ELSEVIER

Artificial Intelligence 83 (1996) 143–166

Artificial Intelligence

Abduction to plausible causes: an event-based model of belief update ★

Craig Boutilier *

Department of Computer Science, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Received March 1994; revised November 1994

Abstract

The Katsuno and Mendelzon (KM) 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 conforming to the KM postulates 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 at least to the origins of the situation calculus [20]. 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 effect changes in the world, and agents must be able to modify their beliefs about the world to reflect such considerations. Furthermore, an agent situated in a dynamic world must be able

★ Some parts of this report appeared in preliminary form as "An event-based abductive model of update", in: *Proceedings Tenth Canadian Conference on AI*, Banff, Alta. (1994).

* E-mail: cebly@cs.ubc.ca.

SSDI 0004-3702(94)00097-2

to reason about changes in the world not simply due to its own actions, but due to the occurrence of exogenous events as well.

One of the most influential theories of belief change has been the *AGM theory* proposed by Alchourrón, Gärdenfors and Makinson [1]. 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 [27] 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 [16], 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 characterized 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). As a concrete example, suppose that someone observes that the grass in front of her house is wet. She is not sure whether she left her book outside on the patio, but concludes that if the book is outside it is wet too. There are two possible worlds admitted by her knowledge, O and $\overline{O}$ (the book is outside or it is not). When the first possibility is updated with the observation of wet grass, a wet book is the result. When the second possibility is updated, the book remains dry. The conditional belief $O \equiv W$ (the book is wet if and only if it was outside) is part of our agent's updated belief set.

In this paper, we present an abductive model of belief change suitable for updating beliefs in response to a changing world. While our semantics induces a class of belief change operators that is somewhat more general than Katsuno–Mendelzon (KM) update operators, the most compelling aspect of our model is the fact that it breaks the KM 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 or easily obtainable from users of a system. In the following, we use the term *update* to describe any process of belief change used to capture changes in belief due to change in the world (not simply those models conforming to the KM postulates).

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. In our example, there are several possible causes of wet grass, among them the sprinkler turning on automatically, or rain. If rain is the most plausible of these causing events, our agent concludes that everything on the patio is wet, including the book *if* it is out there. Had sprinkler been the most plausible explanation, a different conclusion would have been reached: the book would be dry

regardless of its location. It is these considerations that allow an agent to determine just what changes in the world are most plausible. Intuitively, information about the effects of events, as well as their relative plausibility, will be more readily available or easier to assess than a direct ordering of plausibility over possible "evolutions" of the world.

We formalize this notion in an abstract manner obtaining a class of *explanation-change* operators that are similar in spirit and intent to KM update operators, but somewhat more general. We note that explanation has often been closely linked with belief revision [12]. Indeed, Boutilier and Becher [5] present a model of abduction where explanations are determined by explicit belief revision. Given this connection and the fact that update can be viewed as an essentially abductive process, we may also take update to be a certain kind of belief revision. This stands in stark contrast with the accepted wisdom that update and revision are orthogonal forms of belief change. While we could cast our model as a form of belief revision, this would detract from the main point of the paper. However, we do elaborate on this connection in the concluding section.

In Section 2 we review the KM postulates for belief update and the KM semantics. In Section 3 we analyze this semantics more closely, and break it into more basic elements. We describe our abductive view of update and show its relationship to the KM model. In particular, we show that certain semantic assumptions naturally give rise to the KM theory; however, we argue that these assumptions are inappropriate as general update principles. We also briefly describe and characterize a special class of update operators. In Section 4, we analyze our model more deeply and discuss the connections to belief revision. We also argue that proper modification of belief states in response to observations in dynamic settings involves a combination of belief revision and belief update. Finally, we compare our construction to the model of update proposed by del Val and Shoham [8]. Proofs of the main results can be found in Appendix A.

2. The semantics of update

Katsuno and Mendelzon [16] 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, both of which we review here.

We assume the existence of some knowledge base KB, 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 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* [16] 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. We will call such a change in any world an "evolution" of that world. To capture this intuition, Katsuno and Mendelzon postulate a family of preorders

$$\{\leqslant_w:\ w \in W\},$$

where each $\leqslant_w$ is a reflexive, transitive relation over W. We interpret each such relation as follows: if $u \leqslant_w v$ then u is at least as plausible a change relative to w (or an evolution of w) as is v. Finally, a *faithfulness condition* is imposed: for every world w, the preorder $\leqslant_w$ has w as a minimum element; that is, $w <_w v$ for all $v \neq w$. Intuitively, this ensures that w is itself more plausible than any other evolution of w. [1]

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

$$\|KB \diamond O\| = \bigcup_{w \in \|KB\|} \left\{ \min_{\leqslant_w} \{v:\ v \models O\} \right\},$$

where $\min_{\leqslant_w} X$ is the set of minimal elements (w.r.t. $\leqslant_w$) 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 the KM update theory.

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 Fig. 1). If the orderings $\leqslant_w$ and $\leqslant_v$ 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)`,

[1] Katsuno and Mendelzon use the term *persistent* to describe such orderings.

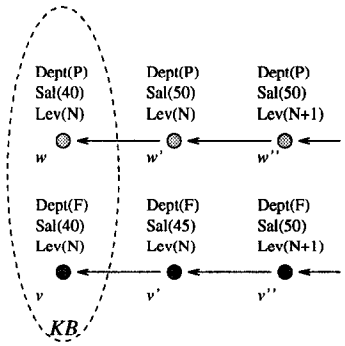

Fig. 1. An update model.

Dept(P) ⊃ Level(N) and Dept(F) ⊃ 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.

The KM semantics shows very clearly one of the main distinctions between update and belief revision. In belief revision (e.g., using the AGM theory), if an observation O is consistent with KB, then the revised $KB * O$ must be equivalent to $KB \cup \{O\}$. This needn't be the case for update. Given KB as above, we may receive an update transaction

$$O \equiv (\texttt{Dept(P)} \supset \texttt{Sal(40)}) \wedge (\texttt{Dept(F)} \supset \texttt{Sal(50)}).$$

While $KB \cup \{O\}$ entails Dept(P), and is captured semantically by the set $\{w\}$, $KB \diamond O$ corresponds to the set $\{w, v''\}$ and does not commit Fred to a particular department. The crucial distinction is update's willingness to consider the evolution of each possible world individually. Belief revision only considers the belief set KB as a whole.

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 a world might change. Given the role of update, this interpretation seems correct: worlds closer to w in the ordering $\leqslant_w$ 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.[2] 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.[3] 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. We do not presume that the agent has direct knowledge of the event occurrence. If such direct knowledge is available the problem becomes much simpler, for the agent can simply predict the effects of this event using some theory of action. This is a very specific update problem, restricting an agent to updating by observations of the form "Event E occurred". No explanation is required.[4]

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 [10] to the situation calculus [20,24] to more sophisticated probabilistic representations [7, 18]. 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.

---

[2] 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.

[3] 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).

[4] This assumption is embodied to a certain extent in the update models of del Val and Shoham [8,9] and Goldszmidt and Pearl [13], as we discuss in Section 4.

This provides a possible interpretation of the update process, and in our view, a very natural one.[5] Furthermore, as we describe in the concluding section (and in detail in [4]), 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 [21] propose a model where an observation that conflicts with the predicted effects of an agent's action causes the agent to infer the existence of some external event occurrence. Shanahan [26] 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 can account for the *update* of a knowledge base. We should point out that Reiter [25] (and personal communication) has informally suggested that update can be viewed as explanation to events causing an observation. We will 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 $e : W \rightarrow 2^W$. 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. In addition, we insist that $e(w) \neq \emptyset$ for each e, w.[6] We emphasize that not only are all possible outcomes of an event captured by the set $e(w)$, but also that each world in $e(w)$ is a legitimate, plausible outcome.

[5] 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.

[6] 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 in 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.

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 [6, 18]. The key feature of these, and other representations, is that each action/event induces a function between worlds (or worlds and sets of worlds).[7] Thus, most action representations will fit within this abstract model. While we do not delve into the representation of actions, our examples will suggest ways in which traditional representations can be augmented with the features of our 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 [4]).

In order to explain certain observations by appeal to plausible event occurrences, we need some metric for ranking 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.[8]

We require that $\preceq_w$ be a preorder for each w, and will occasionally assume that $\preceq_w$ is a total preorder. Once again, we do not expect that this family of orderings will be presented explicitly. Compact representation schemes are possible. For example, in our database example we might suppose that a user can specify the constraint that a `RAISE` event is more plausible than a `PROMOTION` event for employees of the Purchasing department. The relative plausibility need not be asserted explicitly for each world satisfying `Dept(P)`.

We note that there are few restrictions on the relative plausibility of events in any given ordering $\preceq_w$. The only structural basis for the logical comparison of events is through outcome sets, but these provide no logical constraints on relative plausibility. If we have two events e and f such that $e(w) \subseteq f(w)$, we impose no constraints on the relative ordering of e and f in $\preceq_w$. In particular, we cannot insist that an event e with fewer possible outcomes be judged more likely than an event f. For instance, imagine two events, *flipping* a coin and *placing* a coin, such that flipping results in two possible outcomes (heads, tails) and placing has three outcomes (heads, tails, edge). This provides no *a priori* reason to consider flipping or placing more likely than the other.

Putting these ingredients together, we have the following definitions:

[7] 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 [24] induces just such a mapping.

[8] Other models of event orderings are possible, including using a fixed ordering for all worlds, or associating event plausibility with belief sets (or sets of worlds) rather than individual worlds. However, these seem less compelling.

Definition 1. 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 $e : W \rightarrow 2^W$) and $\preceq$ is an indexed family of events orderings $\{\preceq_w : w \in W\}$ (where each $\preceq_w$ is a preorder over E).

Definition 2. 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 3. Let O be some proposition and $w \in W$. The set of *weak explanations* of O relative to w is

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

An event e is a *weak explanation* of O relative to w iff $e \in Expl(O, w)$. If $Expl(O, w) = \emptyset$, 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 outcome 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 of *some* possible outcome. This is often the most we can expect in a domain with nondeterministic events. For example, someone tossing a coin onto a chess board is a quite reasonable explanation for the fact that the coin is on a black square; but knowing the event occurred is not enough to predict that outcome, for it might well have landed on a white square.

A predictive explanation is similar, but we insist that *each* outcome of e satisfies O.

Definition 4. The set of *predictive explanations* of O relative to w is

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

An event e is a *predictive explanation* of O relative to w iff $e \in Expl_P(O, w)$. If $Expl_P(O, w) = \emptyset$, we say that O is *not predictively explainable* relative to w.

The distinction between weak and predictive explanations is very similar to that made between *consistency-based* diagnosis [23] and *predictive* (or *abductive*) diagnosis [22]. This distinction is illustrated in Fig. 2. Both e and f are nondeterministic events. Event e predictively explains O, while f weakly explains O but does not predictively explain O. We are interested here in weak explanations, for these seem most appropriate when dealing with nondeterministic events. However, we note the following:

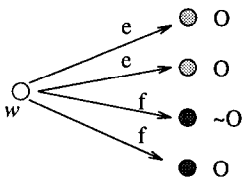

Fig. 2. Weak and predictive explanations.

Proposition 5. *If e is a deterministic event, then e weakly explains O iff e predictively explains O.*

Corollary 6. *If EM is a deterministic event model, O is weakly explainable iff O is predictively explainable.*

For a particular world w, $Expl(O, w)$ denotes those most plausible events that could cause O to be true. The possibilities admitted by such a set of explanations are the possible results of each of these events. To determine these we simply evolve or *progress* w in accordance with these possible event occurrences; that is:

Definition 7. The *progression* of world w given observation O is the set of worlds

$$Prog(w \mid O) = \bigcup \{e(w) \cap \|O\|: \ e \in Expl(O, w)\}.$$

Note that if O is unexplainable relative to w, then $Prog(w \mid O) = \emptyset$. This means that there is no event (among those specified in the model) that could have caused w to evolve into a world that satisfies O. The occurrence of O relative to w is impossible. We also note that if we restrict our attention to predictive explanations, or to deterministic event models, we can rewrite this definition as

$$Prog(w \mid O) = \bigcup \{e(w): \ e \in Expl_P(O, w)\}.$$

Taking a cue from the Katsuno–Mendelzon update semantics, the progression of a knowledge base KB given a particular observation O is obtained by considering all plausible evolutions of each world $w \in \|KB\|$. However, if O is unexplainable for some $w \in \|KB\|$, we take O to be unexplainable relative to KB as a whole.

Definition 8. The *progression* of KB given observation O is the set of worlds

$$Prog(KB \mid O) = \bigcup \{Prog(w \mid O): \ w \in \|KB\|\}.$$

If $Prog(w \mid O) = \emptyset$ for some $w \in \|KB\|$, we let $Prog(KB \mid O) = \emptyset$.

The motivation for this last condition, that O must be explainable relative to every $w \in \|KB\|$, comes from the KM update semantics itself. In the KM theory of update, the updated KB is constructed by considering the possible evolution of *every* possibility admitted by KB. We duplicate this intuition by considering the progression function

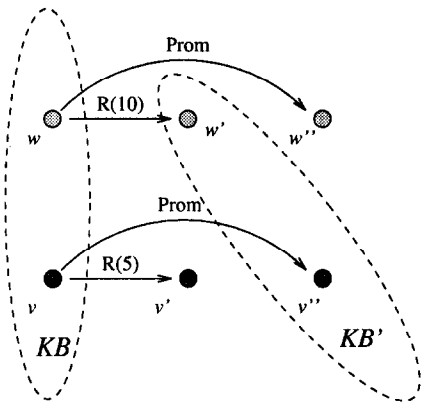

Fig. 3. An event ordering.

of every world in $\|KB\|$. If no such evolution is possible for one of these worlds, we trivialize the result of updating KB. We *might* have allowed the progression of KB to be nontrivial even if some worlds could not evolve so as to satisfy O, and define $Prog(KB \mid O)$ without this last condition. In other words, we might have considered $Prog(KB \mid O)$ to be as in the definition, but simply accept, when $Prog(w \mid O) = \emptyset$, that w contributes nothing to the construction of $Prog(KB \mid O)$. However, we adopt the current approach for two reasons. First, our goal is to pursue the analogy with the KM update semantics. Our definition is a direct adaptation. Second, dropping this restriction has implications for the relationship between belief revision and update. Simply excluding worlds whose progression is empty is, in effect, performing revision in addition to update. While this is generally a good idea, the correct way to bring together revision and update requires more drastic changes in the way update is performed. We elaborate on this in Section 4.

We take progression of KB to be the semantic counterpart of the update of the theory KB. With such a progression function, we can now define the *explanation-change operator* relative to a given event model, which determines the consequences of adopting an observation.

Definition 9. The *explanation-change operator* induced by an event model EM is $\diamond_{EM}$:

$$KB \diamond_{EM} O = \{A \in L_{CPL}: Prog(KB \mid O) \models A\}.$$

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

Purchasing: `RAISE(10)` $\prec$ `PROMOTION` $\prec$ `RAISE(5)`,

Finance: `RAISE(5)` $\prec$ `PROMOTION` $\prec$ `RAISE(10)`.

This is illustrated in Fig. 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 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).

Finally, we can formalize our initial example. We first adopt a conditional STRIPS-like representation of events, using variables to schematically capture a set of propositions, and take each event specification to induce the obvious transformation on possible worlds (see, e.g., [6, 18]). We have two possible events, RAIN and SPRINKLER, with effects as follows:

Event	*Condition*	*Effect*
RAIN	On(grass,x)	Wet(x), Wet(grass), Wet(patio)
	On(patio,x)	Wet(x), Wet(grass), Wet(patio)
	else	Wet(grass), Wet(patio)
SPRINKLER	On(grass,x)	Wet(x), Wet(grass)
	else	Wet(grass)

We also have the proposition O asserting that it is overcast, and influencing the plausibility of these two events. A plausibility ordering might be given as follows:

If O then RAIN $\prec$ SPRINKLER,

If $\neg$O then SPRINKLER $\prec$ RAIN.

Our agent's knowledge base consists of the beliefs

$$\{\mathtt{O}, \neg\mathtt{Wet(book)}, \mathtt{On(patio,book)} \equiv \neg\mathtt{Inside(book)}\}.$$

Given the fact O (overcast), the most plausible explanation for the observation Wet(grass) is RAIN. The effect is then Wet(book) if On(patio,x) and $\neg$Wet(book) if Inside(book). Note that had it not been overcast, SPRINKLER would have been the most plausible explanation and our agent would rest assured that her book is dry.

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 within $\|KB\|$. 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 the Katsuno–Mendelzon theory

We are interested in the question of whether the explanation-change operator satisfies the KM update postulates. This is not the case given the formulation above.

Proposition 10. *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 in 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 if we assume $v \models A$, 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 11. The *null event* is an event n, where $n(w) = \{w\}$ for all $w \in W$.

Definition 12. 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 in any world. This seems to be the crucial assumption underlying postulate (U2).

Proposition 13. *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. As an example, consider an agent monitoring a control system producing some product. It observes a display that indicates whether the system is proceeding normally. If it believes that a normal condition is displayed before the *next* observation, observing that the display (still) indicates normal should not require that its other beliefs *not* change: it may, for instance,

update the number of units produced in response to this observation. 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 $Prog(w \mid O) = \emptyset$ 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 14. 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 15. *If EM is a complete event model then $Prog(KB \mid O) \neq \emptyset$ for any consistent O and KB.*

This condition is sufficient to ensure (U5) and (U8) are satisfied as well.

Proposition 16. *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 17. 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 18. *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 13 and 16, we obtain the main representation result for explanation-change.

Theorem 19. *Let $\diamond_{EM}$ be the explanation-change operator induced by some complete, centered event model. Then $\diamond_{EM}$ satisfies update postulates* (U1)–(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 20. 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: $v \leqslant_w u$ 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 $e_v \preceq_w e_u$.

Intuitively, the more plausible some causing event e_v for world v is (relative to w), the more plausible evolution v of world w is deemed to be (according to $\leqslant_w$).

Theorem 21. *Let $\{\leqslant_w: w \in W\}$ be the family of plausibility orderings induced by some complete, centered event model EM. Then*:

(a) *Each relation $\leqslant_w$ is a faithful preorder over W.*
(b) *The change operation determined by $\{\leqslant_w: w \in W\}$ is a KM update operator.*
(c) *The update operator determined by $\{\leqslant_w: w \in W\}$ is equivalent to the explanation-change operator $\diamond_{EM}$.*

We note that if the event model is not centered then the generated preorder is not necessarily faithful. If the model is not complete, then we have only a restricted form of faithfulness. It will be the case that $w <_w v$ for any world v that is a possible evolution of w (i.e., if $v \in e(w)$ for some $e \in E$). However, those worlds that cannot result from the application of some event to w will be unrelated to w. In this case, we can say that $\leqslant_w$ is faithful relative to the *connected component* of $\leqslant_w$ that includes w. Intuitively, we want to ignore those worlds that are not "reachable" from w. To do this we can simply define an update operator using the relation $\leqslant_w$ restricted to such worlds:

$$\{v: v \in e(w) \text{ for some } e \in E\}.$$

This ensures that unrelated worlds are not trivially minimal in the ordering relation $\leqslant_w$.

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

Proposition 22. *Let $EM = \langle W, E, \preceq \rangle$ be a complete event model such that $\preceq_w$ is a total preorder for each $w \in W$. Then each plausibility ordering $\leqslant_w$ induced by EM is a total preorder.*

The circumstance where a set of events is totally preordered by plausibility may arise rather frequently; for instance, events may be ranked according to some integer scale,

or assigned some qualitative probability ranking. Therefore, 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. [3, 19]).

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

Theorem 23. *An update operator* $\diamond$ *satisfies postulates* (U1)–(U9) *iff there exists an appropriate family of faithful* total *preorders* $\{\leqslant_w: w \in W\}$ *that induces* $\diamond$ *(in the usual way).*

Corollary 24. *Let* $\diamond_{EM}$ *be the explanation-change operator induced by some total order event model. Then* $\diamond_{EM}$ *satisfies postulate* (U9).

Indeed, Katsuno and Mendelzon [16] also discuss the possibility of totally ordered plausibility rankings and provide a postulate (U9) related to the one above, and the proof of equivalence is similar to that suggested by them (see also their work on total orders for belief revision [17]).

As a final remark, we note that the converses of Theorems 19 and 21 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 is 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.

4.1. Relationship to belief revision

It has frequently been suggested that abduction can be modeled by appeal to belief revision [12]. Boutilier and Becher [5] present a model of abduction along these lines, whereby an explanation for an observation O, with respect to some KB, is a sentence E such that $KB * E$ entails O. In other words, had an agent believed the explanation E it would have believed the observation O.

The abductive view of update suggests that update may also be viewed as a form of belief revision. Since explanations take the form of event occurrences, an interpretation

using belief revision takes update to be the process of an agent revising its beliefs about whether an event occurred and just what that event was. For instance, in our main example the observation Sal(50000) can be viewed as causing an agent to *give up* its belief that no event has occurred (i.e., the world has not changed) and *accept* the fact that something has happened—in particular, it accepts the belief

$$\texttt{Dept(P)} \supset \textit{Occurred}(\texttt{RAISE(10)}) \wedge \texttt{Dept(F)} \supset \textit{Occurred}(\texttt{PROMOTION}).$$

Of course, we have not provided an explicit logical language for the representation of actions or events, and in particular, have not provided a method for revising beliefs about such occurrences. However, there are a number of ways to model this type of belief revision, including using *histories* or *runs* of a system as our basic semantic objects. A run is essentially a sequence of world states capturing a particular evolution of a system. Using these as semantic primitives one can capture beliefs about the actual state of the world in addition to event occurrences. While not directly suited to our task, the revision model of Friedman and Halpern [11], in which runs are ranked in manner suitable for belief revision, is precisely the type of system upon which a more elaborate model of update, revision and explanation can be built.

When viewed in this way, certain problems with the update model, as formulated by Katsuno and Mendelzon and recast here, become apparent. The types of explanations one is willing to consider are restricted to event occurrences. In other words, an agent is bound to revise its beliefs only about possible event occurrences and their consequences. Thus, an agent making an observation is not allowed to entertain the possibility that its knowledge base *KB* was incomplete or incorrect. It can only change its beliefs about the *post-event* world state. Semantically, this restriction is apparent in our definition of update (as well as Katsuno and Mendelzon's). We require that *every* $w \in \|KB\|$ be progressed according to likely explaining events.

It is just this restriction that calls into question the suitability of update as a "stand-alone" belief change operator. 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 this salary change if Fred is indeed in Finance. Far from being a miraculous occurrence, it suggests that Fred is 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 our definition of update; one might argue that we should simply update those worlds for which explanations exist and ignore the others. This minor adjustment seems 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, when a litmus strip is dipped into it and promptly turns blue. This is not explainable for any *KB*-world (it should turn red) in terms of event effects. As such, the minor adjustment of our definition of update is not sufficient. We may want to update *KB* consistently in circumstances where no possible event could give rise to

the observation *given our current state of belief*. Instead, the intuitive explanation in this example consists of two parts: the first postulates that the event of dipping the paper in the solution occurred; the second suggests that the solution is in fact a base. This requires revision of *KB*—we must change our beliefs about the *pre-event* state of the world in order to modify *KB* correctly.

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. To adequately reflect such considerations, we must have the ability to compare the plausibility of event occurrences with the plausibility of beliefs about the world state. This provides further support for more expressive models and languages in which event occurrences can be reasoned with explicitly.

Issues of this sort make postulate (U3) (and certain aspects of (U5) and (U8)) somewhat questionable, and provides 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 [4]. Roughly, the logic for belief revision set forth in [2] is used to capture the revision process, but is combined with elements of dynamic logic [14] to capture the evolution of the world due to action occurrences.

4.2. Related work

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 [8,9], 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 [13], 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 proposals 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 [4], 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 [8, 13] might also be used to greater advantage.

Appendix A. Proofs of main results

Proposition 10. *Let $\diamond_{EM}$ be the explanation-change operator induced by some event model. Then $\diamond_{EM}$ satisfies postulates* (U1), (U4), (U6) *and* (U7).

Proof. Assume an event model $M = \langle W, E, \preceq \rangle$ and associated update operator $\diamond$ (for simplicity we drop the subscript). We show in turn that each of these postulates is satisfied.

(U1) By definition $Res(A, w) \subseteq \|A\|$ for all w (hence for all $w \in \|KB\|$). Immediately we have $KB \diamond A \models A$.

(U4) Suppose $\models A \equiv B$. Then e explains A w.r.t KB iff e explains B, for any event e. Thus, $KB \diamond A \equiv KB \diamond B$.

(U6) Suppose $KB \diamond A \models B$ and $KB \diamond B \models A$. Then we have $Res(A, KB) \subseteq \|B\|$ and $Res(B, KB) \subseteq \|A\|$. If $v \in Res(A, KB)$, then for some $w \in \|KB\|$ we have a most plausible explaining event e for A such that $v \in e(w)$. However, since $v \in \|B\|$, e must also be a most plausible explaining event for B as well; otherwise there must exist some more plausible event $f \prec e$ that explains B, contradicting the fact that e is most plausible for A (since $Res(B, KB) \subseteq \|A\|$). Therefore, $v \in Res(B, KB)$, so $Res(A, KB) \subseteq Res(B, KB)$. By symmetry the reverse containment holds, so $Res(A, KB) = Res(B, KB)$ and we have $KB \diamond A \equiv KB \diamond B$.

(U7) Let KB be complete so that $\|KB\| = \{w\}$ for some world w. Suppose $v \in Res(A, w) \cap Res(B, w)$. (If there is no such v then $(KB \diamond A) \wedge (KB \diamond B)$ is inconsistent and (U7) holds trivially.) Then there is some e such that $v \in e(w)$ and e is a most plausible explaining event for both A and B. This ensures that e explains $A \vee B$ and that $v \in Res(A \vee B, w)$. Hence, $Res(A, w) \cap Res(B, w) \subseteq Res(A \vee B, w)$. Therefore, $(KB \diamond A) \wedge (KB \diamond B) \models KB \diamond (A \vee B)$. □

Proposition 13. *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).

Proof. Given Proposition 10, we need only show that (U2) is satisfied by centered event models. Assume that $M = \langle W, E, \preceq \rangle$ is such a model, inducing update operator $\diamond$ (for simplicity we drop the subscript). Suppose $KB \models A$. Then for each $w \in \|KB\|$, we have $w \models A$; and the null event is the most plausible explaining event for each such world. Thus $Res(A, KB) = \|KB\|$ and $KB \diamond A$ is equivalent to KB. □

Proposition 16. *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).

Proof. Given Proposition 10, we need only show that (U3), (U5) and (U8) are satisfied by complete models. Assume that $M = \langle W, E, \preceq \rangle$ is such a model, inducing update operator $\diamond$ (for simplicity we drop the subscript).

(U3) Since M is complete, any satisfiable A is explainable for every w. So if KB is satisfiable, $Res(A, KB) \neq \emptyset$ and $KB \diamond A$ is satisfiable.

(U5) Let $v \in Res(A, KB) \cap \|B\|$ (if there is no such v, then (U5) holds trivially). Then for some $w \in \|KB\|$ there is a most plausible explanation e for A (w.r.t. w) such that $v \in e(w)$. This event e must also be a most plausible explanation for $A \wedge B$ w.r.t. w (otherwise some more plausible event would explain $A \wedge B$, hence A). Thus $v \in Res(A \wedge B, KB)$. Notice that since $A \wedge B$ must be explainable for every world, we cannot have that $Res(A \wedge B, KB)$ is set to the empty set. Thus, $Res(A, KB) \cap \|B\| \subseteq Res(A \wedge B, KB)$ and $(KB \diamond A) \wedge B \models KB \diamond (A \wedge B)$.

(U8) Assume that $KB_1 \vee KB_2$ is satisfiable. We have $w \in Res(A, KB_1 \vee KB_2)$ iff there is some $v \in \|KB_1 \vee KB_2\|$ such that $w \in Res(A, v)$. Such a v is either in $\|KB_1\|$ or $\|KB_2\|$, so this holds iff $w \in Res(A, KB_1) \cup Res(A, KB_2)$. Therefore, $(KB_1 \vee KB_2) \diamond A \equiv (KB_1 \diamond A) \vee (KB_2 \diamond A)$. □

Theorem 21. *Let $\{\leqslant_w : w \in W\}$ be the family plausibility orderings induced by some complete, centered event model EM. Then:*

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

(b) *The change operation determined by $\{\leqslant_w : w \in W\}$ is a* KM *update operator.*

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

Proof. Let w be some world in a complete, centered event model $EM = \langle W, E, \preceq \rangle$, and let $\leqslant_w$ be an induced ordering.

(a) That $\leqslant_w$ is reflexive and transitive follows immediately from the definition $\leqslant_w$ in terms of the event ordering $\preceq_w$ and the fact that $\preceq_w$ is itself reflexive and transitive. Hence $\leqslant_w$ is a preorder. Since null action $n \prec_w e$ for all $e \neq n$ and $n(w) = \{w\}$, we have $w \prec_w v$ for all $v \neq w$ (if $v \in e_v(w)$ for some event e_v). Thus $\leqslant_w$ is faithful. Finally, since EM is complete, for any v there is some e_v such that $v \in e_v(w)$. Thus $\leqslant_w$ is persistent.

(b) The representation theorem of Kastuno and Mendelzon ensures that the family of orderings $\{\leqslant_w: w \in W\}$ generates an update operator satisfying (U1)–(U8).

(c) Denote by $\diamond$ the update operator generated by $\{\leqslant_w: w \in W\}$. We will show that $KB \diamond A \equiv KB \diamond_{EM} A$ for any consistent KB and A. Assume $v \in \|A\|$. We have

$$
\begin{aligned}
v \in \|KB \diamond A\| \quad \text{iff} \quad & v \in \bigcup_{w \in \|KB\|} \left\{ \min_{\leqslant_w} \{v: v \models A\} \right\} \\
\text{iff} \quad & \text{for some } w \in \|KB\|, \text{ if } u \leqslant_w v, \text{ then } u \not\models A \\
\text{iff} \quad & \text{for some } w \in \|KB\| \text{ and event } e,\ v \in e(w) \text{ and} \\
& \text{for all } e' \prec_w e, \text{ we have } e'(w) \cap \|A\| = \emptyset \\
\text{iff} \quad & \text{for some } w \in \|KB\|,\ v \in Res(A, w) \\
\text{iff} \quad & v \in Res(A, KB) \\
\text{iff} \quad & v \in \|KB \diamond A\|. \qquad \square
\end{aligned}
$$

Proposition 22. *Let $EM = \langle W, E, \preceq \rangle$ be a complete event model such that $\preceq_w$ is a total preorder for each $w \in W$. Then each plausibility ordering $\leqslant_w$ induced by EM is a total preorder.*

Proof. Theorem 21 ensures that $\leqslant_w$ is a preorder. For any world u, let e_u denote any event that has outcome u relative to w; i.e., $u \in e_u(w)$ (such an event must exist since EM is complete). Consider two worlds u and v. Suppose $v \not\leqslant_w u$. Then there must be some e_u such that for all e_v, $e_v \not\preceq_w e_u$. Since $\preceq_w$ is a total preorder, $e_u \not\preceq_w e_v$ for all such e_v; and $u \leqslant_w v$. Thus, $\leqslant_w$ is a total preorder. $\square$

Theorem 23. *An update operator $\diamond$ satisfies postulates* (U1)–(U9) *iff there exists an appropriate family of faithful* total *preorders $\{\leqslant_w: w \in W\}$ that induces $\diamond$ (in the usual way).*

Proof. We first assume a suitable family of preorders. The representation result of Katsuno and Mendelzon ensures that the induced update operator $\diamond$ satisfies (U1)–(U8). We now show that it also satisfies (U9). Let KB be complete with $\|KB\| = \{w\}$. Suppose $KB \diamond A \not\models \neg B$ and $KB \diamond A \models C$. Let $\min(A)$ denote the set $\min_{\leqslant_w}\{v: v \models A\}$. Then we have $\min(A) \subseteq \|C\|$ and $\min(A) \cap \|B\| \neq \emptyset$. Since $\leqslant_w$ is a total preorder,

$$\min(A \wedge B) \subseteq \min(A) \cap \|B\| \subseteq \|C\|,$$

so $KB \diamond (A \wedge B) \models C$. Therefore (U9) is satisfied.

Now we suppose $\diamond$ satisfies postulates (U1)–(U9). To prove that a suitable family of orderings exist, we adopt the basic technique of Katsuno and Mendelzon [15]. However, the orderings are constructed in a rather different fashion to ensure that the preorders are total. As preliminary notation, for any set of worlds X, we write Φ_X to denote some sentence such that $\|\Phi_X\| = X$. To emphasize that this will be the object of revision, we write $KB_{\{w\}}$ instead of $\Phi_{\{w\}}$. We note that (U1) and (U3) ensure that $KB_{\{w\}} \diamond \Phi_X \equiv \Phi_{X'}$ for some $X' \subseteq X$. We will also make use of the fact that $\Phi_{X'} \models \Phi_X$ for any $X' \subseteq X$. We can now define a family of ordering relations based on $\diamond$ as follows:

$$v \leqslant_w u \quad \text{iff} \quad v \in \|KB_{\{w\}} \diamond \Phi_{\{v,u\}}\|.$$

We first show that $\leqslant_w$ is a faithful persistent preorder. Clearly, $\leqslant_w$ is reflexive, since (U1) and (U3) ensure that $v \in \|KB_{\{w\}} \diamond \Phi_{\{v\}}\|$. Similarly, at least one of v or u is in $\|KB_{\{w\}} \diamond \Phi_{\{v,u\}}\|$, so either $v \leqslant_w u$ or $u \leqslant_w v$. It is easy to verify that $\leqslant_w$ is faithful and persistent due to (U2) and (U3). It simply remains to verify that $\leqslant_w$ is transitive. So suppose $v \leqslant_w u$ and $u \leqslant_w t$. This ensures that $v \in \|KB_{\{w\}} \diamond \Phi_{\{v,u\}}\|$ and $u \in \|KB_{\{w\}} \diamond \Phi_{\{u,t\}}\|$.

(a) Suppose $KB_{\{w\}} \diamond \Phi_{\{v,u,t\}} \equiv \Phi_{\{t\}}$. Then

$$KB_{\{w\}} \diamond \Phi_{\{v,u,t\}} \not\models \neg\Phi_{\{u,t\}} \quad \text{and} \quad KB_{\{w\}} \diamond \Phi_{\{v,u,t\}} \models \Phi_{\{t\}}.$$

By (U9),

$$KB_{\{w\}} \diamond (\Phi_{\{v,u,t\}} \wedge \Phi_{\{u,t\}}) \models \Phi_{\{t\}},$$

or equivalently

$$KB_{\{w\}} \diamond \Phi_{\{u,t\}} \models \Phi_{\{t\}}.$$

But this contradicts the fact that $u \in \|KB_{\{w\}} \diamond \Phi_{\{u,t\}}\|$.

(b) Suppose $KB_{\{w\}} \diamond \Phi_{\{v,u,t\}} \not\models \neg\Phi_{\{u\}}$ and $KB_{\{w\}} \diamond \Phi_{\{v,u,t\}} \models \neg\Phi_{\{v\}}$. Then

$$KB_{\{w\}} \diamond \Phi_{\{v,u,t\}} \not\models \neg\Phi_{\{v,u\}} \quad \text{and} \quad KB_{\{w\}} \diamond \Phi_{\{v,u,t\}} \models \Phi_{\{u,t\}}.$$

By (U9),

$$KB_{\{w\}} \diamond (\Phi_{\{v,u,t\}} \wedge \Phi_{\{v,u\}}) \models \Phi_{\{u,t\}},$$

or equivalently

$$KB_{\{w\}} \diamond \Phi_{\{v,u\}} \models \Phi_{\{u,t\}}.$$

But this contradicts the fact that $v \in \|KB_{\{w\}} \diamond \Phi_{\{v,u\}}\|$ and $v \notin \|\Phi_{\{u,t\}}\|$. Thus, if $u \in \|KB_{\{w\}} \diamond \Phi_{\{v,u,t\}}\|$ then $v \in \|KB_{\{w\}} \diamond \Phi_{\{v,u,t\}}\|$.

By (a) and (b), we know that $KB_{\{w\}} \diamond \Phi_{\{v,u,t\}}$ is not equivalent to $\Phi_{\{t\}}$, $\Phi_{\{u\}}$ or $\Phi_{\{u,t\}}$. Thus, $v \in \|KB_{\{w\}} \diamond \Phi_{\{v,u,t\}}\|$.

Now by (U5) we have $(KB_{\{w\}} \diamond \Phi_{\{v,u,t\}}) \wedge \Phi_{\{v,t\}} \models KB_{\{w\}} \diamond \Phi_{\{v,t\}}$. Since $v \in \|KB_{\{w\}} \diamond \Phi_{\{v,u,t\}}\|$, this conjunction is consistent, thus $v \in \|KB_{\{w\}} \diamond \Phi_{\{v,t\}}\|$. Therefore $v \leqslant_w t$ and $\leqslant_w$ is transitive.

Finally, we demonstrate that

$$\|KB \diamond A\| = \bigcup_{w \in \|KB\|} \left\{ \min_{\leqslant_w}\{v\colon\ v \models A\} \right\}.$$

The remainder of the proof follows closely that of Katsuno and Mendelzon, but we include it for completeness. We assume KB and A are consistent, for the relations hold trivially otherwise. We first show that this relation holds for any complete KB. Assume $\|KB\| = \{w\}$. We use $\min(A)$ to denote the set of minimal A-worlds in $\leqslant_w$.

Suppose $v \models KB \diamond A$ but $v \notin \min(A)$. Then there is some $u <_w v$ such that $u \models A$. By (U5), $(KB \diamond A) \wedge \Phi_{\{v,u\}} \models KB \diamond \Phi_{\{v,u\}}$; and by definition of $\leqslant_w$, $KB \diamond \Phi_{\{v,u\}} \equiv \Phi_{\{u\}}$. But then $v \not\models KB \diamond A$. So v must be in $\min(A)$. Thus, $\|KB \diamond A\| \subseteq \min(A)$.

Now suppose $v \in \min(A)$. Let $\|A\| = \{u_1, \ldots, u_n\}$. We have that

$$A \equiv \Phi_{\{v,u_1\}} \vee \cdots \vee \Phi_{\{v,u_n\}}.$$

And since $v \leqslant u_i$ for each $i \leqslant n$ (since $v \in \min(A)$), we have $v \in \|KB \diamond \Phi_{\{v,u_i\}}\|$ for each $i \leqslant n$. That is, v satisfies

$$(KB \diamond \Phi_{\{v,u_1\}}) \wedge \cdots \wedge (KB \diamond \Phi_{\{v,u_n\}}).$$

By (U7), v therefore satisfies

$$KB \diamond (\Phi_{\{v,u_1\}} \vee \cdots \vee \Phi_{\{v,u_n\}}).$$

That is, $v \in \|KB \diamond A\|$. Therefore, $\min(A) \subseteq \|KB \diamond A\|$.

The result holds for any complete KB. However, any KB is equivalent to the disjunction of some finite set of complete KBs. Thus, by (U8) we have

$$\|KB \diamond A\| = \bigcup_{w \in \|KB\|} \left\{ \min_{\leqslant_w}\{v\colon\ v \models A\} \right\}. \qquad \square$$

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 and David Makinson for well-considered comments on an earlier draft of this paper. Thanks also to the referees for suggestions that helped clarify the presentation. This research was supported by NSERC Research Grant OGP0121843.

References

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

[2] C. Boutilier, Unifying default reasoning and belief revision in a modal framework, *Artif. Intell.* **68** (1994) 33–85.

[3] C. Boutilier, Conditional logics of normality: a modal approach, *Artif. Intell.* **68** (1994) 87–154.
[4] C. Boutilier, Generalized update: Belief change in dynamic settings, Manuscript (1994); shortened version in: *Proceedings IJCAI-95*, Montreal, Que. (1995) 1104–1111.
[5] C. Boutilier and V. Becher, Abduction as belief revision, *Artif. Intell.* **77** (1995) 43–94.
[6] C. Boutilier and R. Dearden, Using abstractions for decision-theoretic planning with time constraints, in: *Proceedings AAAI-94*, Seattle, WA (1994) 1016–1022.
[7] T. Dean, L.P. Kaelbling, J. Kirman and A. Nicholson, Planning with deadlines in stochastic domains, in: *Proceedings AAAI-93*, Washington, DC (1993) 574–579.
[8] A. del Val and Y. Shoham, Deriving properties of belief update from theories of action, in: *Proceedings AAAI-92*, San Jose, CA (1992) 584–589.
[9] A. del Val and Y. Shoham, Deriving properties of belief update from theories of action II, in: *Proceedings IJCAI-93*, Chambery (1993) 732–737.
[10] R.E. Fikes and N.J. Nilsson, Strips: A new approach to the application of theorem proving to problem solving, *Artif. Intell.* **2** (1971) 189–208.
[11] N. Friedman and J.Y. Halpern, A knowledge-based framework for belief change, part II: revision and update, in: *Proceedings Fourth International Conference on Principles of Knowledge Representation and Reasoning*, Bonn, Germany (1994) 190–201.
[12] P. Gärdenfors, *Knowledge in Flux: Modeling the Dynamics of Epistemic States* (MIT Press, Cambridge, MA, 1988).
[13] M. Goldszmidt and J. Pearl, Rank-based systems: a simple approach to belief revision, belief update, and reasoning about evidence and actions, in: *Proceedings Third International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, MA (1992) 661–672.
[14] D. Harel, Dynamic logic, in: D. Gabbay and F. Guenthner, eds., *Handbook of Philosophical Logic* (Reidel, Dordrecht, Netherlands, 1984) 497–604.
[15] H. Katsuno and A.O. Mendelzon, On the difference between updating a knowledge database and revising it, Tech. Rept. KRR-TR-90-6, University of Toronto, Toronto, Ont. (1990).
[16] H. Katsuno and A.O. Mendelzon, On the difference between updating a knowledge database and revising it, in: *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, MA (1991) 387–394.
[17] H. Katsuno and A.O. Mendelzon, Propositional knowledge base revision and minimal change, *Artif. Intell.* **52** (1991) 263–294.
[18] N. Kushmerick, S. Hanks and D. Weld, An algorithm for probabilistic least-commitment planning, in: *Proceedings AAAI-94*, Seattle, WA (1994) 1073–1078.
[19] D. Lehmann, What does a conditional knowledge base entail? in: *Proceedings First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, Ont. (1989) 212–222.
[20] J. McCarthy and P. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: B. Meltzer and D. Michie, eds. *Machine Intelligence* **4** (Edinburgh University Press, 1969) 463–502.
[21] L. Morgenstern and L.A. Stein, Why things go wrong: A formal theory of causal reasoning, in: *Proceedings Seventh National Conference on Artificial Intelligence*, St. Paul, MN (1988) 518–523.
[22] D. Poole, A logical framework for default reasoning, *Artif. Intell.* **36** (1988) 27–47.
[23] R. Reiter, A theory of diagnosis from first principles, *Artif. Intell.* **32** (1987) 57–95.
[24] R. Reiter, The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression, in: V. Lifschitz, ed., *Artificial Intelligence and Mathematical Theory of Computation (Papers in Honor of John McCarthy)* (Academic Press, San Diego, CA, 1991) 359–380.
[25] R. Reiter, On specifying database updates, Tech. Rept. KRR-TR-92-3, University of Toronto, Toronto, Ont. (1992).
[26] M. Shanahan, Explanation in the situation calculus, in: *Proceedings IJCAI-93*, Chambery (1993) 160–165.
[27] M. Winslett, Reasoning about action using a possible models approach, in: *Proceedings AAAI-88*, St. Paul, MN (1988) 89–93.