Distance semantics for Belief Revision *

Karl Schlechta† Daniel Lehmann‡ Menachem Magidor§

Abstract

A vast and interesting family of natural semantics for Belief Revision is defined. Suppose one is given a distance d between any two models. One may define the revision of a theory K by a formula a as the theory defined by the set of all those models of a that are closest, by d, to the set of models of K. This family is characterized by a set of rationality postulates that extends the AGM postulates. The new postulates describe properties of iterated revisions.

1 Introduction

Intelligent agents must gather information about the world, elaborate theories about it and revise those theories in view of new information that, sometimes, contradicts the beliefs previously held. Belief revision is therefore a central topic in Knowledge Representation. It has been studied in different forms: numeric or symbolic, procedural or declarative, logical or probabilistic. One of the most successful framework in which belief revision has been studied has been proposed by Alchourrón, Gärdenfors and

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

†Université de Provence, 3, Place Victor Hugo, F-13331 Marseille Cédex 3, France, ks@cmi.univ-mrs,fr

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

§Institute of Mathematics, Hebrew University, 91904 Jerusalem, Israel, menachem@sunrise.ma,huji.ac.il

Makinson, and is known as the AGM framework. It deals with operations of revision that revise a theory (the set of previous beliefs) by a formula (the new information). It proposes a set of rationality postulates that any reasonable revision should satisfy. A large number of researchers in AI have been attracted by and have developed this approach further: both in the abstract and by devising revision procedures that satisfy the AGM rationality postulates. Even though iterated revisions are a central concern in AI applications, those postulates do not mention iterated revisions explicitly. A closer look shows that the AGM postulates do not say anything on the way a revision may depend on its first argument, the theory being revised. If one considers iterated revisions, shouldn't one put some requirement on this dependence? Many authors ([3, 2, 4, 12, 11, 9]) have recently expressed interest in iterated revisions and proposed additional postulates about them. The last of the references above has proposed an alternative, more liberal, syntax. The present work is couched in the traditional AGM framework, but the main ideas may be useful in the framework proposed in [9]: further work on semantics based on a notion of distance for updates is in progress.
The AGM framework, defined in [1], studies revision operations, denoted $*$, that operate on two arguments: a theory (any set of formulas closed under logical deduction) K on the left and a formula a on the right. Thus $K{*}a$ is the result of revising theory K by formula a, using revision method $*$. The original AGM rationality postulates are the following. We use the customary notation $Cn(X)$ for the set of all logical consequences of a set X of formulas. The logical connectives always have precedence over the $*$ operation.

K*1 $K{*}a$ is a theory.

K*2 $a \in K{*}a$.

K*3 $K{*}a \subseteq Cn(K, a)$.

K*4 If $\neg a \notin K$, then $Cn(K, a) \subseteq K{*}a$.

K*5 If $K{*}a$ is inconsistent,

then a is a logical contradiction.

K*6 If $\models a \leftrightarrow b$, then $K{*}a = K{*}b$.

K*7 $K * a \wedge b \subseteq Cn(K{*}a, b)$.

K*8 If $\neg b \notin K{*}a$, then $Cn(K{*}a, b) \subseteq K * a \wedge b$

2 Distances

The AGM revisions, the study of which was initiated in [1], have not, to this day, received a satisfactory semantic treatment. The main result about those revisions, their very close connection with the rational relations of [10], showed in [6], cannot be considered as suitable semantics. The purpose of this paper is to propose such semantics.
Let us assume that a language $\mathcal{L}$ is given, and let $\mathcal{M}$ be the corresponding set of models. We shall assume that, between any two models m and n, a distance $d(m,n)$ is defined. More precisely, we assume a set $\mathcal{D}$ of distances is given. We shall not say much about this set (but the reader may assume it is the set of real numbers greater or equal to zero), except that it is a totally ordered set ($<$ is the ordering), and that it has a smallest element denoted 0. For any two models m and n, $d(m,n)$ is an element of $\mathcal{D}$. The assumption that $\mathcal{D}$ is totally ordered is similar to the assumption, in economics, that utilities are totally ordered. Our assumptions about d are the following.

- **(A0)** $d(m,n) = 0$ iff $m = n$
- **(Symmetry)** d(m , n) = d(n , m)
- **(Limit assumption)** Given any two non-empty sets of models A and B, there are models $m \in A$ and $n \in B$ such that, for any models $m' \in A$ and $n' \in B$, one has $d(m,n) \leq d(m',n')$.

Assumption A0 expresses the fact that $d(m,m)$ is zero, and if $m \neq n$, then $d(m,n) > 0$. The Limit Assumption assumes the distance between any two sets, defined as an infimum, is in fact a minimum. If the set $\mathcal{M}$ of models is finite, the Limit Assumption is certainly satisfied. Note that we do require the function d to be symmetric. It turns out that we could as well weaken this requirement to:

(Weak symmetry) $d(m,n) < d(s,t) \Rightarrow d(n,m) < d(t,s)$.

The representation result will show that we could also have required the distance d to satisfy the triangular inequality. We do not know, at the moment, how to treat the case of a pseudo-distance, that is a d that satisfies A0 and the Limit assumption but does not necessarily satisfy Weak symmetry. We

do not know either how to treat the case of an ultrametric distance, i.e., satisfying $d(m,n) \leq \max(d(m,t), d(t,n))$.
Intuitively, the distance from m to n, $d(m,n)$ represents the "cost" or the "difficulty" of a change from the situation represented by m to the situation represented by n. M. Winslett has considered one such distance: the distance between two propositional worlds is the number of atomic propositions on which they differ. Another example of such a distance is the trivial distance: $d(m,n)$ is zero if $m = n$ and one otherwise. Both those distances satisfy the triangular inequality. In applications dealing with reasoning about actions and change, one may want to consider the distance between two models to represent how difficult, or unexpected, the transition is. In such a case, a natural d may well not be symmetric. Further work is needed to characterize such revisions.
By the limit assumption, one may define the distance between any two non-empty sets of models: $d(A,B)$ is the minimal $d(m,n)$ for $m \in A$ and $n \in B$. We shall demote $d(\{m\},B)$ and $d(A,\{n\})$ by $d(m,B)$ and $d(A,n)$ respectively. Let us now define a revision $*$, based on a distance d. A theory K defines a set B of models: the set of all models of K. Similarly a formula a defines the set A of all models that satisfy a. Let A and B be non-empty sets of models and $B|A$ be defined as:

$$B|A = \{m \in A \mid d(B,m) = d(B,A)\}.$$

The revision defined by distance d is:

$$K{*}a = \{b \in \mathcal{L} \mid \forall m \in B|A,\ m \models b\}.$$

In other words, the revision of K by a is the theory defined by the set $B|A$ of models. The definition above takes care of the case K and a are consistent (separately). In all other cases define $K{*}a$ to be $\mathcal{C}n(a)$.
The main goal of this work is to characterize the properties, i.e., rationality postulates satisfied by revisions defined by distances. A first easy result is: any such revision satisfies the AGM postulates **K∗1**–**K∗8**. For this result, the symmetry property is not needed: revisions defined by pseudo-distances satisfy the AGM postulates. But any revision defined by a pseudo-distance also satisfies some properties that do not follow from the AGM postulates. Consider, for example, the set $C = (B_1 \cup B_2)|A$. If $d(B_1,A) < d(B_2,A)$, then $C = B_1|A$. If $d(B_2,A) < d(B_1,A)$, we have $C = B_2|A$. If $d(B_1,A) = d(B_2,A)$,

then we have $C = (B_1|A) \cup (B_2|A)$. It follows that any revision defined by a pseudo-distance satisfies: $(K_1 \cap K_2)*a$ is equal to $K_1*a \cap K_2*a$, to K_1*a, or to K_2*a. This property does not follow from the AGM postulates, but seems a very natural property. When revising a disjunction $K_1 \vee K_2$ by a formula a, there are two possibilities. First, it may be the case that our indecision concerning K_1 or K_2 persists after the revision, and, in this case, the revised theory is naturally the disjunction of the revisions. But it may also be the case that the new information a makes us *revise backwards* and conclude that it must be the case that K_1 or, respectively, K_2 was (before the new information) the better theory and, in this case, the revised theory should be K_1*a or K_2*a.
Notice that, unrelated to the discussion above, it follows from the AGM postulates that $K*(a_1 \vee a_2)$ is equal to $K*a_1 \cap K*a_2$, to $K*a_1$ or to $K*a_2$. This property concerns a disjunction in the right-hand argument. The property described above and satisfied by revisions generated by distances is an analogue for the left-hand argument.
One can conclude that any revision defined by a pseudo-distance satisfies the following properties, that deal with iterated revisions:

$$\text{if } d \in K*a*c \text{ and } d \in K*b*c, \text{ then } d \in K*(a \vee b)*c$$

and

$$\text{if } d \in K*(a \vee b)*c, \text{ then , either } d \in K*a*c \text{ or } d \in K*b*c.$$

Those properties seem intuitively *right*. If after any one of two sequences of revisions that differ only at step i (step i being a in one case and b in the other), one would conclude that d holds, then one should conclude d after the sequence of revisions that differ from the two revisions only in that step i is a revision by the disjunction $a \vee b$, since knowing which of a or b is true cannot be crucial. This property is an analogue for the left argument of the Or property of [8]. Similarly, if one concludes d from a revision by a disjunction, one should conclude it from at least one of the disjuncts. This property is an analogue for the left argument of the Disjunctive Rationality property of [8], studied in [5]. It is easy to see that the property (C1) of Darwiche and Pearl [4], i.e., $K*a*(a \wedge b) = K*(a \wedge b)$ is not satisfied by all revisions defined by distances. The next section will precisely characterize those revisions that are defined by distances.

3 Representation result

We shall assume, for now, that the language $\mathcal{L}$ is a propositional language on a finite number of atomic propositions. Each model may be characterized by a complete formula, a formula it is the only model to satisfy.
This assumption is a natural one if one supposes that the theories to be revised are finitely generated, i.e., equivalent to a formula, and it has been accepted by many researchers, following [7]. Let us notice that, if one is not willing to make the assumption that theories are finitely generated, there no reason to suppose either that one revises by formulas, and one should probably consider revising arbitrary theories by arbitrary theories.
Our main result is a characterization of revisions defined by distances. We shall concentrate here on quasi-semantic properties to characterize those revisions. Further work needs to be done to provide a presentation more elegant and more in line with the AGM style. The first postulate needed deals with the limit cases of an inconsistent K or an inconsistent a:

(P0) If K or a is inconsistent then $K{*}a = \mathcal{C}n(a)$.

We shall, from now on, assume both K and a are consistent. The AGM postulates **K∗1**–**K∗6** are part of our characterization. The main auxiliary notion to express those postulates and to prove the representation result is the notion of a critical pair. A pair of models (m, n) is said to be critical for theory K and formula a iff:

1. $\neg\chi_n \notin K{*}a$
2. $\neg\chi_m \notin \mathcal{C}n(\chi_n){*}\chi_K$.

Above, χ_m denotes the characteristic formula of model m and χ_K the characteristic formula of theory K, i.e., the conjunction of all formulas of K. Intuitively the pair (m, n) is critical for K and a iff the distance between them is the distance between the sets of worlds that satisfy K and a respectively.
Notice that the postulates introduced above imply that, if (m, n) is a critical pair for K and a: m is a model of K and n is a model of a. They also imply that: $b \in K{*}a$ iff for all critical pairs (m, n) for K, a, $n \models b$.
The symmetry property of distances translates into an additional postulate:

(Sym) (m, n) is a critical pair for K, a

iff (n, m) is a critical pair for $\mathcal{C}n(a), \chi_K$.

The main postulate is formally similar to the Loop property of [8]: it states that it is never the case that $d(m, n) < d(m, n)$.
Let us define a binary relation, R between pairs of complete formulas. We shall identify a complete formula with the model it satisfies and with the theory of that model. We say that $(m, n)R(s, t)$ iff there are a theory K and a formula a such that, $s \models K$, $t \models a$ and (m, n) is a critical pair for K, a.
Note that $(m, n)R(s, t)$ implies that the distance between m and n is less or equal to the distance between s and t.
Our postulate is:

(L) if $(m_0, n_0)R(m_1, n_1)R \ldots R(m_n, n_n)R(m_0, n_0)$,

then for any K, a such that $m_0 \models K, m_1 \models K, n_0 \models a, n_1 \models a$

(m_0, n_0) is critical for K, a iff (m_1, n_1) is.

The main result of this work is the following representation theorem: a revision $*$ is defined by some distance d iff it satisfies **(P0)**, **K∗1–K∗6**, **(Sym)** and **(L)**.
The proof considers the relation $<$ defined by: $(m, n) < (s, t)$ iff $(m, n)R(s, t)$ and $(s, t)\not R(m, n)$. The transitive closure of $<$ relation is irreflexive, and is therefore a strict partial order. It may be extended to a total order. This total order is used as the set $\mathcal{D}$. Since this total order may obviously be embedded in the real interval $[0, 1]$ is such a way that non-zero distances are at least equal to $\frac{1}{2}$, one may ensure the triangular inequality is satisfied.

4 Further work

It follows from the representation result that any revision that satisfies the postulates also satisfies **K∗7**, **K∗8** and the properties discussed in the last part of Section 2. A direct proof would enable us to understand better the meaning of postulates **(Sym)** and **(L)**, and perhaps enable us to find an equivalent set of postulates that are more intuitive.

It should be noticed that if (m, n) is critical for K, a and $s \models K$, $t \models a$, then (m, n) is critical for $Cn(\chi_m \vee \chi_s)$, $\chi_n \vee \chi_t$ and $s \models Cn(\chi_m \vee \chi_s)$, $t \models \chi_n \vee \chi_t$. This means that, for the definition of R one may consider only theories K and formulas a that have at most two models.
Since we have seen that, in many cases, the natural (pseudo) distance between models is not symmetric, the most important open question is probably the characterization of revisions defined by pseudo-distances. The difficulty seems to lie in the characterization of critical pairs.
In a different direction, one may try to characterize revisions defined by ultra-metric distances.

References

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

[2] Craig Boutilier. Revision sequences and nested conditionals. In Ruzena Bajcsy, editor, *Proceedings of the 13th I.J.C.A.I.*, pages 519–525. Morgan Kaufmann, Chambéry, Savoie, France, August 1993.

[3] Craig Boutilier and Moisés Goldszmidt. Revision by conditional beliefs. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI)*, pages 649–654. Morgan Kaufmann, Washington, D.C., July 1993.

[4] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. In Ronald Fagin, editor, *Proceedings of the fifth Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 5–23. Morgan Kaufmann, Pacific Grove, CA, March 1994.

[5] Michael Freund. Injective models and disjunctive relations. *Journal of Logic and Computation*, 3(3):231–247, 1993.

[6] Peter Gärdenfors and David Makinson. Nonmonotonic inference based on expectations. *Artificial Intelligence*, 65(1):197–245, January 1994.

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

[8] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1–2):167–207, July 1990.

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

[10] Daniel Lehmann and Menachem Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, May 1992.

[11] Abhaya C. Nayak, Norman Y. Foo, Maurice Pagnucco, and Abdul Sattar. Changing conditional beliefs unconditionally. private communication, 1995.

[12] Mary-Anne Williams. Transmutations of knowledge systems. In Jon Doyle, Erik Sandewall, and Pietro Torasso, editors, *Proceedings of the fourth International Conference on Principles of Knowledge Representation and Reasoning*, pages 619–629. Morgan Kaufmann, Bonn, Germany, May 1994.