Explaining Updates by Minimal Sums ⋆

Karl Schlechta

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

Jürgen Dix [1]

The University of Manchester,
Dept. of Computer Science
Oxford Road, Manchester, M13 9PL, UK
dix@cs.man.ac.uk

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* 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 [2] is an important problem in Artificial Intelligence, as has been recognized very early. Much of human

⋆ Preliminary results of this paper appeared in the Proceedings of the 19th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'99), C. Pandu Rangan and V. Ramam and R. Ramanujan (editors), Madras, India, 1999. Lecture Notes in Computer Science 1738, pp. 142–152.

[1] Responsible Author: Please address all mail to me.

[2] The term *situation* is not to be confused with the same term in *situation calculus*.

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 ([2]). E.g. circumscriptive approaches try to minimize change by circumscribing certain predicates (see [16,10]). Default logics formalize persistency by stating special *default rules* ([9]). 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 [3,6,8]).

In this paper we generalize a particular approach introduced in [14,15] and [4]. 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 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: Th induces the set $\{\mathcal{A} : \mathcal{A} \models Th\} \subseteq W$.*

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

Definition 2 (Sequence $\boldsymbol{\sigma}$, 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 *inertia*. 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 [12].
(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 [13]. 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. 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 3 (Plausible Explanations) *Sequences that contain loops of the form $\langle \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_1\rangle$ and thus are unneccessary 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 4 (Preference Relation $\prec$) *A preference relation $\prec$ is any re-*

lation 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 [13,1], 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 [14,15]. The idea is to assume the notion of a distance between arbitrary worlds.

Definition 5 (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 [14,15] 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 6 (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*

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

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 [7], 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 prove in Section 3 a quite general fact, Proposition 13, which will be important for our overall solution of the problem. It is an abstract representation result (about constructing ranked orders) that can be instantiated to the situation we are considering. Our main results are contained in Section 4, where we (1) prove Proposition 14 (stating that the preferred sequences are completely determined by the endpoints of certain intermediate sequences), and (2) formulate our main Theorem 19 and state the conditions (Criteria 15, 18 and Definition 17) under which the equivalence holds. Finally, Section 5 contains the proof of Theorem 19. This proof consists mainly of an application of our results in Section 3 to the situation here and its combination with a variation of a result in [7]. We conclude with Section 6 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 7 ($\mathbf{\Pi_{fin}(2^W)}$) *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 $\Sigma_{\restriction i_0}$ the restriction of Σ to the first i_0 components.

If there is a distance dist or, more generally, a preference relation $\prec$ defined on W, we can determine the set of those sequences σ with $\sigma_i \in \Sigma_i$ for which $\mathsf{sum\text{-}dist}(\sigma)$ (see Definition 6) is minimal. We call such sequences dist-preferred

sequences. Analogously, 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 8 (dist- and $\prec$-Preferred Sequences) *Let a sequence $\Sigma := \langle \Sigma_1, \ldots, \Sigma_n \rangle$ of situations be given. We denote by*

$$\mathsf{Pref}_{\mathsf{dist}}(\Sigma) \quad (\textit{resp. } \mathsf{Pref}_{\prec}(\Sigma))$$

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) $\mathsf{sum\text{-}dist}(\sigma)$ 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 9 ($\mathsf{End}_{\prec}, \mathsf{End}_{\mathsf{dist}}$) *We define the following functions, depending on the underlying preference relation* dist *or $\prec$.*

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

[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. In fact, we will prove a theorem of the form

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

We would like to emphasize that our approach only assumes knowledge about $\mathsf{End}_{dist}(\langle \Sigma_1, \ldots, \Sigma_n \rangle)$, i. e. about the endpoints which are contained 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 14 (Section 4) that

knowledge of the endpoints of intermediate sequences allows us to completely reconstruct $\mathsf{End}_{dist}(\Sigma)$.

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

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

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 11 (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 An Abstract Representation Result

We start our formal exposition with an abstract approach, which has proved useful in many situations. The main result itself, Proposition 13, is neither conceptually nor technically deep, but it serves very well as a guideline to prove general representation theorems (in the finite case) for operations based on distances.

Informally, our result shows that for the existence of a *distance* dist, it suffices to show two properties (Ω_1) and (Ω_2). These two properties are sufficiently close to the operation considered to give an idea how to build the proof (or to see which properties one still has to add for completeness).

Definition 12 (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 13 (1)), we construct a ranked 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 $\lhd$, i.e.

$$\Omega_{\lhd}(X) := \{x \in X : \neg\exists x' \in X\, x' \lhd x\}.$$

In the second part (Proposition 13 (2)), we show the same for a suitably defined distance function and a total order $<$.

Before stating the main proposition of this section, we introduce two notions:

- An ordering $\lhd$ 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

$$u \lhd u' \text{ if and only if } rank(u) <_T rank(u').$$

- We say that $u \perp u'$ if, by definition, u is incomparable with u' with respect to the order $\lhd$.

Proposition 13 (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) $d(u,v) \prec d(x,y) \rightarrow \mathsf{dist}(u,v) < \mathsf{dist}(x,y)$,
$d(u,v) \preceq d(x,y) \rightarrow \mathsf{dist}(u,v) \leq \mathsf{dist}(x,y)$,
(c) for all $A \subseteq U$, $a \in A$: $[\![a]\!] \cap \Omega(A) = \emptyset$ if and only if $[\![a]\!] \cap \Omega_{<}(A) = \emptyset$.

The proof of this proposition uses two notes that are independent of it and will be shown at the end of this section.

PROOF. The proofs of (1) and (2) are very close, and have a common beginning.

Let $\prec^+$ and $\preceq^+$ be the closures of $\preceq$ under reflexivity and $\prec$ / $\preceq$ under transitivity, more precisely:

- $a \preceq^+ a$,
- $a \preceq b$ implies $a \preceq^+ b$, and $a \prec b$ implies $a \prec^+ b$,
- $a \preceq^+ b \preceq^+ c$ implies $a \preceq^+ c$,
- ($a \preceq^+ b \prec^+ c$ or $a \prec^+ b \preceq^+ c$ or $a \prec^+ b \prec^+ c$) implies ($a \prec^+ c$ and $a \preceq^+ c$).

We define the following relation on U^2:

$$a \approx b \quad \textbf{if and only if} \quad (a \preceq^+ b \text{ and } b \preceq^+ a).$$

$\approx$ is an equivalence relation. We denote the equivalence classe of a with $\tilde{a}$.

Furthermore, let

$$Z =_{\text{def}} \{\tilde{a} : a \in U\}.$$

We define $\prec$ on Z^2 by $\tilde{a} \prec \tilde{b}$ if and only if $a \preceq^+ b$, but $\tilde{a} \neq \tilde{b}$ (thus $b \not\preceq^+ a$). This is well-defined, and $\prec$ on Z is transitive and free of cycles too. (For the latter, e.g. $\tilde{a} \prec \tilde{b} \prec \tilde{a}$ implies $\tilde{a} = \tilde{b}$.)

We now turn to the two parts of the proof.

***(1)*:** We first extend $\prec$ on Z to a strict total order $<$ on Z. Then we define the relation $\lhd$ on U^2:

$$a \lhd b \text{ if, by definition, } \tilde{a} < \tilde{b}. \tag{2}$$

$\lhd$ is a ranked order on U (via $a \mapsto \tilde{a}$).
We have to show $[\![a]\!] \cap \Omega(A) = \emptyset$ if and only if $[\![a]\!] \cap \Omega_{\lhd}(A) = \emptyset$. Let $a \in A$.

(a) Let $[\![a]\!] \cap \Omega(A) = \emptyset$. By (Ω_0) there is a $b \in A$ with $[\![b]\!] \cap \Omega(A) \neq \emptyset$. By Ω_1, there is $b' \in [\![b]\!] \cap A$ and $b' \prec^* a$. By Note 1 we have $b' < a$ and $a \notin \Omega_<(A)$, therefore $[\![a]\!] \cap \Omega_{\lhd}(A) = \emptyset$.

(b) Suppose $[\![a]\!] \cap \Omega_{\lhd}(A) = \emptyset$, but $[\![a]\!] \cap \Omega(A) \neq \emptyset$. Choose $a' \in [\![a]\!] \cap A$ which is $\lhd$ -minimal in $[\![a]\!] \cap A$ (i.e. there is no $a'' \in [\![a]\!] \cap A$ with $a'' \lhd a'$). This is possible, as $[\![a]\!] \cap A$ is finite and $\lhd$ is free from cycles. As $[\![a]\!] \cap \Omega_{\lhd}(A) = \emptyset$, there is $b \in A$, $b \lhd a'$. If $[\![b]\!] \cap \Omega(A) = \emptyset$, then by (Ω_1) there is $a'' \in [\![a]\!] \cap A$, $a'' \prec^* b$, so by Note 1 $a'' \lhd b \lhd a'$, which is a contradiction.

If $[\![b]\!] \cap \Omega(A) \neq \emptyset$, then by (Ω_2) there is $a'' \in [\![a]\!] \cap A$, $a'' \preceq^* b$. By Note 1, $a'' \lhd b$ or $a'' \perp b$ or $a'' = b$. If $a'' \perp b$, then, using rankedness $a'' \lhd a'$. If $a'' \lhd b$, then $a'' \lhd a'$ by transitivity. This is a contradiction.

***(2)*:** As in Step *1* above, we extend $\prec$ on Z to a strict total order $<$ on Z and define

$$\mathsf{dist}(x, y) := \widetilde{d(x, y)}. \tag{3}$$

The rest of the proof for *(2)* is almost verbatim the same as the one for *(1)*: It remains to show $[\![a]\!] \cap \Omega(A) = \emptyset$ if and only if $[\![a]\!] \cap \Omega_<(A) = \emptyset$. Let $a \in A$.

(a) Let $[\![a]\!] \cap \Omega(A) = \emptyset$. By (Ω_0) there is a $b \in A$ with $[\![b]\!] \cap \Omega(A) \neq \emptyset$. By Ω_1, there is $b' \in [\![b]\!] \cap A$ and $b' \prec^* a$. Therefore (Note 2) we have $\mathsf{dist}(b') < \mathsf{dist}(a)$ and $a \notin \Omega_<(A)$, therefore $[\![a]\!] \cap \Omega_{\lhd}(A) = \emptyset$.

(b) Suppose $[\![a]\!] \cap \Omega_<(A) = \emptyset$, but $[\![a]\!] \cap \Omega(A) \neq \emptyset$. Choose $a' \in [\![a]\!] \cap A$ $< -$*minimal* in $[\![a]\!] \cap A$ (i.e. there is no $a'' \in [\![a]\!] \cap A$ $\mathsf{dist}(a'') < \mathsf{dist}(a')$). This is possible, as $[\![a]\!] \cap A$ is finite and $<$ is free from cycles. As $[\![a]\!] \cap \Omega_<(A) = \emptyset$, there is $b \in A$, $\mathsf{dist}(b) < \mathsf{dist}(a')$. If $[\![b]\!] \cap \Omega(A) = \emptyset$,

then by (Ω1) there is $a'' \in [\![a]\!] \cap A$, $a'' \prec^* b$, so $\mathsf{dist}(a'') < \mathsf{dist}(b) < \mathsf{dist}(a')$—this is a contradiction. If $[\![b]\!] \cap \Omega(A) \neq \emptyset$, then by ($\Omega$2) there is $a'' \in [\![a]\!] \cap A$, $a'' \preceq^* b$. So by Note 2, (b) $\mathsf{dist}(a'') \leq \mathsf{dist}(b)$, *contradiction.*

Note that we use in both cases for the representation result essentially Note 1 or Note 2 respectively, independent of the details of the construction of the order $\leq$ / $<$ from $\preceq$ / $\prec$. $\square$

Note 1 *Let $a, b \in W$ and let $\lhd$ as defined in Equation (2).*

(1) $a \prec^* b \rightarrow a \lhd b$,
(2) $a \preceq^* b \rightarrow a \lhd b$ *or* $a \perp b$ *or* $a = b$.

PROOF.

(1) $a \prec^* b \rightarrow a \preceq^+ b$, but not $b \preceq^+ a$ (otherwise there is a cycle involving $\prec$), so $\tilde{a} \prec \tilde{b}$, so $\tilde{a} < \tilde{b}$, so $a \lhd b$.
(2) $a \preceq^* b \rightarrow a \preceq^+ b$. If $b \preceq^+ a$ too, then $a \approx b$, and $a \perp b$ or $a = b$. Otherwise $\tilde{a} < \tilde{b}$, so $a \lhd b$. $\square$

Note 2 *Let $u, u', v, v' \in U$ and let* dist *as defined in Equation (3).*

(1) $d(u,v) \prec^* d(u',v')$ *implies* $\mathsf{dist}(u,v) < \mathsf{dist}(u',v')$,
(2) $d(u,v) \preceq^* d(u',v')$ *implies* $\mathsf{dist}(u,v) \leq \mathsf{dist}(u',v')$,
(3) The definition $0 =_{def} \mathsf{dist}(x,x)$ *for any $x \in W$ is well defined. In addition, the following holds:*
(a) $0 \leq \mathsf{dist}(x,y)$ *for any $x, y \in W$,*
(b) $0 < \mathsf{dist}(x,y)$ *if and only if $x \neq y \in W$.*

PROOF.

1. $d(u,v) \prec^* d(u',v')$ implies $d(u,v) \preceq^+ d(u',v')$, but not $d(u',v') \preceq^+ d(u,v)$ (otherwise there is a cycle involving $\prec$), so $\widetilde{d(u,v)} \prec \widetilde{d(u',v')}$, so $\widetilde{d(u,v)} < \widetilde{d(u',v')}$, so $\mathsf{dist}(u,v) < \mathsf{dist}(u',v')$.

2. $d(u,v) \preceq^* d(u',v')$ implies $d(u,v) \preceq^+ d(u',v')$. If $d(u',v') \preceq^+ \widetilde{d(u,v)}$ too, then $d(u,v) \approx d(u',v')$, and $\mathsf{dist}(u,v) = \mathsf{dist}(u',v')$. Otherwise $\widetilde{d(u,v)} < \widetilde{d(u',v')}$, so $\mathsf{dist}(u,v) < \mathsf{dist}(u',v')$.

3. 0 is well defined by (d_2) and *2.* $0 \leq \mathsf{dist}(x,y)$ holds by (d_1), (d_2), *1.*, *2.*. $0 < \mathsf{dist}(x,y)$ if and only if $x \neq y$ holds for the same reasons. $\square$

4 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 14 ($\mathsf{Pref}_{\mathsf{dist}}(\Sigma)$ Induced By Intermediate $\mathsf{End}_{dist}(\Sigma_{|i})$) *Let Σ be a sequence in the sense of Definition 8.*

$\mathsf{Pref}_{\mathsf{dist}}(\Sigma)$ *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 e.g. $|\mathsf{End}_{dist}(\Sigma_{|i-1})| = 1$, we apply Case 1 to $i-1$. So 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:

$$a_i \in \mathsf{End}_{dist}(\Sigma|_i)$$
$$a_{i-1} \in \mathsf{End}_{dist}(\Sigma[i/\{a_i\}]|_{i-1})$$
$$a_{i+1} \in \mathsf{End}_{dist}(\Sigma[i/\{a_i\}]|_{i+1})$$

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

Our next theorem is the main result of this paper. In Section 2 we have shown 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 (like in Definition 12). 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 15 (Important Conditions)

(C1) $\mathsf{End}_{\prec}(\Sigma) \subseteq \Sigma_n$, *if* Σ_n *is the last component of* Σ,
(C2) $\Sigma \neq \emptyset \to \mathsf{End}_{\prec}(\Sigma) \neq \emptyset$,
(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 [5] (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.

One of the important ingredients of Proposition 13 is the 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 16 (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 18.

Definition 17 (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 $\{a_\sigma + \ldots + d_{\sigma'}) : \sigma, \sigma'$ sequences $\}$. In the following we will formulate conditions to constrain the interaction between $+$ and $\prec$. (The terms a_σ, b_τ correspond to one sequence. When they are compared, they are of equal length. ($\overset{\leq}{\underset{>}{\equiv}}$ stands for $\preceq$ and $\succeq$ simultaneously.)*

(+1) $a_\sigma + a_\tau \overset{\leq}{\underset{>}{\equiv}} a_\tau + a_\sigma$,

(+2) $(a_\sigma + a_\tau) + a_\eta \mathrel{\substack{\leqq\\ \geqq}} a_\sigma + (a_\tau + a_\eta)$,
(+3) $a_\sigma \mathrel{\substack{\leqq\\ \geqq}} a_{\sigma'} \to ((a_\tau \mathrel{\substack{\leqq\\ \geqq}} a_{\tau'}) \leftrightarrow (a_\sigma + a_\tau) \mathrel{\substack{\leqq\\ \geqq}} (a_{\sigma'} + a_{\tau'}))$,
(+4) $a_\sigma \mathrel{\substack{\leqq\\ \geqq}} a_{\sigma'} \to (a_\tau \prec a_{\tau'} \leftrightarrow (a_\sigma + a_\tau \prec a_{\sigma'} + a_{\tau'})$,
(+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 18 (Loop) *The (smallest) relation defined by* (R1)–(R6), (+1)–(+5) *(see Definition 17) 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.

We are now ready to give a precise solution of our original problem.

Theorem 19 (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 the rationals* $\mathbb{Q}$*, 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 15 and 18.*

Corollary 20 *The previous theorem also holds if "into the rationals* $\mathbb{Q}$*" is replaced by "into an ordered abelian group".*

5 Proof of the Main Result

As the proof of Theorem 19 is quite involved, we first give a sketch of it.

5.1 Sketch of Proof

The proof is an application of Proposition 13. The direction from *(1)* to *(2)* is trivial. It remains to show that *(2)* implies *(1)*.

Assume a function $\mathsf{End}_\prec$ satisfying Criteria 15 and 18. 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 [7]. The crucial loop criterion is used to ensure that a solution exists. It then remains to show that $\mathsf{End}_\prec(\Sigma) = \mathsf{End}_{dist}(\Sigma)$.

5.2 Auxiliary Lemmas

What do the conditions stated Definitions 17 tell us? It is worth noting that in general, we cannot "observe" sums. By this we mean the following: if $x = \mathsf{dist}(a,b)$ and $y = \mathsf{dist}(c,e)$, it is not necessarily true that there exists a sequence σ with $\mathsf{sum\text{-}dist}(\sigma) = x + y$. In particular, it is not guaranteed that there is f with $\mathsf{dist}(b,f) = y$.

This is the reason why we pack the conditions *(+1)*–*(+5)* of Definition 17 into the relation and **(Loop)**, and do not use conditions like (for $\Sigma_n = \Sigma'_1$, $\sigma_n = \sigma'_1$) :

- If $\sigma \in \mathsf{End}_\prec(\Sigma)$, $\sigma' \in \mathsf{End}_\prec(\Sigma')$, then $\sigma\sigma' \in \mathsf{End}_\prec(\Sigma\Sigma')$, if $\sigma \in \mathsf{End}_\prec(\Sigma)$, $\sigma' \notin \mathsf{End}_\prec(\Sigma')$, then $\sigma\sigma' \notin \mathsf{End}_\prec(\Sigma\Sigma')$,
- if $\sigma \notin \mathsf{End}_\prec(\Sigma)$, $\sigma' \in \mathsf{End}_\prec(\Sigma')$, then $\sigma\sigma' \notin \mathsf{End}_\prec(\Sigma\Sigma')$, if $\sigma \notin \mathsf{End}_\prec(\Sigma)$, $\sigma' \notin \mathsf{End}_\prec(\Sigma')$, then $\sigma\sigma' \notin \mathsf{End}_\prec(\Sigma\Sigma')$.

Such conditions are much weaker, because they apply only to those sums which are really observable. We could, of course, stipulate a general condition of homogeneity of the space: This can be done by performing sufficient translations to guarantee concatenability. However, this would impose a restriction on the models we consider.

Lemmas 21–23 are auxiliary lemmas. The latter two show the essential prerequisites of the abstract representation result of Section 2 (Proposition 13).

The only thing still to show is the treatment of sums, which is done in the final proof at the end of this section.

Lemma 21

(1) $B' \subseteq B \rightarrow \Sigma \times B \preceq \Sigma \times B'$
(2) $b \in \mathsf{End}_{\prec}(\Sigma \times B) \rightarrow \Sigma \times b \preceq \Sigma \times B$
(3) $b \in B$, $b \notin \mathsf{End}_{\prec}(\Sigma \times B) \rightarrow (\Sigma \times B) \prec (\Sigma \times b)$
(4) If $b \in B$, $b \notin \mathsf{End}_{\prec}(\Sigma \times B)$, *then there is* $\sigma' \in \Sigma$ *s.t.* $\forall \Sigma' \subseteq \Sigma$ (Σ' *legal,* $\sigma' \in \Sigma' \rightarrow b \notin \mathsf{End}_{\prec}(\Sigma' \times B)$)
(5) If $b \in \mathsf{End}_{\prec}(\Sigma \times B)$, *then there is* $\sigma' \in \Sigma$ *s.t.* $\forall \Sigma' \subseteq \Sigma$ (Σ' *legal,* $\sigma' \in \Sigma' \rightarrow b \in \mathsf{End}_{\prec}(\Sigma' \times B)$).

PROOF.

(1) trivial by *(R1), (C1), (C2).*
(2) trivial by *(R1).*
(3) trivial by *(R2).*
(4) If not, then $\forall \sigma'$ there exists $\Sigma' \subseteq \Sigma$ (Σ' legal, $\sigma' \in \Sigma'$, $b \in \mathsf{End}_{\prec}(\Sigma' \times B)$). Let then $\sigma \in \Sigma$. For $\sigma' \neq \sigma$, $\sigma' \in \Sigma$, there is $\Sigma'_{\sigma'}$ with $\sigma' \in \Sigma'_{\sigma'} \subseteq \Sigma$, $b \in \mathsf{End}_{\prec}(\Sigma'_{\sigma'} \times B)$. Then $\Sigma = \{\sigma\} \cup \bigcup \{\Sigma'_{\sigma'} : \sigma' \neq \sigma\}$, but $b \notin \mathsf{End}_{\prec}(\Sigma \times B)$. Thus $\sigma \times B \prec_{(R5)} \Sigma'_{\sigma'} \times B \preceq_{(R3)} \sigma' \times B$ for all $\sigma' \neq \sigma$. Using the argument twice shows that $\prec$ contains a cycle.
(5) If not, then $\forall \sigma'$ there exists $\Sigma'_{\Sigma'} \subseteq \sigma$ ($\Sigma'_{\sigma'}$ legal, $\sigma' \in \Sigma'_{\sigma'}$, $b \notin \mathsf{End}_{\prec}(\Sigma'_{\sigma'} \times B)$), but $\Sigma = \bigcup \Sigma'_{\sigma'}$, so this contradicts *(C3).* □

Lemma 22 $b, b' \in \mathsf{End}_{\prec}(\Sigma \times B)$, $\sigma' \in \Sigma$ *implies* $\exists \sigma \in \Sigma\, (\sigma b \preceq^* \sigma' b')$.

PROOF.

If $b \in \mathsf{End}_{\prec}(\sigma' \times B)$, then $\sigma' b \preceq_{\text{Lemma 21 (2)}} \sigma' \times B \preceq_{\text{Lemma 21 (1)}} \sigma' b'$.

Suppose $b \notin \mathsf{End}_{\prec}(\sigma' \times B)$. By Lemma 21 (5) there is σ s.t. $\sigma \in \Sigma' \subseteq \Sigma$ implies $b \in \mathsf{End}_{\prec}(\Sigma' \times B)$. Thus the set of σ s.t. $b \in \mathsf{End}_{\prec}([\sigma', \sigma] \times B)$ is not empty. Choose such σ with minimal Hamming distance from σ'. Then $[\sigma', \sigma] = \bigcup \{[\sigma', \sigma\prime'] : \sigma'' \in [\sigma', \sigma], \sigma'' \neq \sigma\} \cup \{\sigma\}$. Moreover, for each $\sigma'' \in [\sigma', \sigma]$, $\sigma\prime\prime \neq \sigma$ $b \notin \mathsf{End}_{\prec}([\sigma', \sigma''] \times B)$. Thus, $\sigma \times B \preceq_{(R4)} [\sigma', \sigma'] \times B = \sigma' \times B \preceq_{\text{Lemma 21 (1)}} \sigma' b'$. As $b \in \mathsf{End}_{\prec}([\sigma', \sigma] \times B)$, there must be by Lemma 21 (5) $\sigma'' \in [\sigma', \sigma]$ s.t. $\forall \Sigma'' \subseteq [\sigma\prime, \sigma]$ ($\sigma'' \in \Sigma''$ implies $b \in \mathsf{End}_{\prec}(\Sigma'' \times B)$). Choice of σ shows that this σ'' can only be σ. Thus, in particular, $b \in \mathsf{End}_{\prec}(\sigma \times B)$. Thus $(\sigma, b) \preceq_{\text{Lemma 21 (2)}} (\sigma \times B)$. □

Lemma 23 *$b \in \mathsf{End}_\prec(\Sigma \times B)$, $b' \notin \mathsf{End}_\prec(\Sigma \times B)$, $\sigma' \in \Sigma$ implies $\exists \sigma \in \Sigma$: $(\sigma b \prec^* \sigma' b')$.*

PROOF.

(a) If $b \in \mathsf{End}_\prec(\sigma' \times B)$, $b' \notin \mathsf{End}_\prec(\sigma' \times B)$, then $\sigma' b \preceq_{\text{Lemma 21 (2)}} \sigma' \times B \prec_{\text{Lemma 21 (2)}} \sigma' b'$.

(b) If $b' \in \mathsf{End}_\prec(\sigma' \times B)$, then there is by Lemma 21 (4) $\sigma \in \Sigma$ s.t. $\sigma \in \Sigma' \subseteq \Sigma$ implies $b' \notin \mathsf{End}_\prec(\Sigma' \times B)$.

Thus, the set of $\sigma \in \Sigma$ s.t. $b' \notin \mathsf{End}_\prec([\sigma', \sigma] \times B)$ is not empty, let σ be such with minimal Hamming distance from σ'. Then, as in the proof of Lemma 22, $\sigma \times B \prec [\sigma', \sigma\prime] \times B$ by (R5) for any $\sigma'' \in [\sigma', \sigma]$, $\sigma'' \neq \sigma$, so $\sigma \times B \prec [\sigma', \sigma'] \times B = \sigma' \times B \preceq_{\text{Lemma 21 (1)}} \sigma' b'$. If $b \in \mathsf{End}_\prec(\sigma \times B)$, then $\sigma b \preceq_{\text{Lemma 21 (2)}} \sigma \times B$, and we are done. If $b \notin \mathsf{End}_\prec(\sigma \times B)$, then by an argument as above, using Lemma 21 (5), we find σ^+ with minimal Hamming distance from σ s.t. $b \in \mathsf{End}_\prec([\sigma, \sigma^+] \times B)$. As above, we see that $\sigma^+ \times B \preceq [\sigma, \sigma] \times B$, and as in the proof of Lemma 22, we see that $b \in \mathsf{End}_\prec(\sigma^+ \times B)$, so $\sigma^+ b \preceq_{\text{Lemma 21 (2)}} \sigma^+ \times B$. Thus, we have $\sigma^+ b \preceq \sigma^+ \times B \preceq \sigma \times B \prec \sigma' \times B \preceq \sigma' b'$.

(c) If $b, b' \notin \mathsf{End}_\prec(\sigma' \times B)$, then $\sigma' \times B \prec_{\text{Lemma 21 (3)}} \sigma' b'$. Choose as above σ with least Hamming distance from σ' s.t. $b \in \mathsf{End}_\prec([\sigma', \sigma] \times B)$. As above, we see $b \in \mathsf{End}_\prec(\sigma \times B)$, and $\sigma b \preceq_{\text{Lemma 21 (2)}} \sigma \times B \preceq \sigma' \times B \prec \sigma' b'$. □

5.3 Final Proof

Proof of the main Theorem 19. The direction from *(1)* to *(2)* is trivial. It remains to show that *(2)* implies *(1)*.

We consider the relations $\preceq$ / $\prec$ restricted to $\sigma' s$, i.e. we neglect the Σ''s.

Part 1: The algorithm below shows, by **Loop**, that the resulting system of inequalities has a solution, and, moreover, constructs this solution, with all $\mathsf{dist}(m, m')$ and thus all $\mathsf{sum\text{-}dist}(\sigma)$ in $\mathbb{Q}$. In particular, the original $\preceq$ / $\prec$ between $\sigma' s$ are respected by the assignment of values.

Part 2: It then remains to show that $\mathsf{End}_\prec(\Sigma) = \mathsf{End}_{\mathsf{dist}}(\Sigma)$ As before, we use Lemmas 21, 22, 23 and the strategy of the proof of Proposition 13.

Note that by *(R6)* the system of inequalities contains $0 \prec \mathsf{dist}(a, b)$ for all $a \neq b$. This can be seen by applying *(R6)* to the set $\{a\} \times \{a, b\}$ (thus $\mathsf{dist}(a, a) < \mathsf{dist}(a, b)$ and, analogously, $\mathsf{dist}(b, b) < \mathsf{dist}(b, a)$) and to the set $\{a, b\} \times \{a, b\}$ (thus $\mathsf{dist}(a, a) = \mathsf{dist}(b, b)$).

Part 1

The following algorithm is a modification of an algorithm communicated by S. Koppelberg, Berlin. The original algorithm seems to be due to [7].

We have a system of inequalities and equalities of the types

$$x_{1,1} + \ldots + x_{1,m} \prec x_{2,1} + \ldots + x_{2,m}$$

or

$$x_{1,1} + \ldots + x_{1,m} \preceq x_{2,1} + \ldots + x_{2,m}$$

or

$$x_{1,1} + \ldots + x_{1,m} \leqq\!\!\!\!\!\!= x_{2,1} + \ldots + x_{2,m},$$

where m can differ. The sum $x_{1,1} + \ldots + x_{1,m}$ corresponds to one term $\delta(\sigma)$. Each $x_{i,j}$ stands for two neighboring worlds in a sequence, thus $x_{i,j} = \mathsf{dist}(a, b)$.

The last one can be transformed to

$$x_{1,1} + \ldots + x_{1,m} \preceq x_{2,1} + \ldots + x_{2,m}$$

and

$$x_{2,1} + \ldots + x_{2,m} \preceq x_{1,1} + \ldots + x_{1,m}.$$

As we can determine whether $x_{i,k}$ is 0 ($x_{i,k} = \mathsf{dist}(a, a)$ for some a in some sequence σ), we can denote these $x_{i,k}$'s by 0.

Let the remaining system contain $x_1, \ldots, x_n$. We eliminate by induction all but one of the $x_{i,k}$. The procedure will be successful (by the **Loop** condition), and tells us how to assign positive rationals to the $x_{i,k}$.

Assume without loss of generality that the left hand side is always less or equal the right hand side.

The procedure eliminates x_n by induction, and the simplified system of inequalities $\mathcal{S}'$ has a solution if and only if the original one $\mathcal{S}$ has.

Without loss of generality, x_n does not occur on both sides of the same inequality (otherwise, subtract one each on both sides repeatedly—this is justified by (+3) and (+4)).

Case 1: x_n does not occur in $\mathcal{S}$—we are done.

Case 2: x_n occurs only on the right hand side. Let $\mathcal{S}' \subseteq \mathcal{S}$ be the set of those inequalities, where x_n does not occur. If $\mathcal{S}'$ has a solution, choose x_n big enough to make $\mathcal{S}$ true.

Case 3: x_n occurs only on the left hand side. Then replace $x_m + x_n \preceq x_k + x_l$ by $x_m + 0 \prec x_k + x_l$, replace $x_m + x_n \prec x_k + x_l$ by $x_m + 0 \prec x_k + x_l$, and let the other inequalities unchanged. Let this modified system $\mathcal{S}'$ have a solution. Then the difference in the modified inequalities is at least some minimum, where we can put x_n in.

Case 4: x_n occurs on both the left and the right hand sides. Let $\mathcal{S}_l$ be the set of inequalities, where x_n occurs on the left hand side, let $\mathcal{S}_r$ be the set of inequalities, where x_n occurs on the right hand side.

Informally, we isolate x_n and transform all $\mathsf{sum\text{-}dist}()_i \in \mathcal{S}_l$ into $x_n \preceq R$ or $x_n \prec R$, and all $\mathsf{sum\text{-}dist}()_j \in \mathcal{S}_r$ into $L \preceq x_n$ or $L \prec x_n$, e.g. $x_3 + x_4 \preceq x_n + x_5$ will become $x_3 + x_4 - x_5 \preceq x_n$. We then consider all inequalities of the form $L \preceq R$ or $L \prec R$ resulting from $L \preceq x_n \preceq R$ etc., and *squeeze* x_n into a solution of the system of $L \preceq R$ and $L \prec R$. As $0 \prec x_n$ is among the original inequalities, we can find a positive solution for x_n. In general, this procedure will use subtraction, which is not observable and does not figure among the conditions $(+i)$. So, instead we consider the sums $\mathsf{sum\text{-}dist}()_i + \mathsf{sum\text{-}dist}()_j$ where $\mathsf{sum\text{-}dist}()_i \in \mathcal{S}_l$, $\mathsf{sum\text{-}dist}()_j \in \mathcal{S}_r$, and eliminate x_n from both sides. These are *legal* operations covered by the $(+i)$. We then solve this system, have numbers, and *squeeze* x_n into the inequalities.

Let $\mathcal{S}'$ be the set of inequalities where x_n does not occur, and all sums $\mathsf{sum\text{-}dist}()_i + \mathsf{sum\text{-}dist}()_j$, $\mathsf{sum\text{-}dist}()_i \in \mathcal{S}_l$, $\mathsf{sum\text{-}dist}()_j \in \mathcal{S}_r$.

For instance, for

$$x_n \preceq x_1,\ x_n \preceq x_2,\ x_3 \prec x_n,\ x_3 + x_4 \preceq x_n + x_5$$

we consider

$$\begin{aligned} &x_n + x_3 \prec x_1 + x_n, \\ &x_n + x_3 + x_4 \preceq x_1 + x_n + x_5, \\ &x_n + x_3 \prec x_2 + x_n, \\ &x_n + x_3 + x_4 \preceq x_2 + x_n + x_5. \end{aligned}$$

This is justified by $(+3) - (+5)$. Now we can eliminate x_n on both sides, where it still occurs. This is justified by $(+3)$ and $(+4)$.

In our example, $x_3 \prec x_1$, $x_3 + x_4 \preceq x_1 + x_5$, $x_3 \prec x_2$, $x_3 + x_4 \preceq x_2 + x_5$.

Let this $\mathcal{S}'$ have a solution. Then $x_3 \prec x_1$, $x_3 + x_4 - x_5 \preceq x_1$, $x_3 \prec x_2$,

$x_3 + x_4 - x_5 \preceq x_2$. We have to fit x_n into $[max(x_3 + x_4 - x_5, x_3), min(x_1, x_2)]$.

Finally, we have

$$0 + \ldots + 0 + x_1 + \ldots + x_1 \prec 0 + \ldots + 0,$$

namely sums of equal length on both sides, or

$$0 + \ldots + 0 + x_1 + \ldots + x_1 \preceq 0 + .. + 0,$$

or

$$0 + \ldots + 0 \prec 0 + \ldots + 0 + x_1 + \ldots + x_1,$$

or

$$0 + \ldots + 0 \preceq 0 + \ldots + 0 + x_1 + .. + x_1.$$

But the transformations we applied to go from $\mathcal{S}$ to $\mathcal{S}'$ were legal, covered by the conditions $(+i)$ in Definition 17, and thus preserved freedom from **Loop**, and by $0 \prec x_1$, the first two possibilities lead to a cycle—which was excluded. The latter two show that the final system has a solution, which can be transformed into one for the original system as indicated.

So the algorithm defines a distance compatibel with $+$. 0 does what it should.

Part 2
It remains to show that the distance represents $\mathsf{End}_\prec$, i.e. $\mathsf{End}_\prec(\Sigma) = \mathsf{End}_{\mathsf{dist}}(\Sigma)$.

For better readability, we separate the last component from Σ.

Let $b \in \mathsf{End}_\prec(\Sigma \times B)$. This implies, by Lemma 22, $\forall \sigma' b' \exists \sigma . \sigma b \preceq \sigma' b'$ which implies $\forall \sigma' b' \exists \sigma . \mathsf{sum\text{-}dist}(\sigma b) \leq \mathsf{sum\text{-}dist}(\sigma' b')$ which implies, by finiteness, $b \in \mathsf{End}_{\mathsf{dist}}(\Sigma \times B)$. Let $b \notin \mathsf{End}_\prec(\Sigma \times B)$, $\sigma \in \Sigma$. So there are (by $\mathsf{End}_\prec(\Sigma \times B) \neq \emptyset$ and Lemma 23) $b' \in \mathsf{End}_\prec(\Sigma \times B)$, $\sigma' \in \Sigma$ s.t. $\sigma' b' \prec \sigma b$, so $\mathsf{sum\text{-}dist}(\sigma' b') < \mathsf{sum\text{-}dist}(\sigma b)$, so $b \notin \mathsf{End}_{\mathsf{dist}}(\Sigma \times B)$. $\square$

6 Conclusion

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 [13] and [4].
(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 [1].
(3) In the present paper, we push the idea of inertia further, minimizing the sum of changes involved in a sequence of worlds ([14,15]).

We have shown in this paper the exact relationship between these approaches. We developed a general representation result in the spirit of [11], Theorem 19, 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.

Acknowledgements

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 of FSTTCS '99 read the paper extremely carefully and his suggestions were greatly appreciated.

References

[1] Shai Berger, Daniel Lehmann, and Karl Schlechta. Preferred history semantics for iterated updates. *Journal of Logic and Computation*, 9(6):817–833, 1999.

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

[3] 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., 2001. Shortened version also appeared in Dix, Pereira, Przymusinski (Eds.), *Logic Programming and Knowledge Representation*, Springer LNAI 1471, pages 1–55, 1998.

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

[5] Menachem Magidor Daniel Lehmann and Karl Schlechta. Distance semantics for belief revision. *Journal of Symbolic Logic*, to appear, 2001.

[6] Jürgen Dix, Ulrich Furbach, and Ilkka Niemelä. Nonmonotonic Reasoning: Towards Efficient Calculi and Implementations. In Andrei Voronkov and Alan Robinson, editors, *Handbook of Automated Reasoning*. Elsevier-Science-Press, 2001.

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

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

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

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

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

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

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

[14] Marianne Winslett. Reasoning about action using a possible models approach. In *Proc. 7th National Conference on Artificial Intelligence (AAAI-88)*, pages 89–93, 1988.

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

[16] 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)*, volume 1, pages 718–723. Morgan Kaufmann, 1993.