Explaining Updates by Minimal Sums*

Jürgen Dix[1]** and Karl Schlechta[2]

[1] University of Maryland, College Park, MD 20752, USA
dix@cs.umd.edu,
WWW home page: http://www.uni-koblenz.de/~dix
[2] Laboratoire d'Informatique de Marseille, CNRS ESA 6077, CMI,
39 rue Joliot Curie, F-13453 Marseille Cedex 13, France
ks@gyptis.univ-mrs.fr
WWW home page: http://protis.univ-mrs.fr/~ks

Abstract. Human reasoning about developments of the world involves always an assumption of *inertia*. We discuss two approaches for formalizing such an assumption, based on the concept of an *explanation*: *(1)* there is a general preference relation $\prec$ given on the set of all explanations, *(2)* there is a notion of a *distance* dist between models and explanations are *preferred* if their sum of distances is minimal. Each *distance* dist naturally induces a preference relation $\prec_{\mathsf{dist}}$. We show exactly under which conditions the converse is true as well and therefore both approaches are equivalent modulo these conditions. Our main result is a general representation theorem in the spirit of Kraus, Lehmann and Magidor.

1 Introduction

Reasoning about *developments* or *changing* situations[1] is an important problem in Artificial Intelligence, as has been recognized very early. Much of human reasoning about these problems is based on the assumption that the world is relatively *static*. We will, for instance, hesitate to accept an explanation as *plausible* which involves many and unmotivated changes.

Generally, there is always an assumption of *inertia* formalizing that certain properties tend to *persist over time*. Many nonmonotonic logics have been used to formalize persistency ([BDK97]). E.g. circumscriptive approaches try to minimize change by circumscribing certain predicates (see [ZF93,KL95]). Default logics formalize persistency by stating special *default rules* ([HM87]). In logic programming, various versions of *negation-as-failure* have been defined to specify that fluents persist if other fluents *can not be proved* to hold (see [BD99,GL93]).

In this paper we generalize a particular approach introduced in [Win88,Win89] and [Dal88]. The overall framework is propositional logic with respect to an underlying signature $\mathcal{L}$. We denote by $Mod_{\mathcal{L}}$ the set of all propositional models

* A full version of this paper, including detailed proofs, can be obtained under URL:http://www.cs.umd.edu/~dix/pub_journals.html.
** Currently on leave from University of Koblenz-Landau.
[1] The term *situation* is not to be confused with the same term in *situation calculus*.

with respect to $\mathcal{L}$. In this paper, however, we do not make use of the fact that $Mod_{\mathcal{L}}$ is induced by $\mathcal{L}$. We abstract from this and view this set simply as a set of *worlds* denoted by W. The actual world can then be simply represented as an element of W. In most cases, however, we do not know the actual world. All we know is the current *situation* which is a set of worlds.

Definition 1 (Situation S). *A situation S is a set of worlds: $S \subseteq W$. As usual, S can be also viewed as a set Th of $\mathcal{L}$-formulae: via Gödel's completeness theorem, Th induces the set $\{\mathcal{A} : Th \models \mathcal{A}\} \subseteq W$.*

How the world actually evolves (while certain actions occur) will be described by a *sequence of worlds.*

Definition 2 (Sequence σ, Explanation Expl). *A sequence, denoted by σ, $\langle \mathcal{A}_1, \ldots, \mathcal{A}_n \rangle$, is a finite list of worlds: $\mathcal{A}_i \in W$. We also denote it by $\sigma := \langle \sigma_1, \ldots, \sigma_n \rangle$. We say that the sequence σ* explains *the change from situation S to situation $S^\star$, if, by definition, $\sigma_1 \in S$ and $\sigma_n \in S^\star$.*

The sequence σ is also called an explanation *(this use was suggested by Daniel Lehmann and the authors adopt it) for the change of S to $S^\star$. We denote by $\mathsf{Expl}(S, S^\star)$ the set of all such explanations. By Expl we mean the set*

$$\mathsf{Expl}(2^W, 2^W) := \bigcup_{S, S^\star \subseteq W} \mathsf{Expl}(S, S^\star)$$

of all explanations of all possible pairs $(S, S^\star)$.

Thus, a sequence σ describes the development of the world. We note that in general, the change from an initial situation S to another situation $S^\star$ may be described by several different developments, even if both S and $S^\star$ consist of just one world. Before going on with the technical definitions, some general comments about our approach are in order:

1. We assume discrete time, and a sequence of observations: At time 1, we observed S_1, at time 2 S_2 etc., where the S_i are (usually not complete) theories corresponding to sets of worlds.
2. An explanation of this sequence is a sequence of worlds σ_1, σ_2, etc., with $\sigma_i \in S_i$.
3. Thus, given a fixed sequence of observations, its explanations all have the same length.
4. We are interested to single out the *best* or *more plausible* explanations, and do this by an assumption of a *distance*. A *distance* between worlds reflects the "cost" or "probability" of a change from one world to the other.
5. Consequently, we consider sequences of worlds with a small sum of distances between the individual worlds of this sequence as more probable or plausible than those whose sum of distances is big. Thus, such "minimal" explanations will be considered best explanations of a given sequence of observations.
6. A classical example of a distance between worlds is the Hamming distance, i.e. the number of propositional variables in which they differ. Other distances are considered e.g. in [Sch95a].

7. Depending on our assumptions about the world, a number of approaches are possible. First, we can assume an abstract, arbitrary order between explanations, this idea was pursued in [Sch95b]. Second, we can assume that explanations *with repetitions* (i.e. the world has not changed at a certain moment) are better than those without repetitions. Thus, the sequence $\langle w, w\rangle$ is considered better than the sequence $\langle w, w'\rangle$—provided both explain a given sequence of observations. This idea was pursued in [BLS98]. In the present paper, we push the idea of inertia further, minimizing the sum of changes involved in a sequence of worlds.

Such sequences from S to S^*, or explanations, may represent different *grades* of plausibility: some sequences are less plausible than others. This leads to the notion of a *plausible* explanation illustrated in the next example.

Example 1 (Plausible Explanations).
Sequences that contain loops of the form $\langle \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_1\rangle$ and thus are unnecessary long, should not be considered as *plausible* explanations. A criterion of *inertia* is needed in order to rule out the unmotivated sequences and to define the set of plausible explanations.

Of course, the most general approach is to just assume any preference relation between explanations.

Definition 3 (Preference Relation $\prec$). *A preference relation $\prec$ is any relation on the set of all explanations* Expl

$$\prec \subseteq \mathsf{Expl} \times \mathsf{Expl}.$$

We call an explanation σ $\prec$-preferred*, if by definition, σ is minimal with respect to $\prec$, i. e. there is no other explanation $\sigma' \neq \sigma$ with $\sigma \prec \sigma'$.*

In [Sch95b,BLS98], the authors state general representation results for preference relations between arbitrary sequences of models.

A more intuitive approach to exclude such examples is due to [Win88,Win89]. The idea is to assume the notion of a distance between arbitrary worlds.

Definition 4 (Distance dist). *A distance* dist *is a function that associates to any two worlds a nonnegative rational number:*

$$\mathsf{dist} : W \times W \longrightarrow \mathbb{Q}; (\mathcal{A}_1, \mathcal{A}_2) \mapsto \mathsf{dist}(\mathcal{A}_1, \mathcal{A}_2)$$

The idea of [Win88,Win89] was to *measure* the sum of all distances in a sequence and to consider those sequences as most plausible that correspond to minimal sums.

Definition 5 (Plausible Explanations Induced by dist: $\prec_{\mathsf{dist}}$). *Let a distance be given on W^2. Let also two situations S, $S^\star$ and two explanations for the change from S to $S^\star$ $\sigma := \langle \sigma_1, \ldots, \sigma_n\rangle$, $\sigma^\star := \langle \sigma_1^\star, \ldots, \sigma_m^\star\rangle$ be given (i. e. σ_1 and $\sigma_1^\star$ are both contained in S and σ_n and $\sigma_m^\star$ are both contained in $S^\star$).*

We say that σ is more plausible than $\sigma^\star$, *denoted by* $\sigma \prec_{\mathsf{dist}} \sigma^\star$ *if, by definition*

$$\sum_{i=1}^{n-1} \mathsf{dist}(\sigma_i, \sigma_{i+1}) \leq \sum_{i=1}^{m-1} \mathsf{dist}(\sigma^\star_i, \sigma^\star_{i+1}). \tag{1}$$

The most plausible explanations for the change from S *to* $S^\star$ *are those whose sum of distances is minimal.*

For a sequence σ *we denote by* $\mathsf{sum\text{-}dist}(\sigma)$ *the number* $\sum_{i=1}^{n-1} \mathsf{dist}(\sigma_i, \sigma_{i+1})$.

Thus, if the notion of a *distance* is available, we can immediately define an induced preference relation $\prec_{\mathsf{dist}}$. But is the converse also true? I.e. given an arbitrary preference relation $\prec$ between possible explanations, *does there exist a measure of distance* dist *on* W *such that* $\prec = \prec_{\mathsf{dist}}$?.

The aim of this paper is to completely solve this question by characterizing those preference relations $\prec$ which can be generated by a distance dist. To do this we use (an adaptation of) an old algorithm, going back to [Far02], to determine whether a set of inequalities of sums has a solution.

The plan of the paper is as follows. After introducing some additional terminology in Section 2, we start in Section 3 with Proposition 1, stating that the preferred sequences are completely determined by the endpoints of certain intermediate sequences. We then give a precise formulation of our main theorem and end with a sketch of the proof of this result. This proof depends heavily on a abstract representation result shortly reviewed in the appendix. We conclude with Section 4 by citing related approaches.

2 Terminology

As already mentioned above, we assume discrete time, given by the integers $\mathbb{N}$. We also assume that we we have only *incomplete information* about the state of affairs at times $t_1, \ldots, t_n$. This information is given by a sequence of situations (i. e. sets of models)

Definition 6 ($\Pi_{\mathsf{fin}}(2^W)$). *A sequence* $\Sigma := \langle \Sigma_1, \ldots, \Sigma_n \rangle$ *is a finite list of situations:* $\Sigma_i \subseteq W$. *Equivalently, we can view* Σ *as the product* $\Pi_{i=1}^n \Sigma_i$. *A sequence* Σ *represents our knowledge about the world at times* $t_1, \ldots, t_n$. *We denote by* $\Pi_{fin}(2^W)$ *the union of all finite products of situations (sets of worlds) in* W*:*

$$\Pi_{fin}(2^W) := \bigcup_{n \in \mathbb{N}} \{ \Pi_{i=1}^n \Sigma_i^j \mid \emptyset \neq \Sigma_i^j \subseteq W \}$$

We denote by $\Sigma_{|i_0}$ *the restriction of* Σ *to the first* i_0 *components.*

If there is a distance dist we can determine the set of those sequences σ with $\sigma_i \in \Sigma_i$ for which $\mathsf{sum\text{-}dist}(\sigma)$ (see Definition 5) is minimal. We call such sequences dist-preferred sequences. Analogously, given a preference relation $\prec$ defined on W, we call sequences $\prec$-preferred if there are no other sequences σ with $\sigma_i \in \Sigma_i$ that are smaller with respect to the relation $\prec$.

Definition 7 (dist- and $\prec$-Preferred Sequences). *Let a sequence of situations $\Sigma := \langle \Sigma_1, \ldots, \Sigma_n \rangle$ be given. We denote by $\mathsf{Pref}_{\mathsf{dist}}(\Sigma)$ (resp. $\mathsf{Pref}_{\prec}(\Sigma)$) the set of* dist*-preferred (resp. $\prec$-preferred) sequences of worlds that are compatible with Σ: those sequences σ satisfying*

1. *$\sigma_i \in \Sigma_i$ (in particular σ and Σ have the same length),*
2. *sum-dist(σ) is minimal (resp. σ is $\prec$-preferred) among all sequences satisfying (1).*

Note that $\mathsf{Pref}_{\mathsf{dist}}(\Sigma)$ (resp. $\mathsf{Pref}_{\prec}(\Sigma)$) are plausible explanations for the change of situation Σ_1 to Σ_n.

We now associate to any sequence of situations Σ the set of endpoints of dist- (resp. $\prec$-) preferred sequences compatible with Σ.

Definition 8 ($\mathsf{End}_{\prec}, \mathsf{End}_{\mathsf{dist}}$). *We define the following functions, depending on the underlying preference relation* dist *or $\prec$.*

$$\begin{array}{lll} \mathsf{End}_{\mathsf{dist}} : & \Pi_{fin}(2^W) & \to 2^W; \\ & \Sigma & \mapsto \{\mathcal{A} \in W \mid \exists \sigma \in \mathsf{Pref}_{\mathsf{dist}}(\Sigma) \text{ s.t.}^{a}\ \sigma_n = \mathcal{A}\} \\ \mathsf{End}_{\prec} : & \Pi_{fin}(2^W) & \to 2^W; \\ & \Sigma & \mapsto \{\mathcal{A} \in W \mid \exists \sigma \in \mathsf{Pref}_{\prec}(\Sigma) \text{ s.t.}^{a}\ \sigma_n = \mathcal{A}\} \end{array}$$

[a] Note that Σ and σ have the same length, say n.

The function $\mathsf{End}_{dist}(\cdot)$ for given dist has certain properties, which we will later use to completely characterize it. This means we will prove a theorem of the form

> *If the function $\mathsf{End}_{\prec} : \Pi_{fin}(2^W) \to 2^W$ satisfies certain properties, then there is a distance* dist *on W^2 such that $\mathsf{End}_{\prec}(\Sigma) = \mathsf{End}_{dist}(\Sigma)$ for all $\Sigma \in \Pi_{fin}(2^W)$.*

We would like to emphasize that our approach only assumes knowledge about $\mathsf{End}_{dist}(\Sigma)$, i. e. about the endpoints in Σ_n. We do not assume anything about the *endpoints of intermediate sequences of length less than n*: $\mathsf{End}_{dist}(\Sigma_{|i})$ for $i < n$.

Consequently, from $\mathsf{End}_{dist}(\Sigma)$ the set of all dist-preferred sequences can not be reconstructed. On the other hand we show in Proposition 1 (Section 3) that knowledge of the endpoints of intermediate sequences allows us to completely reconstruct $\mathsf{Pref}_{\mathsf{dist}}(\Sigma)$.

Although it is not needed to formulate our problem, the following extension of $\prec$ from a relation between sequences σ to a relation between sequences Σ is very important in the proof of our main result.

Remark 1 (Extending $\prec$, $\prec_{\mathsf{dist}}$ to sequences Σ). The relations $\prec$, (resp. $\prec_{\mathsf{dist}}$) can be straightforwardly extended to relations between sequences Σ:

$$\Sigma \prec \Sigma' \text{ if, by definition, } \sigma \prec \sigma' \text{ for all } \sigma \in \mathsf{Pref}_{\prec}(\Sigma), \sigma' \in \mathsf{Pref}_{\prec}(\Sigma'), \quad (\textit{resp. } \sigma \in \mathsf{Pref}_{\mathsf{dist}}(\Sigma), \sigma' \in \mathsf{Pref}_{\mathsf{dist}}(\Sigma')).$$

We assume that we have the *information* $\mathsf{End}_{dist}(\Sigma)$ only about products Σ of sets of models, but not about arbitrary sets of sequences. Thus,

$$\mathsf{End}_{dist}(\{a, a'\} \times \{b, b'\})$$

will be given, but, if $a \neq a'$ and $b \neq b'$, then $\mathsf{End}_{dist}(\{\langle a, b\rangle, \langle a', b'\rangle\})$ may not necessarily be defined—the sequences $\langle a, b'\rangle$, $\langle a', b\rangle$ are missing. On the other hand, we assume that we can *reason* about unions of sets of sequences, in particular if a union of products of sets is itself a product of sets, like

$$\{a, a'\} \times \{b, b'\} = (\{a\} \times \{b\}) \cup (\{a\} \times \{b'\}) \cup (\{a'\} \times \{b\}) \cup (\{a'\} \times \{b'\}).$$

Definition 9 (Legal Sets of Sequences). *We call a set of sequences (of situations)* legal*, if this set is a product of sets.*

Thus, we can reason about arbitrary sets of sequences, but the *world* does not give us information about arbitrary, only about *legal* sets of sequences. It seems a natural hypothesis that the language of reasoning may be stronger than the language of observation.

Obviously, the Σ_n are in a stronger position than the other intermediate Σ_i, by definition of $\mathsf{End}_{dist}(\Sigma)$. This corresponds to the fact that, considering a development into the future, we are probably most interested in the final outcome. Conversely, given a development from the past to the present, we might have most information about the present.

There are, however, other directions of possible interest, and the reader will see how to adapt our conditions and proofs to the case which interests him. We examine in this paper the two extremes—all $\mathsf{End}_{dist}(\Sigma_{|i})$ are known, and, only one $\mathsf{End}_{dist}(\Sigma_{|i})$ is known. It should not be too difficult to modify our results and techniques accordingly.

3 Updating by Minimal Sums

Before formulating our main results, we need some additional notation: If σ is a sequence and a a point, σa will be the concatenation of σ with a. Consequently

1. $\sigma \times A$ will denote the set of all sequences σa, $a \in A$.
2. $\Sigma \times A$ will denote the set of all sequences σa, $\sigma \in \Sigma$, $a \in A$. Likewise $\Sigma \times a$ by abuse of notation.

The following lemma illustrates that, if we also know the preferences for suitable intermediate observations, we can totally determine the preferred sequences. The meaning of "suitable" will become clear in the proof of the lemma.

Proposition 1 ($\mathsf{Pref}_{\mathsf{dist}}(\Sigma)$ Induced By Intermediate $\mathsf{End}_{dist}(\Sigma_{|i})$). *Let Σ be a sequence in the sense of Definition 7.*

$\mathsf{Pref}_{\mathsf{dist}}(\Sigma)$ *is reconstructible from* $\mathsf{End}_{dist}(\Sigma'_{|i})$ *for suitable Σ' with $\Sigma'_i \subseteq \Sigma_i$.*

Proof. Fix i.

Case 1: $\mathsf{End}_{dist}(\Sigma_{|i}) = \{a_i\}$. Then for all $x \in \mathsf{End}_{dist}(\Sigma_{|i-1})$ there is a preferred sequence containing $\langle x, a_i\rangle$ as a subsequence. Likewise for $y \in \mathsf{End}_{dist}(\Sigma_{|i+1})$.

Case 2: $|\mathsf{End}_{dist}(\Sigma_{|i})| > 1$. If $|\mathsf{End}_{dist}(\Sigma_{|i-1})| = 1$, we apply Case 1 to $i-1$. Suppose $|\mathsf{End}_{dist}(\Sigma_{|i-1})| > 1$, and, for the same reason, $|\mathsf{End}_{dist}(\Sigma_{|i+1})| > 1$. Fix $a_i \in \mathsf{End}_{dist}(\Sigma_{|i})$, and consider $\Sigma[i/\{a_i\}]$, where Σ_i has been replaced by $\{a_i\}$, i.e.

$$\Sigma[i/\{a_i\}] =_{\text{def}} \Sigma_1 \times \ldots \times \{a_i\} \times \ldots \times \Sigma_n.$$

If $a_{i-1} \notin \mathsf{End}_{dist}(\Sigma[i/\{a_i\}]_{|i-1})$, then there is no preferred sequence through $\langle a_{i-1}, a_i\rangle$ in Σ : Any such sequence σ' through $\langle a_{i-1}, a_i\rangle$ is already in $\Sigma[i/\{a_i\}] \subseteq \Sigma$, and there is a better one in $\Sigma[i/\{a_i\}] \subseteq \Sigma$.
Suppose $a_{i-1} \in \mathsf{End}_{dist}(\Sigma[i/\{a_i\}]_{|i-1})$. As $a_i \in \mathsf{End}_{dist}(\Sigma_{|i})$, there is a preferred sequence in Σ through a_i. It is already in $\Sigma[i/\{a_i\}]$. But in $\Sigma[i/\{a_i\}]$, there is one through all $a_{i-1} \in \mathsf{End}_{dist}(\Sigma[i/\{a_i\}]_{|i-1})$. By rankedness, all are preferred in Σ. So there is a preferred sequence in Σ through $\langle a_{i-1}, a_i\rangle$ for all $a_{i-1} \in \mathsf{End}_{dist}(\Sigma[i/\{a_i\}]_{|i-1})$. The same argument applies to $i+1$.
Suppose now $\sigma, \sigma' \in \mathsf{Pref}_{\mathsf{dist}}(\Sigma)$, and $\sigma_i = \sigma'_i$. Let $\sigma = \sigma^{\#}\Sigma\sigma*$, where $\sigma^{\#} = \sigma_1 \ldots \sigma_i$, $\sigma_* = \sigma_{i+1} \ldots \sigma_n$. Likewise let $\sigma' = \sigma'^{\#}\sigma'^{*}$. Then also $\sigma^{\#}\sigma'^{*}$ and $\sigma'^{\#}\sigma^{*} \in \mathsf{Pref}_{\mathsf{dist}}(\Sigma)$. For if not, then e.g. $\mathsf{sum\text{-}dist}(\sigma^{\#}) > \mathsf{sum\text{-}dist}(\sigma'^{\#})$, as $\sigma' \in \mathsf{Pref}_{\mathsf{dist}}(\Sigma)$, but then $\mathsf{sum\text{-}dist}(\sigma^{\#}) + \mathsf{sum\text{-}dist}(\sigma^{*}) > \mathsf{sum\text{-}dist}(\sigma'^{\#}) + \mathsf{sum\text{-}dist}(\sigma^{*})$, contradicting $\sigma \in \mathsf{Pref}_{\mathsf{dist}}(\Sigma)$.
Thus, any sequence constructed as follows:

$$\begin{array}{l} a_i \in \mathsf{End}_{dist}(\Sigma_{|i}) \\ a_{i-1} \in \mathsf{End}_{dist}(\Sigma[i/\{a_i\}]_{|i-1}) \\ a_{i+1} \in \mathsf{End}_{dist}(\Sigma[i/\{a_i\}]_{|i+1}) \end{array}$$

belongs to $\mathsf{Pref}_{\mathsf{dist}}(\Sigma)$, and no others. □

Our next theorem is the main result of this paper. In Section 2 we noted that a preference relation $\prec$ between worlds implies the existence of a function

$$\mathsf{End}_{\prec} : \Pi_{fin}(2^W) \to 2^W.$$

In general, the properties of this function depend on the underlying $\prec$ relation. Indeed, if there is a distance dist then the induced function

$$\mathsf{End}_{\mathsf{dist}} : \Pi_{fin}(2^W) \to 2^W$$

has a lot of properties, due to this distance function. In our main theorem we want to completely characterize the general function $\mathsf{End}_{\prec}(\cdot)$ by suitable such properties.

For the following, let therefore $\mathsf{End}_\prec(\cdot)$ be any function from $\Pi_{fin}(2^U)$ to 2^U.

We are looking for conditions on $\mathsf{End}_\prec(\cdot)$ which guarantee the existence of a distance with suitable order and addition on the values and which singles out $\mathsf{End}_\prec(\Sigma)$ exactly for all *legal* Σ. If the relation $\prec$ is induced from a distance dist then the following holds:

Criterion 10 (Important Conditions)
(C1) $\mathsf{End}_\prec(\Sigma) \subseteq \Sigma_n$, *if* Σ_n *is the last component of* Σ,
(C2) $\Sigma \neq \emptyset \rightarrow \mathsf{End}_\prec(\Sigma) \neq \emptyset$,
(C3) $\mathsf{End}_\prec((\bigcup \Sigma_i) \times B) \subseteq \bigcup \mathsf{End}_\prec(\Sigma_i \times B)$.

There is one last condition that we need in order to prove our equivalence result: the (Loop) criterion. Before giving the technical details, we give some intuitive explanations:

1. If a choice function can be represented by a distance, then the relation generated by it must be *free of loops*. So it must not be possible to conclude from the given information that $a+b \preceq c+d \prec a+b$, otherwise, there would be no distances a, b, c, d and addition $+$ representing it. Thus the Loop condition constrains the general preference relation $\prec$. As we put sufficiently many operations into the loop condition, the Farkas algorithm used in our proof will terminate and generate the representing distance.
2. Note that the central conditions for representability in [LMS99] (conditions $(|\ S1)$, $(|\ A2)$, $(|\ A3)$, $(*S1)$, $(*A2)$, $(*A3)$) are also essentially loop conditions. This is not surprising, as the problem there is similar to the one posed here: we try to embed a partial order into a total order, and this can only be done if the strict part of the partial order does not contain any loops.

The proof of our main result uses an abstract representation theorem which is given in the appendix (to make the paper self-contaned). One of the important ingredients of this representation theorem is a certain equivalence relation $\equiv$. We define this relation on the set of all sequences of worlds as follows:

$$\sigma \equiv \sigma' \text{ if, by definition } \sigma \text{ and } \sigma' \text{ have the same endpoint.}$$

1. If σ, σ' have the same length, then $[\sigma, \sigma'] := \{\sigma'' : \sigma''_i \in \{\sigma_i, \sigma'_i\} \text{ for all } i\}$. Note that $[\sigma, \sigma']$ is a (legal) product of sets (of size ≤ 2). Likewise, if Σ is a legal set of sequences, and σ a sequence, both of same length, then $[\Sigma, \sigma] := \{\sigma' : \sigma'_1 \in \Sigma_i \cup \{\sigma_i\}\}$.
2. If σ, σ' are two sequences with $\sigma_n = \sigma'_1$, then $\sigma\sigma'$ denotes their concatenation $\langle \sigma_1, \ldots, \sigma_n, \sigma'_n, \ldots, \sigma'_{n'} \rangle$. We write this also as $\sigma_1 \times \ldots \times \sigma_n \times \sigma'_1 \times \ldots \times \sigma'_{n'}$.

Definition 11 (Hamming Distance). *If* Σ *is a set of sequences,* σ *a sequence, both of the same length, then the* Hamming distance $\mathsf{hamm}(\Sigma, \sigma)$ *will be the minimum of the Hamming distances* $\mathsf{hamm}(\sigma', \sigma)$, $\sigma' \in \Sigma$. *(The Hamming distance between two sequences* σ, σ' *of equal length is the number of* i*'s s.t.* $\sigma_i \neq \sigma'_i$*).*

We have to state the following definition which contains the important conditions $(R_1) - -(R_6)$ and $(+1) - -(+5)$ used in the **(Loop)**-condition of Criterion 13.

Definition 12 (Constructing $\prec$ and $\preceq$). *Originally, $\prec$ is only a relation between sequences σ, σ'. Here we extend $\prec$ to (1) a relation between arbitrary sums of sequences, and (2) to a relation between sequences Σ.*

$\prec$, $\preceq$ and Addition: *Let us consider in an abstract setting arbitrary sums of distances of sequences σ. I.e. we start with a set $\{a_{\sigma'} : a_{\sigma'}$ is a subsequence of σ, σ is a sequence $\}$ and equip it with a binary function $+$. So we consider the set $\{a_\sigma + \ldots + a_{\sigma'}) : \sigma, \sigma'$ sequences $\}$. In the following we will formulate conditions to constrain the interaction between $+$ and $\prec$. (The terms a_σ, a_τ correspond to one sequence. When they are compared, they are of equal length. ($\overset{\leqq}{\geqq}$ stands for $\preceq$ and $\succeq$ simultaneously.)*

(+1) $a_\sigma + a_\tau \overset{\leqq}{\geqq} a_\tau + a_\sigma$,

(+2) $(a_\sigma + a_\tau) + a_\eta \overset{\leqq}{\geqq} a_\sigma + (a_\tau + a_\eta)$,

(+3) $a_\sigma \overset{\leqq}{\geqq} a_{\sigma'} \to ((a_\tau \overset{\leqq}{\geqq} a_{\tau'}) \leftrightarrow (a_\sigma + a_\tau) \overset{\leqq}{\geqq} (a_{\sigma'} + a_{\tau'}))$,

(+4) $a_\sigma \overset{\leqq}{\geqq} a_{\sigma'} \to (a_\tau \prec a_{\tau'} \leftrightarrow (a_\sigma + a_\tau \prec a_{\sigma'} + a_{\tau'})$,

(+5) $(a_\sigma \prec a_{\sigma'} \wedge a_\tau \prec a_{\tau'}) \to (a_\sigma + a_\tau \prec a_{\sigma'} + a_{\tau'})$.

$\prec$, $\preceq$ and Comparisons: *Here we extend $\prec$ to a relation between sequences Σ. This is done by using the function $\mathsf{End}_\prec$. (In* (R4), (R5) *i ranges over some index set I.)*

(R1) $\Sigma \times B \preceq \Sigma \times B'$ *if* $\mathsf{End}_\prec(\Sigma \times (B \cup B')) \cap B \neq \emptyset$,

(R2) $\Sigma \times B \prec \Sigma \times B'$ *if* $\mathsf{End}_\prec(\Sigma \times (B \cup B')) \cap B' = \emptyset$.

For the left hand side:

(R3) $\Sigma \times B \preceq \Sigma' \times B$ *if* $\Sigma' \subseteq \Sigma$,

(R4) $(\forall i \in I\colon \Sigma' \times B \preceq \Sigma_i \times B)$ *if* $\mathsf{End}_\prec((\Sigma' \cup \bigcup \Sigma_i) \times B) \not\subseteq \bigcup \mathsf{End}_\prec(\Sigma_i \times B)$.

(R5) $(\forall i \in I\colon \Sigma' \times B \prec \Sigma_i \times B)$ *if* $\bigcap \mathsf{End}_\prec(\Sigma_i \times B) \not\subseteq \mathsf{End}_\prec((\Sigma' \cup \bigcup \Sigma_i) \times B)$.

(R6) If $\bigcap_{i \in I} \Sigma_i \neq \emptyset$, *then* $\mathsf{End}_\prec(\Sigma) = \bigcap_{i \in I} \Sigma_i$.

With the help of the notions introduced in the last definition, we define the *(Loop)*-criterion:

Criterion 13 (Loop) *The (smallest) relation defined by* (R1)–(R6), (+1)–(+5) *(see Definition 12) contains no loops involving $\prec$ (i.e. loops involving $\preceq$ are allowed, but no loops with the "strictly less" relation $\prec$). In other words, the transitive closure of this relation is antisymmetric.*

Again, if the relation $\prec$ is induced from a distance dist then the (Loop) criterion is satisfied, as can be easily checked.

Theorem 1 (Representation Theorem). *Let W, the set of explanations* Expl *and a relation $\prec \subseteq \mathsf{Expl} \times \mathsf{Expl}$ be given. Then the following are equivalent:*

1. *There is a distance* dist *from $W \times W$ into an ordered abelian group, such that* $\mathsf{End}_\prec(\Sigma) = \mathsf{End}_{\mathsf{dist}}(\Sigma)$ *and* $\mathsf{Pref}_{\mathsf{dist}}(\Sigma) = \mathsf{Pref}_\prec(\Sigma)$.
2. *The function* $\mathsf{End}_\prec$ *satisfies the conditions of Criteria 10 and 13.*

Proof (Sketch). Due to lack of space, we can provide the reader only with a small sketch of the proof. For the detailed proof, which is rather involved and uses the abstract representation result given in the appendix, we refer to URL:http://www.cs.umd.edu/~dix/pub_journals.html.

The direction from *(1)* to *(2)* is trivial. It remains to show that *(2)* implies *(1)*.

Assume a function $\mathsf{End}_{\prec}$ satisfying Criteria 10 and 13. Suppose the required distance function $\mathsf{dist}(i,j)$ between two neighbouring worlds in a sequence is modelled by the variable $x_{i,j}$. Then, for a sequence $\sigma = \langle 1, 2, ..., m \rangle$, the distance function would yield the sum

$$x_{1,2} + ... + x_{m-1,m}.$$

A similar sum is built up for a sequence τ.

If $\sigma \prec \tau$, this leads to an inequality for the two sums. In this way, a system of inequalities is built up. We solve this system by using a modification of an algorithm communicated by S. Koppelberg, Berlin. The original algorithm seems to be due to [Far02]. The crucial loop criterion is used to ensure that a solution exists. It then remains to show that $\mathsf{End}_{\prec}(\Sigma) = \mathsf{End}_{dist}(\Sigma)$.

4 Conclusions and Acknowledgements

One of the most distinguishing features of classical reasoning as applied in mathematics and *human reasoning* as applied in everyday life, is the treatment of how the world changes over time. Humans use the fact, often induced by context, that certain properties *persist over time*. Frameworks for studying the formalization of this persistence are very important to develop reasoning calculi that can be applied for realistic scenario. The many frameworks for belief revision—as studied in the last 15 years—all treat this problem.

There have been proposed a lot of systems for dealing with this persistence problem. For example, depending on our assumptions about the world, a number of approaches are possible:

1. We can assume an abstract, arbitrary order between explanations, this idea was pursued in [Sch95b] and [Dal88].
2. We can assume that explanations *with repetitions* (i.e. the world has not changed at a certain moment) are better than those without repetitions. Thus, the sequence $\langle w, w \rangle$ is considered better than the sequence $\langle w, w' \rangle$—provided both explain a given sequence of observations. This idea was pursued in [BLS98].
3. In the present paper, we push the idea of inertia further, minimizing the sum of changes involved in a sequence of worlds ([Win88,Win89]).

We have shown in this paper the exact relationship between these approaches. We developed a general representation result in the spirit of [KLM90], Theorem 1, stating under exactly what conditions an arbitrary *preference ordering* is induced by a *distance on the underlying models*.

We note in particular that although the main theorem can be stated without too much technical machinery, its proof requires quite a bit of technical notation. We also note our use of an old result of Farkas: this shows once again that mathematical results considered quite exotic still find their applications in modern computer science.

We owe special thanks to three anonymous referees for their careful reading and their numerous suggestions to improve this paper. In particular, one of the referees read the paper extremely carefully and his suggestions were greatly appreciated.

References

[BD99] Gerhard Brewka and Jürgen Dix. Knowledge representation with extended logic programs. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd Edition, Volume 6, Methodologies*, chapter 6. Reidel Publ., 1999.

[BDK97] Gerd Brewka, Jürgen Dix, and Kurt Konolige. *Nonmonotonic Reasoning: An Overview*. CSLI Lecture Notes 73. CSLI Publications, Stanford, CA, 1997.

[BLS98] Shai Berger, Daniel Lehmann, and Karl Schlechta. Preferred history semantics for iterated updates. Technical Report TR-98-11, Leibniz Center for Research in Computer Science, Hebrew University, Givat Ram, Jerusalem 91904, Israel, July 1998. to appear in *Journal of Logic and Computation*.

[Dal88] M. Dalal. Investigations into a theory of knowledge bases revisions. In *NCAI'88*, pages 475–479. Morgan Kaufmann, 1988.

[LMS99] Menachem Magidor, Daniel Lehmann and Karl Schlechta. Distance semantics for belief revision. *Journal of Symbolic Logic*, to appear, 1999.

[Far02] J. Farkas. Theorie der einfachen Ungleichungen. *Crelles Journal für die Reine und Angewandte Mathematik*, 124:1–27, 1902.

[GL93] Michael Gelfond and Vladimir Lifschitz. Representing Actions and Change by Logic Programs. *Journal of Logic Programming*, 17:301–322, 1993.

[HM87] S. Hanks and D. McDermott. Nonmonotonic Logic and Temporal Projection. *Artificial Intelligence*, 33:379–412, 1987.

[KL95] G. N. Kartha and V. Lifschitz. A Simple Formalization of Actions Using Circumscription. In Chris S. Mellish, editor, *14th IJCAI*, volume 2, pages 1970–1975. Morgan Kaufmann, 1995.

[KLM90] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic Reasoning, Preferential Models and Cumulative Logics. *Artificial Intelligence*, 44(1):167–207, 1990.

[Sch95a] Karl Schlechta. Logic, topology and integration. *Journal of Automated Reasoning*, 14:353–381, 1995.

[Sch95b] Karl Schlechta. Preferential choice representation theorems for branching time structures. *Journal of Logic and Computation*, 5:783–800, 1995.

[Win88] Marianne Winslett. Reasoning about action using a possible models approach. In *Proc. 7th AAAI-88*, pages 89–93, 1988.

[Win89] Marianne Winslett. Sometimes updates are circumscription. In *Proc. 11th Int. Joint Conf. on Artificial Intelligence (IJCAI'89)*, pages 859–863, 1989.

[ZF93] Yan Zhang and Norman Y. Foo. Reasoning about persistence: A theory of actions. In Ruzena Bajcsy, editor, *Proc. of the 13th Int. Joint Conf. on Artificial Intelligence (IJCAI'93)*, pages 718–723. Morgan Kaufmann, 1993.

A An Abstract Representation Result

Definition 14 (The Abstract Framework). *Let the following be given:*

1. *a nonempty universe U, an arbitrary set,*
2. *a function $\Omega : \Pi_{fin}(2^U) \to 2^U$,*
3. *an equivalence relation $\equiv$ on U (we write $[\![u]\!]$ for the equivalence class of $u \in U$ under $\equiv$) such that $[\![u]\!]$ is finite for all $u \in U$,*
4. *two relations $\prec$ and $\preceq$ on U with $\prec \subseteq \preceq$. We denote by $\preceq^*$, (resp. $\prec^*$) the transitive closure of $\preceq$ (resp. $\prec$).*

We also assume that the following holds for Ω, $\prec$ and $\preceq$:

(Ω_0) *We assume two conditions: $\Omega(A) \subseteq A$, and "$A \neq \emptyset$ implies $\Omega(A) \neq \emptyset$",*
(Ω_1) *if $a \in A$, $[\![a]\!] \cap \Omega(A) = \emptyset$, $[\![b]\!] \cap \Omega(A) \neq \emptyset$, then there is $b' \in [\![b]\!] \cap A$, $b' \prec^* a$,*
(Ω_2) *if $a \in A$, $[\![a]\!] \cap \Omega(A) \neq \emptyset$, $[\![b]\!] \cap \Omega(A) \neq \emptyset$, then there is $b' \in [\![b]\!] \cap A$, $b' \preceq^* a$.*

In the first part (Proposition 2 (1)), we construct a ranked[2] order $\lhd$ on U by extending the relation $\prec$ (and $\preceq$), and show that $\Omega = \Omega_\lhd$, where $\Omega_\lhd$ is the minimality operation induced by $\lhd$: $\Omega_\lhd(X) := \{x \in X : \neg\exists x' \in X\, x' \lhd x\}$.

Proposition 2 (Constructing Ranked Orders).

(1) *If the relation $\preceq$ is free from cycles containing $\prec$, then $\prec$ can be extended to a ranked order $\lhd$ s.t. for all $A \subseteq U$ and $a \in A$:*

$$[\![a]\!] \cap \Omega(A) = \emptyset \quad \textit{if and only if} \quad [\![a]\!] \cap \Omega_\lhd(A) = \emptyset.$$

(2) *If, in addition, U is a set of abstract distances $d(\cdot,\cdot)$ over some space W, i.e. $U = \{d(x,y) : x, y \in W\}$ s.t., in addition to the conditions $\Omega_0, \Omega_1, \Omega_2$ the following holds:*

(d_1) $\forall x, y \in W$: *$x \neq y$ implies $d(x,x) \prec d(x,y)$,*
(d_2) $\forall x, y \in W$: $d(x,x) \preceq d(x,y)$

and the relation $\preceq$ is free from cycles containing $\prec$, then there is a totally ordered set $(Z, <)$ with a minimal element 0 and a distance function $\mathsf{dist} : W \times W \to Z$ s.t.

(a) $0 = \mathsf{dist}(x,x)$ *for any $x \in W$,*
(b) $d(u,v) \prec d(x,y) \to \mathsf{dist}(u,v) < \mathsf{dist}(x,y)$,
$d(u,v) \preceq d(x,y) \to \mathsf{dist}(u,v) \leq \mathsf{dist}(x,y)$,
(c) *for all $A \subseteq U$, $a \in A$: $[\![a]\!] \cap \Omega(A) = \emptyset$ if and only if $[\![a]\!] \cap \Omega_<(A) = \emptyset$.*

[2] $\lhd$ on U is called *ranked*, if, by definition, there exists a function $rank : U \to T$ from U to a strict total order $(T, <_T)$ such that $u \lhd u'$ if and only if $rank(u) <_T rank(u')$.