of belief update from theories of action. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 584–589, San Jose.

del Val, A. and Shoham, Y. 1993. Deriving properties of belief update from theories of action (ii). In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 732–737, Chambery, FR.

Fikes, R. E. and Nilsson, N. J. 1971. Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208.

Goldszmidt, M. and Pearl, J. 1992. Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions. In *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, pages 661–672, Cambridge.

Harel, D. 1984. Dynamic logic. In Gabbay, D. and Guenthner, F., editors, *Handbook of Philosophical Logic*, pages 497–604. D. Reidel, Dordrecht.

Katsuno, H. and Mendelzon, A. O. 1991. 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*, pages 387–394, Cambridge.

Kushmerick, N., Hanks, S., and Weld, D. 1993. An algorithm for probabilistic planning. Technical Report 93-06-04, University of Washington, Seattle.

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

Morgenstern, L. and Stein, L. A. 1988. Why things go wrong: A formal theory of causal reasoning. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 518–523, St. Paul.

Poole, D. 1988. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95.

Reiter, R. 1991. The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. In Lifschitz, V., editor, *Artificial Intelligence and Mathematical Theory of Computation (Papers in Honor of John McCarthy)*, pages 359–380. Academic Press, San Diego.

Reiter, R. 1992. On specifying database updates. Technical Report KRR-TR-92-3, University of Toronto, Toronto.

Shanahan, M. 1993. Explanation in the situation calculus. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 160–165, Chambery, FR.

Winslett, M. 1988. Reasoning about action using a possible models approach. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 89–93, St. Paul.

Acknowledgements

Discussions with Ray Reiter have helped to clarify my initial thoughts on update. Thanks to Richard Dearden and David Poole for helpful discussions on this topic and to Alvaro del Val for well-considered comments on an earlier draft of this paper. This research was supported by NSERC Research Grant OGP0121843.