Preferred History Semantics for Iterated Updates *

Shai Berger† Daniel Lehmann‡ Karl Schlechta§

Shai's corrections 24/2/99

Abstract

We give a semantics to iterated update by a preference relation on possible developments. An iterated update is a sequence of formulas, giving (incomplete) information about successive states of the world. A development is a sequence of models, describing a possible trajectory through time. We assume a principle of inertia and prefer those developments, which are compatible with the information, and avoid unnecessary changes. The logical properties of the updates defined in this way are considered, and a representation result is proved.

1 Introduction

1.1 Overview

We develop in this article an approach to update based on an abstract distance or ranking function. An agent has (incomplete, but reliable) information (observations) about a changing situation in the form of a sequence of formulas. At time 1, α_1 holds, at time 2, α_2 holds, ..., at time n, α_n holds. We are thus in a situation of iterated update. The agent tries to reason about the most likely outcome, i.e. to sharpen the information α_n by plausible reasoning. He knows that the real world has taken some trajectory, or history, that can be described by a sequence of models $< m_1, \ldots, m_n >$, where $m_i \models \alpha_i$ (remember the observations were supposed to be reliable). We say that such a history explains the observations. For his reasoning, he makes two assumptions: First, an assumption of inertia: histories that stay constant are more likely than histories that change without

*This work was partially supported by the Jean and Helene Alfassa fund for research in Artificial Intelligence and by grant 136/94-1 of the Israel Science Foundation on "New Perspectives on Nonmonotonic Reasoning".

†Platonix Technologies Ltd., 44 Petach Tikva Road, Tel Aviv 66183, Israel, shai@platonix.com

‡Institute of Computer Science, Hebrew University, Jerusalem 91904, Israel, lehmann@cs.huji.ac.il

§Laboratoire d'Informatique de Marseille, ESA CNRS 6077, Universite de Provence, CMI, 39 rue Joliot-Curie, F-13453 Marseille Cedex 13, France, ks@gyptis.univ-mrs.fr

necessity. For instance, if $n = 2$, and α_1 is consistent with α_2, $m_1 \models \alpha_1 \wedge \alpha_2$, $m_2 \models \alpha_2$, then the history $\langle m_1, m_1 \rangle$ is preferred to the history $\langle m_1, m_2 \rangle$. We do NOT assume that $\langle m_1, m_1 \rangle$ is more likely than some $\langle m_3, m_2 \rangle$, i.e. we do not compare the cardinality of changes, we only assume "sub-histories" to be more likely than longer ones. Second, the agent assumes that histories can be ranked by their likelihood, i.e. that there is an (abstract) scale, a total order, which describes this ranking. These assumptions are formalized in Section 1.4.

The agent then considers those models of α_n as most plausible, which are endpoints of preferred histories explaining the observations. Thus, his reasoning defines an operator $[\,]$ from the set of sequences of observations to the set of formulas of the underlying language, s.t. $[\alpha_1, \ldots, \alpha_n] \models \alpha_n$.

The purpose of this article is to characterize the operators $[\,]$ that correspond to such reasoning. Thus, we will give conditions for the operator $[\,]$ which all operators based on history ranking satisfy, and, conversely, which allow to construct a ranking r from the operator $[\,]$ such that the operator $[\,]_r$ based on this ranking is exactly $[\,]$. The first part can be called the soundness, the second the completeness part.

Before giving a complete set of conditions in Section 3, we discuss in Section 2 some logical and intuitive properties of ranking based operators, in particular those properties which are related to the Alchourrón, Gärdenfors, Makinson postulates for theory revision, or to the update postulates of Katsuno, Mendelzon. In Section 3, we give a full characterization of these operators. We start from a result of Lehmann, Magidor, Schlechta [7] on distance based revision, which has some formal similarity, and refine the techniques developed there.

In the rest of this section, we first compare briefly revision and update, and then emphasize the relevance of epistemic states for iterated update, i.e. that belief sets are in general insufficient to determine the outcome of our reasoning about iterated update. We then make our approach precise, give some basic definitions, and recall the AGM and KM postulates.

1.2 Revision and Update

Intuitively, belief revision (also called theory revision) in the sense of AGM deals with evolving knowledge about a static situation. Update, on the other hand, deals with knowledge about an evolving situation.

In the case of belief revision, an agent receives successively different information about a situation, e.g. from different sources, and the union of this information may be inconsistent. The theory of belief revision describes "rational" ways to incorporate new information into a body of old information, especially when the new and old information together are inconsistent.

In the case of update, an agent is informed that at time t, a formula ϕ held, at time t' ϕ', etc. The agent tries, given this information and some background assumptions (e.g. of

inertia: that things do not change unless forced to do so) to draw plausible conclusions about the probable development of the situation.
This distinction goes back to Katsuno and Mendelzon [5].
There is another distinction between belief revision and update, which is best seen on the semantic level. Revising a theory K with ϕ results in the theory defined by those models of ϕ in some sense closest to the set (as a whole) of the K-models. Updating K by ϕ results in those models of ϕ, which are in some sense closest to some individual K-model. Revision in this sense is formalized in Lehmann, Magidor and Schlechta [9], elaborated in [7]. The authors have devised there a family of semantics for revision based on minimal change, where change is measured by distances between models of formulae of the background logic. More precisely, the operator defined by such functions revises a belief set T according to a new observation α by picking the models of α that are closest to models of T. Such revision operators have been shown to satisfy some of the common rationality postulates. Several weak forms of a distance function ("pseudo-distances") have been studied in [7], and representation theorems for abstract revision operators by pseudo-distances have been proved. Since, in the weak forms, none of the notions usually connected with a distance (e.g. symmetry, the triangle inequality) are used, such pseudo-distances are actually no more than a preference function, or a ranked order, over pairs of models.
In the present article, we consider a setting that intuitively has an ontology of update. It follows the spirit of Katsuno and Mendelzon's [5] by setting a single belief change system for all sequences of observations (where the AGM revision operator is in general dependent upon the belief set being revised). Since it is based on possible development of the world, and not the knowledge base, all pseudo-distances are between individual models as in the KM update semantics, and not between a theory and a model as in the revision semantics.
On the other hand, where update semantics takes a "local" approach and incorporates in the new belief set the best updating models for each model in the old belief set, we take a "global" approach and pick for the updated belief set only the ending models of the best histories in all. As a result, our system validates all the AGM postulates, and not all the KM ones. The introduction of an update system which follows the AGM postulates for revision may have some interesting ontological consequences, but these will not be dealt with in this work.
The formal definitions and assumptions are to be found in Definitions 1.1, 1.2, 1.3, and Assumption 1.1. We prove a representation theorem for update operators based on rankings of histories, similar to those in [7].
Note also that a consequence of this semantic distinction between revision and update is that revision is close to preferential models, which essentially measure the distance from a fixed ideal point, and that update is close to counterfactual conditionals and their Lewis-Stalnaker semantics of closest worlds, which measures distances from individual worlds.

Finally, note that the approach taken here is more specific than that of [10], which considers arbitrary (e.g. not necessarily ranked) preference relations between histories.

1.3 Epistemic States are not Belief Sets

The epistemic state of an agent, i.e., the state of its mind, is understood to include anything that influences its actions, its beliefs about what is true of the world, and the way it will update or revise those beliefs, depending on the information it gathers. The belief set of an agent, at any time, includes only the set of propositions it believes to be true about the world at this time. One of the components of epistemic states must therefore be the belief set of the agent. One of the basic assumptions of the AGM theory of belief revisions is that epistemic states are belief sets, i.e., they do not include any other information. At least, AGM do not formalize in their basic theory, as expressed by the AGM postulates, any incorporation of other information in the belief revision process. In particular an agent that holds exactly the same beliefs about the state of the world, at two different instants in time is, at those times, in the same epistemic state and therefore, if faced with the same information, will revise its beliefs in the same way. Recent work on belief revision and update has shown this assumption has very powerful consequences, not always welcome [6, 1, 2]. Earlier work on belief base revision (see e.g. [8]) expresses a similar concern about the fundamentals of belief revision.
We do not wish to take a stand on the question of whether this identification of epistemic states with belief sets is reasonable for the study of belief revision, but we want to point out that, in the study of belief update, with its natural sensitivity (by the principle of inertia) to the order in which the information is gathered, it is certainly unreasonable. This is illustrated by the following observation: Let ϕ and ψ be different atomic, i.e. logically independent, formulas. First scenario: update the trivial belief set (the set of all tautologies) by ϕ, then by ψ, then by $\neg\phi \vee \neg\psi$. Second scenario: update the trivial belief set by ψ, then by ϕ, then by $\neg\phi \vee \neg\psi$. We expect different belief sets: we shall most probably try to stick to the piece of information that is the most up-to-date, i.e., ψ in the first scenario and ϕ in the second scenario. But there is no reason for us to think that the belief sets obtained, in both scenarios, just before the last updates, should be different. We expect them to be identical: the consequences of $\phi \wedge \psi$. The same agent, in two different epistemic states, updates differently the same beliefs in the light of the same information.

1.4 Preferred History Semantics

We now make our approach more precise. The basic ontology is minimal: The agent makes a sequence of observations and interprets this sequence of observations in terms of possible histories of the world explaining the observations.
Assume a set $\mathcal{L}$ of formulas and a set $\mathcal{U}$ of models for $\mathcal{L}$. Formulas will be denoted by

Greek letters from the beginning of the alphabet: α, β and so on, and models by m, n, and so on. We do not assume formulas are indexed by time. An observation is a *consistent* formula. (We assume observations to be consistent for two reasons: First, observations are assumed to be reliable; second, as we work with histories made of models, we need some model to explain every observation. Working with unreliable information would be the subject of another paper.) Observing a formula means observing the formula holds.
A sequence of observations is here a *finite* sequence of observations. Sequences of observations will be denoted by Greek letters from the end of the alphabet: σ, τ and concatenation of such sequences by $\cdot$. Notice the empty sequence is a legal sequence of observations. We shall identify an observation with the sequence of observations of length one that contains it. What does a sequence of observations tell us about the present state of the world?
A history is a finite, non-empty sequence of models. Histories will be denoted by h, f, and so on.

Definition 1.1 *A history* $h = \langle m_0, \ldots, m_n \rangle$ explains *a sequence of observations* $\tau = \langle \alpha_0, \ldots, \alpha_k \rangle$ *iff there are subscripts* $0 \leq i_0 \leq i_1 \leq \ldots \leq i_k \leq n$ *such that for any* j, $0 \leq j \leq k$, $m_{i_j} \models \alpha_j$.

Thus, a history explains a sequence of observations if there is, in the history, a model that explains each of the observations in the correct order. Notice that n is in general different from k, that many consecutive i_j's may be equal, i.e., the same model may explain many consecutive observations, that some models of the history may not be used at all in the explanation, i.e., l, $0 \leq l \leq n$ may be equal to none of the i_j's, and that we do not require that j_k be equal to n, or that j_0 be equal to 0, i.e., there may be useless models even at the start or at the end of a history. Note also that if h explains a sequence σ of observations, it also explains any subsequence (not necessarily contiguous) of σ.
The set of histories that explain a sequence of observations give us information about the probable outcome. Monotonic logic is useless here: if we consider all histories explaining a sequence of observations, we cannot conclude anything. It is reasonable, therefore to assume the agent restricts the set of histories it considers to a subset of the explaining histories.
We shall assume the agent has some preferences among histories, some histories being more natural, simpler, more expected, than others. A sequence of observations defines thus a subset of the set of all histories that explain it: the set of all preferred histories that explain it. This set defines the set of beliefs that result from a sequence of observations: the set of formulas satisfied in all the models that may appear as last elements of a preferred history that explains the sequence. The beliefs held depend on the preferences, concerning histories, of the agent. The logical properties of update depend on the class of preferences we shall consider.
Formally, one assumes the agent's preferences are represented by a binary relation $<$ on histories. Intuitively, $h < f$ means that history h is strictly preferred, e.g. strictly more

natural or strictly simpler, than history f. Note that our relation is on histories, not on models. We may now define preferred histories.

Definition 1.2 *A history h is a preferred history for a sequence σ of observations iff*

- *h explains σ*
- *there is no history h' that explains σ such that $h' < h$.*

In this work two assumptions are made concerning the preference relation $<$. First, we assume $<$ is a strict modular, well-ordering, i.e., $<$ is irreflexive, transitive, if $h < h'$, then for any f, either $h < f$ or $f < h'$ and there is no infinite descending chain. Secondly, we assume that partial histories (i.e., sub-histories) are preferred over (longer) histories:

Assumption 1.1 *If $h = \langle m_0, \ldots, m_n \rangle$ and $h' = \langle m_{j_0}, \ldots, m_{j_k} \rangle$, for $0 \leq j_0 < j_1 < \ldots < j_k \leq n$ and $0 \leq k < n$, then $h' < h$.*

For instance, $h' = \langle m_2, m_4 \rangle$ is preferred to $h = \langle m_1, m_2, m_3, m_4 \rangle$.
This assumption is justified both by an epistemological concern: simpler explanations are better, and by an assumption of inertia concerning the way the universe evolves: things tend to stay as they are. This assumption, in a finite setting, trivializes the well-ordering assumption.
Finally, we formally define our operator [ ]:

Definition 1.3 *After a sequence σ of observations, the agent holds the beliefs $[\sigma]$, defined by $\alpha \in [\sigma]$ iff for every model m and every history h, if h is a preferred history explaining σ and m is the last element of h, then $m \models \alpha$.*

Notice that, since histories are non-empty, this definition always makes sense.
The remainder of this paper will show that the assumptions above have far reaching consequences: they are strong assumptions. Our purpose is indeed to look for a powerful logic, not for the minimal logic that agrees with any possible ontology.

1.5 Basic Definitions and Notation

We are dealing here only with finite and complete universes, so theories have logically equivalent formulas (and vice versa), and are isomorphic to sets of models, and we will use them in these senses interchangeably. We will see sequence concatenation also as an outer product (with respect to concatenation) of sets of histories. For technical reasons, most of the discussion will relate to sets of models. An exception, for easier readability and comparison with the AGM and KM conditions is Section 2. We do not consistently differentiate singletons from the members that comprise them, because it is always clear from context which of them we are referring to.

A theory, or belief set, will be a deductively closed set of formulas.
For the convenience of the reader, we recall the AGM postulates for belief revision, (see e.g. [3]), and the Katsuno-Mendelzon postulates for update (see e.g. [5]):

$(K*1)$ $K*\alpha$ is a deductively closed set of formulas.
$(K*2)$ $\alpha \in K*\alpha$.
$(K*3)$ $K*\alpha \subseteq Cn(K,\alpha)$.
$(K*4)$ If $\neg\alpha \notin K$, then $Cn(K,\alpha) \subseteq K*\alpha$.
$(K*5)$ If $K*\alpha$ is inconsistent then α is a logical contradiction.
$(K*6)$ If $\models \alpha \leftrightarrow \beta$, then $K*\alpha = K*\beta$.
$(K*7)$ $K*\alpha\wedge\beta \subseteq Cn(K*\alpha,\beta)$.
$(K*8)$ If $\neg\beta \notin K*\alpha$, then $Cn(K*\alpha,\beta) \subseteq K*\alpha\wedge\beta$.

(U1) $\models (\psi\cdot\mu) \rightarrow \mu$.
(U2) If $\models \psi \rightarrow \mu$, then $\models (\psi\cdot\mu) \leftrightarrow \psi$.
(U3) If both ψ and μ are satisfiable, then so is $\psi\cdot\mu$.
(U4) If $\models \psi_1 \leftrightarrow \psi_2$ and $\models \mu_1 \leftrightarrow \mu_2$, then $\models (\psi_1\cdot\mu_1) \leftrightarrow (\psi_2\cdot\mu_2)$.
(U5) $\models ((\psi\cdot\mu)\wedge\phi) \rightarrow (\psi\cdot(\mu\wedge\phi))$.
(U6) If $\models (\psi\cdot\mu_1) \rightarrow \mu_2$ and $\models (\psi\cdot\mu_2) \rightarrow \mu_1$, then $\models (\psi\cdot\mu_1) \leftrightarrow (\psi\cdot\mu_2)$.
(U7) If ψ is complete, then $\models (\psi\cdot\mu_1)\wedge(\psi\cdot\mu_2) \rightarrow (\psi\cdot(\mu_1\vee\mu_2))$.
(U8) $\models ((\psi_1\vee\psi_2)\cdot\mu) \leftrightarrow (\psi_1\cdot\mu)\vee(\psi_2\cdot\mu)$.

The following is a slight reformulation of the AGM postulates in the spirit of Katsuno-Mendelzon, taken from [7]. We consider here a symmetrical version, in the sense that K and α can both be theories, and simplify by considering only consistent theories.

$(*0)$ If $\models T \leftrightarrow S$, $\models T' \leftrightarrow S'$, then $T*T' = S*S'$,
$(*1)$ $T*T'$ is a consistent, deductively closed theory,
$(*2)$ $T' \subseteq T*T'$,
$(*3)$ If $T\cup T'$ is consistent, then $T*T' = Cn(T\cup T')$,
$(*4)$ If $T*T'$ is consistent with T'', then $T*(T'\cup T'') = Cn((T*T')\cup T'')$.

Finally, we recall the definition of a pseudo-distance from [7].

Definition 1.4
$d : U\times U \rightarrow Z$ is called a pseudo-distance on U iff Z is totally ordered by a relation $<$.

2 Some Important Logical Properties of Updates

A number of logical properties of the operator $[\ ]$ will now be described and discussed. The reasons why those properties hold are varied: some depend on very little of our assumptions, some on almost all of them. We shall try to make the appropriate distinctions.

Lemma 2.1 *For any σ, $[\sigma]$ is a theory.*

This property is analogous to AGM's $(K * 1)$ and is implicit in Katsuno-Mendelzon's presentation [5]. This depends only on the fact that Definition 1.3 defines $[\sigma]$ as the set of all formulas that hold for all the models in a given set. Indeed, this is a property that is expected to hold by the structure of belief sets, not by the definition of explanation or certain properties of the preference relation.
The following properties hold by the definition of explanation, i.e., Definition 1.1.

Lemma 2.2 *If α and α' are logically equivalent, then for any sequences σ, τ: $[\sigma \cdot \alpha \cdot \tau] = [\sigma \cdot \alpha' \cdot \tau]$.*

This property is analogous to AGM's $(K * 6)$ but notice that, there, it is needed only for the second argument of the revision operation, since it is implicit for the first, a theory. It parallels (U4) in Katsuno-Mendelzon's [5]. Lemma 2.2 follows from Definition 1.1, that implies that the histories that explain $\sigma \cdot \alpha \cdot \tau$ are exactly those that explain $\sigma \cdot \alpha' \cdot \tau$. The preferred histories are therefore the same. The next property is more original.

Lemma 2.3 *If $\beta \models \alpha$, then for any sequences σ, τ:*

$$[\sigma \cdot \alpha \cdot \beta \cdot \tau] = [\sigma \cdot \beta \cdot \tau] = [\sigma \cdot \beta \cdot \alpha \cdot \tau].$$

This property has no clear analogue in the AGM or KM frameworks, but is closely related to (U2) of [5]. The first equation is property (C1) of Darwiche-Pearl's [1] and (I5) of [6]. The second equation is a weakening of (I4) of [6]. Here we request α to be a logical consequence of β, there we only asked that α be in $[\sigma \cdot \beta]$. Lemma 2.3 is a consequence of the fact that the histories that explain $\sigma \cdot \beta \cdot \tau$, $\sigma \cdot \alpha \cdot \beta \cdot \tau$ and $\sigma \cdot \beta \cdot \alpha \cdot \tau$ are the same.

Corollary 2.1 *For any sequence σ, $[\sigma \cdot \mathbf{true}] = [\sigma]$.*

The next property deals with disjunction.

Lemma 2.4 *If γ is a member both of $[\sigma \cdot \alpha \cdot \tau]$ and $[\sigma \cdot \beta \cdot \tau]$, then it is a member of $[\sigma \cdot \alpha \vee \beta \cdot \tau]$. In other words*

$$[\sigma \cdot \alpha \cdot \tau] \cap [\sigma \cdot \beta \cdot \tau] \subseteq [\sigma \cdot \alpha \vee \beta \cdot \tau].$$

This property is similar to one half of (U8) of [5], and to a consequence of AGM's (K*7), as pointed out in [4], property (3.14): $(K * A) \cap (K * B) \subseteq K * (A \vee B)$.
Lemma 2.4 depends only on Definitions 1.1 and 1.2, but does not depend on any properties of the preference relation. A history h that explains $\sigma \cdot \alpha \vee \beta \cdot \tau$ explains $\sigma \cdot \alpha \cdot \tau$ or $\sigma \cdot \beta \cdot \tau$. A preferred history for $\sigma \cdot \alpha \vee \beta \cdot \tau$ must therefore either be a preferred history for $\sigma \cdot \alpha \cdot \tau$ (since any history explaining the latter explains the former) or preferred history for $\sigma \cdot \beta \cdot \tau$. The next lemma is a strengthening of Lemma 2.4, and it depends on the modularity of the preference relation.

Lemma 2.5 *The theory $[\sigma \cdot \alpha \vee \beta \cdot \tau]$ is equal to $[\sigma \cdot \alpha \cdot \tau]$, equal to $[\sigma \cdot \beta \cdot \tau]$ or is the intersection of the two theories above.*

This property is a weakening of (U8) of [5], compare also to property (3.16) in [4]: $K * (A \vee B) = K * A$ or $K * (A \vee B) = K * B$ or $K * (A \vee B) = (K * A) \cap (K * B)$.

Proof: A history h that explains $\sigma \cdot \alpha \vee \beta \cdot \tau$ explains $\sigma \cdot \alpha \cdot \tau$ or $\sigma \cdot \beta \cdot \tau$. If all preferred histories for $\sigma \cdot \alpha \vee \beta \cdot \tau$ explain $\sigma \cdot \alpha \cdot \tau$, then

- any preferred history for $\sigma \cdot \alpha \vee \beta \cdot \tau$ is a preferred history for $\sigma \cdot \alpha \cdot \tau$ (otherwise there would be a strictly preferred history for $\sigma \cdot \alpha \cdot \tau$, but that explains $\sigma \cdot \alpha \vee \beta \cdot \tau$), and
- any preferred history for $\sigma \cdot \alpha \cdot \tau$ is a preferred history for $\sigma \cdot \alpha \vee \beta \cdot \tau$, otherwise there would be a strictly preferred history for $\sigma \cdot \alpha \vee \beta \cdot \tau$ that satisfies $\sigma \cdot \beta \cdot \tau$.

In this case, $[\sigma \cdot \alpha \vee \beta \cdot \tau]$ is equal to $[\sigma \cdot \alpha \cdot \tau]$.
Similarly, if all preferred histories for $\sigma \cdot \alpha \vee \beta \cdot \tau$ explain $\sigma \cdot \beta \cdot \tau$, then $[\sigma \cdot \alpha \vee \beta \cdot \tau]$ is equal to $[\sigma \cdot \beta \cdot \tau]$.
Let us assume, therefore, that some preferred histories for $\sigma \cdot \alpha \vee \beta \cdot \tau$ explain $\sigma \cdot \alpha \cdot \tau$ and that some explain $\sigma \cdot \beta \cdot \tau$. By modularity of the preference relation, any preferred history for $\sigma \cdot \alpha \cdot \tau$ is a preferred history for $\sigma \cdot \alpha \vee \beta \cdot \tau$. ∎

The next properties follow from the assumption that sub-histories are preferred to more complete histories. Remark first that, if a history h explains a non-empty sequence σ of observations but the last model of h does not satisfy the last observation of σ, then there is a shorter (initial) sub-history of h that explains σ. The history h cannot be, in this case, a preferred history for σ.

Lemma 2.6 *For any sequence σ of observations and any formula α: $\alpha \in [\sigma \cdot \alpha]$.*

This property is similar to AGM's $(K * 2)$ and (U1) of [5]. In [2], Friedman and Halpern question this postulate. Here, it finds a justification, grounded in our preference for shorter explanations.

Lemma 2.7 *For any sequence σ of observations and any formulas α and β, if $\neg\beta \notin [\sigma \cdot \alpha]$, then $[\sigma \cdot \alpha \cdot \beta] = [\sigma \cdot \alpha \wedge \beta] = \mathcal{C}n([\sigma \cdot \alpha], \beta)$.*

This property is analogous to AGM's $(K * 7)$ and $(K * 8)$.
Proof: We show that the preferred histories for $\sigma \cdot \alpha \cdot \beta$ are exactly those preferred histories of $\sigma \cdot \alpha$ whose last element satisfies β. First, clearly, any preferred history for $\sigma \cdot \alpha$ whose last element satisfies β explains $\sigma \cdot \alpha \cdot \beta$ and is a preferred history for it. Secondly, since $\neg\beta \notin [\sigma \cdot \alpha]$, there is a preferred history h for $\sigma \cdot \alpha$ whose last element satisfies β. As we have just seen h is a preferred history for $\sigma \cdot \alpha \cdot \beta$. Let f be a preferred history for $\sigma \cdot \alpha \cdot \beta$. It explains $\sigma \cdot \alpha$. If it were not a preferred history for $\sigma \cdot \alpha$, there would be a history f', $f' < f$ that explains $\sigma \cdot \alpha$. By modularity, we would have $f' < h$ or $h < f$, which are both impossible. We conclude that f is a preferred history for $\sigma \cdot \alpha$. But its last element satisfies β, by Lemma 2.6.
We have shown that $[\sigma \cdot \alpha \cdot \beta] = \mathcal{C}n([\sigma \cdot \alpha], \beta)$.
To conclude the proof, notice that, by the above, any preferred history for $\sigma \cdot \alpha \cdot \beta$ explains $\sigma \cdot \alpha \wedge \beta$ and that any history explaining $\sigma \cdot \alpha \wedge \beta$ also explains $\sigma \cdot \alpha \cdot \beta$. ∎

Corollary 2.2 *If* $\neg\alpha \notin [\sigma]$*, then* $[\sigma \cdot \alpha] = \mathcal{C}n([\sigma], \alpha)$.

This parallels AGM's $(K * 3)$ and $(K * 4)$.
Proof: By Corollary 2.1, $[\sigma] = [\sigma \cdot \mathbf{true}]$. By Lemma 2.7,

$$[\sigma \cdot \mathbf{true} \cdot \alpha] = \mathcal{C}n([\sigma \cdot \mathbf{true}], \alpha) = [\sigma \cdot \mathbf{true} \wedge \alpha].$$

By Corollary 2.1 and Lemma 2.3 $\mathcal{C}n([\sigma], \alpha) = [\sigma \cdot \alpha]$. ▮
Our last property depends on well-foundedness, which is trivial in a finite setting.

Lemma 2.8 *For any sequence* σ*,* $[\sigma]$ *is consistent.*

This parallels AGM's $(K*5)$ and KM's (U3). This property depends on two assumptions. First we assumed observations were consistent formulas. It follows that any sequence of observations is explained by some history. By finiteness, if h explains σ and h is not a preferred history for σ, then there is a preferred history h' for σ (in fact one such that $h' < h$). Therefore $[\sigma]$ is consistent.

3 A Representation Theorem

3.1 Introduction

In Section 2, we have presented several logical properties of operators [ ] based on history ranking. In this Section, we will give a full characterization. We generalize here results about revision reported in [7]. We will first show that a straightforward generalization of the not necessarily symmetrical revision case fails already for sequences of length 3, and will then give a characterization for sequences of arbitrary finite length in Theorem 3.2. A technical problem for the latter is that we have to work with "illegal" sets of histories, which do not correspond to sequences of sets of models. E.g. the set of sequences $\{\langle 0,0,0\rangle, \langle 0,1,1\rangle\}$ is not the product of any sequence of sets - $\{0\} \times \{0,1\} \times \{0,1\}$ contains too many sequences. Our operator is, however, only defined for such sequences of sets. We use the idea of a patch, a cover of such sets of sequences by products of sequences of sets, to show our result.
As mentioned before, we will want to work mainly with sequences of sets of models. Such sets are freely interchangeable with observations, and every sequence thereof also defines a set of histories. We would like the set of all explaining histories to be representable as a sequence of sets of models. Definition 1.1 allows the set of explaining histories to be infinite, and we will limit the sets of histories dealt with by an assumption strengthening Lemma 2.6. While we do not assume full sub-history preference, we assume that a history explains a sequence of observations if they are of the same length, and the ith model in the history models the ith observation in the sequence. This assumption, like the Lemma, is justified by sub-history preference, with an implicit agreement that consecutive repeats

of a model in a history are merely another way to write that the model explains several observations. In other words, to make the formal phrasing and proof a little easier, we will write for each observation in the sequence σ the model that explains it in the history h. A history h' containing h as a sub-history will not be preferable to h, and sub-histories of h are just represented as longer than they are. Intuitively, an assumption is made here that e.g. $\langle m_1, m_1, m_2\rangle$ and $\langle m_1, m_2, m_2\rangle$, both being representations of $\langle m_1, m_2\rangle$, are equally preferred, and are both considered better than $\langle m_1, m_2, m_3\rangle$. This assumption is neither used nor needed in the theorem or its proof. We make no further assumption on the history-preference relation.
Histories and sequences of observations of length or dimension 2 are closely parallel to the not necessarily symmetric case of [7]. We first recall the corresponding representation result in Section 3.2. Then, we show that a simple generalization of this result fails already in the case of length 3 (Section 3.4). Finally, in Section 3.5, we prove a valid representation theorem for the general, n-dimensional case.

3.2 The 2-D Representation Theorem

First, let us quote a theorem characterizing the revision operators representable by a pseudo-distance function (Proposition 2.5 of [7]). The pseudo-distance mentioned here is actually no more than a preference relation over pairs of models (or histories of length 2). The theorem deals with an operator $\mid$ which revises a belief set by an observation, i.e., $A \mid B$ is the belief set held by an agent who has held a belief set A, after observing B. Now, let X be a finite and complete universe (the set of possible models of the language). In such a universe, A and B are interchangeably formulas, theories and sets of models. Let $\mathcal{P}(X)$ designate the set of all *non-empty* subsets of X.

Definition 3.1 *An operation* $\mid : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathcal{P}(X)$ *is* representable *iff there is a pseudo-distance* $d : X \times X \to \mathcal{Z}$ *such that*

$$A \mid B = \{b \in B \mid \exists a \in A \text{ such that } \forall a' \in A, b' \in B, d(a,b) \leq d(a',b')\}.$$

Thus, intuitively, if A and B are sets, $A \mid B$ is the set of those elements of B, which are closest to the set A. By abuse of notation, if A and B are formulas, $A \mid B$ is the set of formulas valid in the set of those models of B, which are closest to the set of models of A.
For this theorem, Lehmann, Magidor and Schlechta define a relation $R_{\mid}$ on pairs from $\mathcal{P}(X) \times \mathcal{P}(X)$, which intuitively means "provably closer" or "provably preferable", i.e., assuming the underlying pseudo-distance exists, this relation represents information about it that may be deduced by examining the revision operation. For instance, if $(A \mid (B \cup C)) \cap B \neq \emptyset$, B is provably at least as close to A, as C is to A. This information can only apply to the best-preferred pairs of models in the pairs of sets, so $(A, B) R_{\mid} (A', B')$ actually means we have evidence that the best pair of models in $A \times B$ is at least as preferable as the best pair in $A' \times B'$.

Definition 3.2 *Given an operation* $\mid$ *, define a relation* $R_{\mid}$ *on pairs from* $\mathcal{P}(X) \times \mathcal{P}(X)$ *by:* $(A, B) R_{\mid} (A', B')$ *iff one of the following two cases obtains:*

1. $A = A'$ *and* $(A \mid (B \cup B')) \cap B \neq \emptyset$*,*
2. $B = B'$ *and* $((A \cup A') \mid B) \neq (A' \mid B)$*.*

In the rest of this subsection we shall write R instead of $R_{\mid}$. As usual, we shall denote by $R^{\star}$ the transitive closure of R.
Now, we can quote the representation theorem:

Theorem 3.1 *An operation* $\mid$ *is representable iff it satisfies the four conditions below for any non-empty sets* $A, A', B, B' \subseteq X$*:*

1. $(A \mid B) \subseteq B$*,*
2. $((A \cup A') \mid B) \subseteq (A \mid B) \cup (A' \mid B)$*,*
3. *If* $(A, B) R^{\star} (A, B')$*, then* $(A \mid B) \subseteq (A \mid (B \cup B'))$*,*
4. *If* $(A, B) R^{\star} (A', B)$*, then* $(A \mid B) \subseteq ((A \cup A') \mid B)$

This is the strongest version of this theorem proven. Fixing the conditions of the theorem, the characterization grows stronger as the definition of R becomes narrower, as it then (possibly) puts less constraints on $\mid$. A weaker characterization (which is also valid) has R defined to be wider, as follows:

Definition 3.3 *We say the relation* $R_{\mid}$ *holds iff at least one of the following cases obtains:*

1. $A \supseteq A', B \supseteq B' \Rightarrow (A, B) R (A', B')$
2. $(A \mid (B \cup B')) \cap B \neq \emptyset \Rightarrow (A, B) R (A, (B \cup B'))$
3. $((A \cup A') \mid B) \neq (A' \mid B) \Rightarrow (A, B) R ((A \cup A'), B)$

3.3 Ultimate Goal: The n-Dimensional Case

We want to prove a theorem analogous to Theorem 3.1 that relates to strings of observations of length n (instead of length 2, if we ignore the difference in role between a previous observation and a previous belief set), that is, we want to characterize representable operations $[\,]\ : \mathcal{P}(X)^n \to \mathcal{P}(X)$:

Definition 3.4 *An operation* $[\,] : \mathcal{P}(X)^n \to \mathcal{P}(X)$ *is* representable *iff there is a totally ordered set* $\mathcal{Z}$ *(the order is* $<$*) and a function* $r : X^n \to \mathcal{Z}$ *(that will be intuitively understood as a history ranking), such that, for any non-empty subsets* $A_1, ...A_n \subseteq X$*,*

$$\begin{aligned} &[A_1 \cdots A_n] = \\ &\{a_n \in A_n \mid \exists a_1 \in A_1 ... a_{n-1} \in A_{n-1} \forall a'_1 \in A_1 ... a'_n \in A_n \; r(a_1, ..., a_n) \leq r(a'_1, ..., a'_n)\} \end{aligned} \tag{1}$$

We will now see that this may not be achieved by straightforward generalization of the tight 2-dimensional characterization, even for just three dimensions.

3.4 Simple Generalization is not Valid

The first attempt at generalizing this theorem is held short at $n = 3$. Let us phrase the suggested theorem and disprove it, starting with a new definition for R:

Definition 3.5 *Given an operation* $[\,]$ *, one defines a relation* $R_{[\,]}$ *on triplets of non-empty subsets of* X *by:* $(A, B, C)R_{[\,]}(A', B', C')$ *iff one of the following cases obtains:*

1. $A = A'$, $B = B'$ *and* $[A \cdot B \cdot (C \cup C')] \cap C \neq \emptyset$.
2. $A = A'$, $C = C'$ *and* $[A \cdot (B \cup B') \cdot C] \neq [A \cdot B' \cdot C]$.
3. $B = B'$, $C = C'$ *and* $[(A \cup A') \cdot B \cdot C] \neq [A' \cdot B \cdot C]$.

¿From here on, unless otherwise stated, R stands for $R_{[\,]}$. Now, the suggested theorem:

Suggested Theorem 3.1 *An operation* $[\,]$ *is representable iff it satisfies the six conditions below for any non-empty sets* $A, A', B, B', C, C' \subseteq X$*:*

1. $[A \cdot B \cdot C] \subseteq C$.
2. $[A \cdot (B \cup B') \cdot C] \subseteq [A \cdot B \cdot C] \cup [A \cdot B' \cdot C]$.
3. $[(A \cup A') \cdot B \cdot C] \subseteq [A \cdot B \cdot C] \cup [A' \cdot B \cdot C]$.
4. $(A, B, C)R^*(A', B, C) \Rightarrow [A \cdot B \cdot C] \subseteq [(A \cup A') \cdot B \cdot C]$.
5. $(A, B, C)R^*(A, B', C) \Rightarrow [A \cdot B \cdot C] \subseteq [A \cdot (B \cup B') \cdot C]$.
6. $(A, B, C)R^*(A, B, C') \Rightarrow [A \cdot B \cdot C] \subseteq [A \cdot B \cdot (C \cup C')]$.

This theorem is not valid.

Proof: There is a counter-example, as follows: Let $X = \{0, 1\}$ and $n = 3$. This seems to be the simplest n-dimensional case for $n > 2$. For convenience, we will write 0 for $\{0\}$, 1 for $\{1\}$, X for $\{0, 1\}$, and $\star$ for any of them. Define $[\,]$ as follows:

$$[\star \cdot \star \cdot 0] = 0\ ;\ [\star \cdot \star \cdot 1] = 1;$$

$$[0 \cdot 0 \cdot X] = 0\ ;\ [0 \cdot 1 \cdot X] = 0;$$

$$[1 \cdot 0 \cdot X] = 1\ ;\ [1 \cdot 1 \cdot X] = X;$$

$$[0 \cdot X \cdot X] = 0\ ;\ [1 \cdot X \cdot X] = 1;$$

$$[X \cdot 0 \cdot X] = 0\ ;\ [X \cdot 1 \cdot X] = X;$$

$$[X \cdot X \cdot X] = X.$$

The cases where $C = \{0\}$ and $C = \{1\}$ are forced by condition 1 of the theorem, and are not very interesting. As for the case $C = X$, one may check for each two triplets that fall under the conditions of the theorem that it holds. The check is simplified by the fact that Definition 3.5 gives no way to show that $(A, B, C)R(A', B', C)$ when $A \subseteq A'$ and $B \subseteq B'$, and the fact that there is no valid union of sets of sequences of different cardinalities, unless one is contained within the other. The operator complies with all the conditions of Suggested Theorem 3.1. But suppose there is a ranking that defines $[\,]$, we reach the following conclusions (using $x \preceq y$ for xRy, and $x \prec y$ for xRy but provably $\neg yR^*x$, and remembering that the conditions of the Suggested Theorem hold):

$$(1 \cdot 1 \cdot X) \preceq (0 \cdot 1 \cdot X)$$

by Definition 3.5, case 3,

$$(1 \cdot 0 \cdot X) \prec (1 \cdot 1 \cdot X)$$

by case 2 of Definition 3.5, and negation of condition 5 of the Suggested Theorem, and

$$(0 \cdot 0 \cdot X) \prec (1 \cdot 0 \cdot X)$$

by case 3 of the definition of R and negation of condition 4 of the Suggested Theorem. Thus, the best history in $\{0\} \cdot \{0\} \cdot X$ (which we know to be $0 \cdot 0 \cdot 0$) would also be the single best history in $X \cdot X \cdot X$, but we find $[X \cdot X \cdot X] = X$ instead, which means there really is no underlying ranking. ∎

The next natural step is to use a definition more like Definition 3.3 in the generalized theorem. If R was defined as follows:

Definition 3.6 *We say the relation $R_{[\,]}$ holds iff any of the following cases obtains:*

1. $A \supseteq A', B \supseteq B', C \supseteq C' \Rightarrow (A, B, C)R(A', B', C')$
2. $[A \cdot B \cdot (C \cup C')] \cap C \neq \emptyset \Rightarrow (A, B, C)R(A, B, C \cup C')$
3. $[A \cdot (B \cup B') \cdot C] \neq [A \cdot B' \cdot C] \Rightarrow (A, B, C)R(A, B \cup B', C)$
4. $[(A \cup A') \cdot B \cdot C] \neq [A' \cdot B \cdot C] \Rightarrow (A, B, C)R(A \cup A', B, C)$

Then, for the operator defined in the counter-example, we could show that

$$(1, 0, X)R(1, X, X)R(X, X, X)R(X, 0, X)$$

but

$$[1 \cdot 0 \cdot X] \not\subseteq [X \cdot 0 \cdot X]$$

so the operator violates condition 4, and is thus not a counter-example. We have not been able to prove a representation theorem using Definition 3.6, and believe it not to be valid. Yet, we have not found any counter-example. We have discovered (by computerized enumeration of the operators) that there are no counter-examples with $n = 3$ and $|X| = 2$, and the question whether it is valid for all n and $|X|$, and if not, for which n and $|X|$ it is valid, remains open at this stage.

3.5 An n-Dimensional Representation Theorem

3.5.1 Patches

We believe that for this theorem, we need a way to work around "illegal" sets of histories, i.e. such that are not sequences of sets of models. We call such sets "illegal" because, not being interchangeable with sequences of observations, our operator cannot be applied to them. For the technique used in this proof, it is especially regretful that for most sequences σ, if we try to present σ as a disjoint union of a "legal" sequence (say, a singleton) τ with the rest of the histories in the sequence $\sigma \setminus \tau$ then the latter will be illegal. Instead, we use a family of legal sequences whose union is $\sigma \setminus \tau$. As it complements τ to σ, we call this family a *patch*.

Definition 3.7 *Given a sequence $A_1 \cdots A_n$ and a sequence $A'_1 \cdots A'_n$ such that $\forall i A'_i \subseteq A_i$, we define the* patch to $A_1 \cdots A_n$ from $A'_1 \cdots A'_n$ *to be the set of sequences comprised of all the sequences $B^i_1 \cdots B^i_n$ for each i such that $A'_i \neq A_i$, where*

1. $\forall j \neq i$, $B^i_j = A_j$
2. $B^i_i = A_i \setminus A'_i$

Let, for instance, $n = 3$, $A := \{0,1\} \cdot \{0,1\} \cdot \{0,1\}$, $A' := \{0\} \cdot \{0,1\} \cdot \{0\}$, then the patch to A from A' is $\{B^1 := \{1\} \cdot \{0,1\} \cdot \{0,1\}, B^3 := \{0,1\} \cdot \{0,1\} \cdot \{1\}\}$.
Note that the (disjoint) union of A' with $\bigcup_i B^i$ is equal to A.
As with the definition of $R_{|}$, this definition is made as tight as possible to make the theorem stronger. We will see later that if we relax the definition a little, allowing many patches from A' to A by letting B_i^i be any set such that $A_i' \cup B_i^i = A_i$, our theorem will become the generalization of the weaker form of Theorem 3.1 (the one using Definition 3.3). Notice also that Definition 3.7 may be easily written in the language of formulae; we do not wish our basic concepts, or the phrasing of the representation theorem, to stray too far from our original problem domain.

3.5.2 The Theorem

With the new tool, the patch, we may define a new, more powerful relation of "provable preferability".

Definition 3.8 *Given an operation* $[\,]$, *define a relation* $R_{[\,]}$ *on sequences of length* n *of non-empty subsets of* X *by:* $(A_1', ..., A_{n-1}', C')R_{[\,]}(A_1, ..., A_{n-1}, C)$ *iff one of the following cases obtains:*

1. $\forall i A_i \subseteq A_i'$ *and* $C \subseteq C'$.
2. $\forall i A_i = A_i'$, $C' \subseteq C$ *and* $[A_1 \cdots A_{n-1} \cdot C] \cap C' \neq \emptyset$.
3. $\forall i A_i \supseteq A_i'$, $C = C'$, $\{B_1^i \cdots B_{n-1}^i\}$ *is the patch to* $A_1 \cdots A_{n-1}$ *from* $A_1' \cdots A_{n-1}'$, *and one of the following holds:*
   (a) $\bigcap_i [B_1^i \cdots B_{n-1}^i \cdot C] \not\subseteq [A_1 \cdots A_{n-1} \cdot C]$
   (b) $[A_1 \cdots A_{n-1} \cdot C] \not\subseteq \bigcup_i [B_1^i \cdots B_{n-1}^i \cdot C]$

Now we are ready to phrase the representation theorem for the n-dimensional case:

Theorem 3.2 *An operation* $[\,]$ *is representable iff it satisfies the three conditions below for any non-empty sets* $A_1, ...A_{n-1}, C \subseteq X$:

1. $[A_1 \cdots A_{n-1} \cdot C] \subseteq C$,
2. *If* $\forall i A_i' \subseteq A_i$ *and* $\{B_1^i \cdots B_{n-1}^i\}$ *is the patch to* $A_1 \cdots A_{n-1}$ *from* $A_1' \cdots A_{n-1}'$ *then* $[A_1 \cdots A_{n-1} \cdot C] \subseteq [A_1' \cdots A_{n-1}' \cdot C] \cup \bigcup_i [B_1^i \cdots B_{n-1}^i \cdot C]$
3. *If* $\forall i A_i' \subseteq A_i$, $C' \subseteq C$ *and* $(A_1' \cdots A_{n-1}' \cdot C')R^\star(A_1 \cdots A_{n-1} \cdot C)$, *then* $[A_1' \cdots A_{n-1}' \cdot C'] \subseteq [A_1 \cdots A_{n-1} \cdot C]$.

Before proving this theorem, we note that when taking the more relaxed version (as mentioned above) of Definition 3.7, and taking Definition 3.3 to define $R_{|}$, Theorem 3.1 becomes the special case of Theorem 3.2 with $n = 2$: Under these definitions, for any sets A and A', A' (or rather, the set comprised of it as a single sequence of length 1) is a patch to $A \cup A'$ from A, so the concept of the patch coincides with union. When the patch has only one sequence (of one set) in it, then $\bigcap[B_1^1 \cdot C] = \bigcup[B_1^1 \cdot C] = [B_1^1 \cdot C]$, and so the cases 3a and 3b of Definition 3.8 together become case 3 of Definition 3.3. And finally, the conditions 3 and 4 of Theorem 3.1 are expressed together as condition 3 of Theorem 3.2. Note that condition 3 in Theorem 3.2 is essentially a loop condition. Let us now prove the more general theorem.
Proof: For the proof of Theorem 3.2, a number of lemmas will be needed. These lemmas will be presented when needed, and their proof inserted in the midst of the main proof. First, we shall deal with the *soundness* of the theorem, and then with the more challenging *completeness*.

Suppose, then, that $[\,]$ is representable. The ranking r of histories may be extended to sequences of sets in the usual way, by taking the minimum over the sets: $r(A_1, ..., A_n) = min_{a_i \in A_i}\{r(a_1, ..., a_n)\}$. One may then write the equation defining representability as

$$[A_1 \cdots A_{n-1} \cdot C] = \{c \in C \mid r(A_1, ..., A_{n-1}, \{c\}) = r(A_1, ..., A_{n-1}, C)\}$$

Let us now show that the three conditions of Theorem 3.2 hold:
Condition 1 is obvious. For Condition 2, notice that, on one hand, $A_1 \cdots A_{n-1} = \bigcup_i \{B_1^i \cdots B_{n-1}^i\} \cup (A'_1 \cdots A'_{n-1})$, and on the other hand, if $r(a_1, ..., a_{n-1}, c) = r(A_1, ..., A_{n-1}, C)$, $\langle a_1, \cdots, a_n, c\rangle \in A''_1 \cdots A''_{n-1} \cdot C''$ for some A''_i, C'', and $A''_1 \cdots A''_{n-1} \cdot C'' \subseteq A_1 \cdots A_{n-1} \cdot C$ then also $r(a_1, ..., a_{n-1}, c) = r(A''_1, ..., A''_{n-1}, C'')$ and $c \in [A''_1 \cdots A''_{n-1} \cdot C'']$. For Condition 3 we need a little lemma:

Lemma 3.1 *For any non-empty sets $A_i, A'_i, C, C' \subseteq X$,*

$$(A'_1, ..., A'_{n-1}, C')R(A_1, ..., A_{n-1}, C) \Rightarrow r(A'_1, ..., A'_{n-1}, C') \leq r(A_1, ..., A_{n-1}, C)$$

Proof: Let us consider the different cases of Definition 3.8:
Case 1 is obvious.
In case 2 the negation of the consequence $r(A_1, ..., A_{n-1}, C) < r(A'_1, ..., A'_{n-1}, C')$ implies $[A_1 \cdots A_{n-1} \cdot C] \cap C' = \emptyset$ which is the negation of the assumption.
Both sub-cases of case 3 rely on the reasoning used for Condition 2 above: In case 3a by assumption there exists c such that $c \in \bigcap_i [B_1^i \cdots B_{n-1}^i \cdot C]$ but $c \notin [A_1 \cdots A_{n-1} \cdot C]$. By the former, for each i there is a history $\langle b_1^i, \cdots, b_{n-1}^i, c\rangle$, $b_j^i \in B_j^i$, with $r(b_1^i, \cdots, b_{n-1}^i, c)$ minimal in $B_1^i \cdots B_{n-1}^i \cdot C$, but by the latter, for none of these histories $r(b_1^i, \cdots, b_{n-1}^i, c)$ is minimal in $A_1 \cdots A_{n-1} \cdot C$. So the histories with r minimal in $A_1 \cdots A_{n-1} \cdot C$ have to be all members of $A'_1 \cdots A'_{n-1} \cdot C$ which implies $r(A'_1, ..., A'_{n-1}, C) \leq r(A_1, ..., A_{n-1}, C)$ as

needed.
In case 3b, by assumption there is at least one history that is one of the best in $A_1 \cdots A_{n-1} \cdot C$, but is not one of the best in (and thus not a member of) $B_1^i \cdots B_{n-1}^i \cdot C$ for any i. Hence this history is in $A'_1 \cdots A'_{n-1} \cdot C$ and $r(A'_1, ..., A'_{n-1}, C) \leq r(A_1, ..., A_{n-1}, C)$ as needed. ∎

We conclude that Condition 3 holds and the characterization is sound.
For the completeness part, assume the operator [ ] complies with the conditions of Theorem 3.2. Using the Generalized Abstract Nonsense Lemma 2.1 of [7], extend R to a total preorder S satisfying

$$xSy, ySx \Rightarrow xR^\star y. \tag{2}$$

Let $\mathcal{Z}$ be the totally ordered set of equivalence classes of $\mathcal{P}(X)^n$ defined by the total pre-order S. Define a function $d : \mathcal{P}(X)^n \to \mathcal{Z}$ to send a sequence of subsets $A_1, ..., A_n$ to its equivalence class under S. We shall define $r : X^n \to \mathcal{Z}$ by $r(a_1, ..., a_n) \stackrel{\text{def}}{=} d(\{a_1\}, ..., \{a_n\})$. While we aim to prove that r represents [ ], we will have to use d for the proof. d has the following obvious properties:

$$\begin{aligned}&(A_1, ..., A_{n-1}, C)R(A'_1, ..., A'_{n-1}, C')\\ &\Rightarrow d(A_1, ..., A_{n-1}, C) \leq d(A'_1, ..., A'_{n-1}, C'),\end{aligned} \tag{3}$$

and, from Equation 2,

$$\begin{aligned}&d(A_1, ..., A_{n-1}, C) = d(A'_1, ..., A'_{n-1}, C')\\ &\Rightarrow (A_1, ..., A_{n-1}, C)R^\star(A'_1, ..., A'_{n-1}, C').\end{aligned} \tag{4}$$

d also has the less obvious property shown by the next lemma, which means it approximates representation as far as the last argument is concerned:

Lemma 3.2 *For any* $A_1, ..., A_{n-1}, C$*,*

$$d(A_1, ..., A_{n-1}, C) = min_{c \in C}\{d(A_1, ..., A_{n-1}, \{c\})\}$$

and

$$[A_1 \cdots A_{n-1} \cdot C] = \{c \in C \mid d(A_1, ..., A_{n-1}, \{c\}) = d(A_1, ..., A_{n-1}, C)\}. \tag{5}$$

Proof: Suppose $c \in C$, then $(A_1, ..., A_{n-1}, C)R(A_1, ..., A_{n-1}, \{c\})$ and from Equation 3 we get $d(A_1, ..., A_{n-1}, C) \leq min_{c\in C}\{d(A_1, ..., A_{n-1}, c)\}$. If moreover $c \in [A_1 \cdots A_{n-1} \cdot C]$, then $[A_1 \cdots A_{n-1} \cdot C] \cap \{c\} \neq \emptyset$, and by Definition 3.8, part 2, $(A_1, ..., A_{n-1}, \{c\})R(A_1, ..., A_{n-1}, C)$ and therefore $d(A_1, ..., A_{n-1}, c) = d(A_1, ..., A_{n-1}, C)$. We have shown that the left hand side of Equation 5 is included in the right hand side.
Since $[A_1 \cdots A_{n-1} \cdot C]$ is not empty, $\exists c \in [A_1 \cdots A_{n-1} \cdot C]$ and, by the previous remark, $d(A_1, ..., A_{n-1}, C) = d(A_1, ..., A_{n-1}, c)$ and therefore we conclude that $d(A_1, ..., A_{n-1}, C) = min_{c\in C}\{d(A_1, ..., A_{n-1}, c)\}$.
To see the converse inclusion, notice that $d(A_1, ..., A_{n-1}, C) = d(A_1, ..., A_{n-1}, c)$ implies $(A_1, ..., A_{n-1}, c)R^\star(A_1, ..., A_{n-1}, C)$ and, by Property 3 of Theorem 3.2, $[A_1 \cdots A_{n-1} \cdot \{c\}] \subseteq [A_1 \cdots A_{n-1} \cdot C]$, so $c \in [A_1 \cdots A_{n-1} \cdot C]$ by Property 1 of the theorem. ∎

We now have to show that

$$\begin{array}{l} [A_1 \cdots A_{n-1} \cdot C] = \\ \{c \in C \mid \exists a_1, ..., a_{n-1},\ s.t.\ \forall a'_1, ..., a'_{n-1}, c' \in C,\ r(a_1, ..., a_{n-1}, c) \leq r(a'_1, ..., a'_{n-1}, c')\} \end{array} \tag{6}$$

(where $\forall i\ a_i, a'_i \in A_i$).
To see that the right hand side is a subset of the left hand side, assume that $a_i \in A_i, c \in C$ are such that for all $a'_i \in A_i, c' \in C$, $r(a_1, ..., a_{n-1}, c) \leq r(a'_1, ..., a'_{n-1}, c')$. We have to show that $c \in [A_1 \cdots A_{n-1} \cdot C]$. We will show by induction on the size of $A'_1 \cdots A'_{n-1}$, where $a_i \in A'_i \subseteq A_i$ for all i, that $c \in [A'_1 \cdots A'_{n-1} \cdot C]$.
For the base of the induction, if $A'_1 \cdots A'_{n-1}$ is a singleton, then $A'_i = \{a_i\}$ and by Lemma 3.2, remembering that in this case r coincides with d, we find $c \in [A'_1 \cdots A'_{n-1} \cdot C]$.
Otherwise, i.e. if not all A'_i are singletons, we may choose for each i an $a'_i \in A'_i$ such that $a'_i \neq a_i$ if $A'_i \neq \{a_i\}$. Since $A'_1 \cdots A'_{n-1}$ is not a singleton, at least one of the set inequalities holds and $\langle a'_1, \cdots, a'_{n-1}\rangle \neq \langle a_1, \cdots, a_{n-1}\rangle$.
Let $\{B^i_1 \cdots B^i_{n-1}\}$ be the patch to $A'_1 \cdots A'_{n-1}$ from $\{a'_1\} \cdots \{a'_{n-1}\}$. For all i we have, by definition of the patch, $|B^i_1 \cdots B^i_{n-1}| < |A'_1 \cdots A'_{n-1}|$. By choice of a'_i, $\langle a_1, \cdots, a_{n-1}\rangle \in B^i_1 \cdots B^i_{n-1}$ for all i and hence by the induction hypothesis, $c \in [B^i_1 \cdots B^i_{n-1} \cdot C]$. Assume now that $c \notin [A'_1 \cdots A'_{n-1} \cdot C]$, then by Property 3a of Definition 3.8, $(a'_1, ..., a'_{n-1}, C)R(A'_1, ..., A'_{n-1}, C)$. There is some $c' \in [a'_1 \cdots a'_{n-1} \cdot C]$ and by Definition 3.8 part 2, $(a'_1, ..., a'_{n-1}, c')R(a'_1, ..., a'_{n-1}, C)$. Also, by Definition 3.8 part 1, $(A'_1, ..., A'_{n-1}, C)R(a_1, ..., a_{n-1}, c)$.
We have established that $(a'_1, ..., a'_{n-1}, c')R^\star(a_1, ..., a_{n-1}, c)$ so by our original assumption we have $d(a'_1, ..., a'_{n-1}, c') = d(a_1, ..., a_{n-1}, c)$. By Equation 4 this implies also $(a_1, ..., a_{n-1}, c)R^\star(a'_1, ..., a'_{n-1}, c')$ and hence $(a_1, ..., a_{n-1}, c)R^\star(A'_1, ..., A'_{n-1}, C)$. But then by Conditions 3 and 1 of Theorem 3.2, we get $c \in [A'_1 \cdots A'_{n-1} \cdot C]$ which is a contradiction. We conclude that the right hand side of Equation 6 is a subset of the left hand side.

For the converse assume that $c \in C$ is such that for all $a_1, ..., a_{n-1}$ ($a_i \in A_i$), there exist $a'_1, ..., a'_{n-1}$ and $c' \in C$ such that $c \neq c'$ and $r(a_1, ..., a_{n-1}, c) \not\leq r(a'_1, ..., a'_{n-1}, c')$. We need to prove that $c \notin [A_1 \cdots A_{n-1} \cdot C]$. Since X is finite, we may change the order of quantifiers in the assumption to

$$\exists a'_1, ..., a'_{n-1}, c' \in C, \forall a_1, ..., a_{n-1},\ r(a_1, ..., a_{n-1}, c) \not\leq r(a'_1, ..., a'_{n-1}, c')$$

Note that $r(a_1, ..., a_{n-1}, c) \not\leq r(a'_1, ..., a'_{n-1}, c') \Rightarrow \neg(a_1, ..., a_{n-1}, c)R^\star(a'_1, ..., a'_{n-1}, c')$.
As above, but switching the roles of a_i with those of a'_i, let $A'_1, ..., A'_{n-1}$ be such that $a'_i \in A'_i \subseteq A_i$, and prove by induction on $| A'_1 \cdots A'_{n-1} |$, i.e. the cardinality of $A'_1 \cdots A'_{n-1}$, that $c \notin [A'_1 \cdots A'_{n-1} \cdot C]$. If $A'_1 \cdots A'_{n-1}$ is a singleton, then $A'_i = \{a'_i\}$, and by Lemma 3.2 $c \notin [A'_1 \cdots A'_{n-1} \cdot C]$. Otherwise choose for each i an $a_i \in A'_i$ with $a_i \neq a'_i$ if possible, and let $\{B^i_1 \cdots B^i_{n-1}\}$ be the patch to $A'_1 \cdots A'_{n-1}$ from $\{a_1\} \cdots \{a_{n-1}\}$. Again we have for all i, $\langle a'_1, \cdots, a'_{n-1} \rangle \in B^i_1 \cdots B^i_{n-1}$ and $| B^i_1 \cdots B^i_{n-1} | < | A'_1 \cdots A'_{n-1} |$ and hence by induction $c \notin \bigcup_i [B^i_1 \cdots B^i_{n-1} \cdot C]$. Now if $c \in [A'_1 \cdots A'_{n-1} \cdot C]$ we get two consequences. First, by Definition 3.8, part 3b, $(a_1, ..., a_{n-1}, C)R(A'_1, ..., A'_{n-1}, C)$. By the same definition, part 1, we have $(A'_1, ..., A'_{n-1}, C)R(a'_1, ..., a'_{n-1}, c')$. Second, from condition 2 of Theorem 3.2, we derive $c \in [a_1 \cdots a_{n-1} \cdot C]$ so that $(a_1, ..., a_{n-1}, c)R(a_1, ..., a_{n-1}, C)$. The above together give us $(a_1, ..., a_{n-1}, c)R^\star(a'_1, ..., a'_{n-1}, c')$ which contradicts our assumption. We have thus shown the converse inclusion. ∎

4 Conclusion

Iterated updates, of central importance for AI, cannot be properly treated in the AGM framework, not because of the update vs. revision distinction as was thought by Katsuno-Mendelzon, but because of the identification of epistemic states with belief sets, identification accepted by Katsuno-Mendelzon. An ontology in which epistemic states are richer than belief sets yields a family of updates that satisfy the AGM assumptions. We have first proved in Section 2 several properties, which are intuitively appealing, and extend the AGM postulates. In Section 3, we have given a complete set of properties, which characterize our approach, and shown a representation theorem (Theorem 3.2). In summary, we have - on the philosophical side - introduced a natural and intuitively interesting ontology for iterated update, and - on the mathematical side - characterized our idea with a set of sound and complete conditions.

5 Acknowledgements

The authors would like to thank two anonymous referees who have helped to make the article much clearer.

References

[1] Adnan Darwiche and Judea Pearl. On the Logic of Iterated Belief Revision. Artificial Intelligence, 89(1-2):1–29, 1997.

[2] Nir Friedman and Joseph Y. Halpern. Belief Revision: A Critique. Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning, KR'96. Edited by Luigia Carlucci Aiello, Jon Doyle and Stuart Shapiro. Morgan Kaufmann, Cambridge, Mass., November 1996 Pages 421–431.

[3] Peter Gärdenfors. Belief revision: An introduction. In Peter Gärdenfors, editor, *Belief Revision*, number 29 in Cambridge Tracts in Theoretical Computer Science, pages 1–28. Cambridge University Press, 1992.

[4] Peter Gärdenfors. Knowledge in Flux. MIT Press, Cambridge, Massachusetts, 1988.

[5] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In Peter Gärdenfors, editor, *Belief Revision*, number 29 in Cambridge Tracts in Theoretical Computer Science, pages 183–203. Cambridge University Press, 1992.

[6] Daniel Lehmann. Belief revision, revised. In *Proceedings of 14th IJCAI*, pages 1534–1541, Montreal, Canada, August 1995. Morgan Kaufmann.

[7] Daniel Lehmann, Menachem Magidor, Karl Schlechta. Distance Semantics for Belief Revision. Hebrew University Technical Report TR-98-10. Department of Computer Science, Hebrew University, Givat Ram, Jerusalem 91904, Israel.

[8] B.Nebel. A Knowledge Level Analysis of Belief Revision. In *Proceed. First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, 1989.

[9] Karl Schlechta, Daniel Lehmann and Menachem Magidor. Distance Semantics for Belief Revision. Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge Edited by Yoav Shoham. Morgan Kaufmann, De Zeeuwse Stromen, The Netherlands, March 1996 Pages 137–145.

[10] Karl Schlechta. Preferential Choice Representation Theorems for Branching Time Structures. Journal of Logic and Computation, Oxford, Vol. 5, pp. 783-800, 1995.