

Theory Revision and Probability

KARL SCHLECHTA*

Abstract The problem of Theory Revision is to “add” a formula to a theory, while preserving consistency and making only minimal changes to the original theory. A natural way to uniquely determine the process is by imposing an order of “epistemic entrenchment” on the formulas, as done by Gärdenfors and Makinson. We improve their results as follows: We define orders which generate unique revision processes too, but in addition, (1) have nice logical properties, (2) are independent of the theory considered, and are thus well suited for iterated revision and computational purposes, and (3) have a natural probabilistic construction. Finally, we show that the completeness problems of Theory Revision carry over to a certain extent to an approach based on revising axiom systems.

1 Introduction Recent years have seen an increasing interest in theory revision, which has partly centered around the work of Gärdenfors and his co-authors ([1], [2], [4–6]). In [4] and [5], the problem of choice in theory revision (more precisely: maxichoice contraction/revision) finds a natural solution in the concept of epistemic entrenchment: An order on the formulas tells us which to choose. The orders of [4] and [5] have, however, theoretical as well as computational drawbacks: (1) They are tailored to fit the requirements of theory revision and will not respect other natural demands like $\phi \leq \psi$ iff $\neg\psi \leq \neg\phi$ (if I tend to believe more in ψ than in ϕ , then I might tend to believe more in $\neg\phi$ than in $\neg\psi$); (2) For each revision of a new theory (in a fixed language) we have to find a new order, so iterated revision means iterated effort in ordering. Note that “theory” is used here in a technical sense: a deductively closed set of formulas (the deductive closure of a database). Thus two theories might be closely related,

*I would like to thank S. Koppelberg for correcting a proof, and S. Pribbenow for discussions, P. Gärdenfors for his suggestions, and D. Makinson for suggestions and discussions and the simplified version of Proposition 2.3.

and are not necessarily as different as, say, a physical and a medical theory. So iterated theory revision in this sense is a very common phenomenon for cognitive systems.

The main aim of this paper is to show how to overcome these drawbacks: (1) To prove that one order of a very simple kind will do for all theories of a given language and thus for all revisions (Propositions 2.2 and 2.4); and (2) To show how to construct such an order with particularly nice and natural properties for countable propositional languages (Proposition 2.8 and 2.11).

For a different treatment of iterated revision, see Spohn's work, e.g., [7].

The last section is independent of the first part. In Sections 1 and 2, we consider what Gärdenfors and his co-authors call maxichoice contraction—choosing a maximal subset K' of K , such that $K' \not\vdash A$. As pointed out in [1], [4], and [6], maxichoice contraction suffers from a completeness problem. Our point in Section 3 is that Theory Revision with underlying axiom sets is plagued by essentially the same problems (and some more). We consider systems $\langle K, A \rangle$, where K is a deductively closed set of formulas, and A is a set of axioms for K . Theory Revision for $\langle K, A \rangle$ will essentially amount to the choice of a suitable subset of A . Proposition 3.1 shows that there is a continuum between too coarse axiom sets (and too coarse revision) and too fine-grained axiom sets (resulting in full completeness at revision).

For the convenience of the reader, we now repeat the (for our purposes) main definitions and results of Gärdenfors and Makinson. But before that, we give an example to point out the basic problem of underdeterminacy.

A problem of theory revision Let T be a theory, i.e. a deductively closed set of formulas. Suppose $\{A, B\} \subseteq T$, thus $A \wedge B \in T$, and we would like to revise T to a maximal theory $T' \subseteq T$ such that $A \wedge B \notin T'$. So $\{A, B\} \subseteq T$ is impossible, and we have to withdraw A, B , or both. Leaving aside such extreme cases as $\vdash A \leftrightarrow B$, “both” is unsatisfactory, as T' should be maximal. So we must choose A or B , but logic won't tell us which. If we have an order $A < B$ telling us that we like A less than B , we are finished.

This is the idea of Gärdenfors' and Makinson's Solution. In the following, we adopt Gärdenfors' and Makinson's terminology to make this article more readable for those familiar with their work. They denote by “theory contraction” the process of removing a formula from a theory, and by “theory revision” adding a formula A to a theory T so that the resulting theory T' is consistent (if A is) and $A \in T'$. This is made precise in the following.

Definition 1.1 Given a language \mathcal{L} , an inference rule \vdash (we will not be more specific here, and the interested reader is referred, e.g., to [4]), and a “knowledge set” K , i.e. a set of formulas of \mathcal{L} closed under \vdash , then

a function $K-$: Formulas of $\mathcal{L} \rightarrow$ Sets of Formulas of \mathcal{L} is called a contraction function for K , iff it satisfies the postulates (K-1) to (K-8) below, and

a function K^* : Formulas of $\mathcal{L} \rightarrow$ Sets of Formulas of \mathcal{L} is called a revision function for K , iff it satisfies the postulates (K*1) to (K*8) below.

Proposition 1.1 Both notions are interdefinable by the following equations:

$$K^*A := (K - \neg A) + A \quad (\text{where } L + B \text{ is the deductive closure of } L \cup \{B\})$$

i.e., if $K-$ is a contraction function, then K^* so defined is a revision function, and

$$K - A := K \cap (K^* \neg A)$$

i.e., if K^* is a revision function, then $K-$ so defined will be a contraction function.

The proofs are straightforward.

We now state the axioms for $K-$ and K^* . Some very short comments are given at the right. The reader will find more motivation, e.g., in [4]:

Definition 1.2

(K-1) $K - A$ is a knowledge set (i.e. deductively closed under \vdash)

(K-2) $K - A \subseteq K$

(K-3) If $A \notin K$, then $K - A = K$ (the desired result already applies to K)

(K-4) If $\not\vdash A$, then $A \notin K - A$ (success, if possible)

(K-5) $K \subseteq (K - A) + A$ (where $L + B$ is the deductive closure of $L \cup \{B\}$, the “postulate of recovery”)

(K-6) If $\vdash A \leftrightarrow B$, then $K - A = K - B$

(K-7) $(K - A) \cap (K - B) \subseteq K - (A \wedge B)$ (a condition of minimality)

(K-8) If $A \notin K - (A \wedge B)$,
then $K - (A \wedge B) \subseteq K - A$ (In general, the more specific a formula is, the less the change necessary for revision. If $A \notin K - (A \wedge B)$, however, then contraction by $A \wedge B$ will do already)

and

(K*1) K^*A is a knowledge set

(K*2) $A \in K^*A$ (success)

(K*3) $K^*A \subseteq K + A$ (the purpose of K^*A is to “add” A to K , if consistently possible)

(K*4) If $\neg A \notin K$, then $K + A \subseteq K^*A$ (see K*3)

(K*5) $K^*A = K_\perp$ (K_\perp the inconsistent theory only if $\vdash \neg A$; preserve consistency, if possible)

(K*6) If $\vdash A \leftrightarrow B$, then $K^*A = K^*B$

(K*7) $K^*(A \wedge B) \subseteq (K^*A) + B$ (consider K*2 and minimality for motivation)

(K*8) If $\neg B \notin K^*A$, then $(K^*A) + B \subseteq K^*(A \wedge B)$ (see K*4)

By the above interdefinability result, it suffices for our purposes to consider contraction only in the following.

As already pointed out, a suitable order on the formulas of \mathcal{L} will give rise to a unique contraction function for maxichoice contraction. Gärdenfors and Makinson consider relations of “epistemic entrenchment”, where $A \leq B$ means

that B is more deeply entrenched, and we are more willing to give up A than to give up B , if need be, and provided we have a choice. This is made precise in

Definition 1.3 Let $\leq = \leq_K$ be a relation *relative to a knowledge set* K on the formulas of \mathcal{L} such that

- (EE1) If $A \leq B$ and $B \leq C$, then $A \leq C$ (transitivity)
- (EE2) If $A \vdash B$, then $A \leq B$ (If $A \vdash B$, then we believe at least as much in B as in A)
- (EE3) For any A and B , $A \leq A \wedge B$ (Essentially this property makes \leq a total order, and gives the necessary decision for contraction).
- (EE4) When $K \neq K_\perp$ (the set of all formulas of \mathcal{L}), then $A \notin K$ iff $A \leq B$ for all B (It is here that K matters)
- (EE5) If $B \leq A$ for all B , then $\vdash A$. (Only Truth is maximally entrenched).

We then call \leq a *relation of epistemic entrenchment for* K . We may read \leq as strength of belief, where everything outside K is not believed at all. Again, we have an interdefinability result:

Proposition 1.2 *The function $K-$ and the ordering \leq_K are interdefinable in the following sense:*

Define $K - A$ by $B \in K - A :\leftrightarrow B \in K$ and $(A < A \vee B \text{ or } \vdash A)$
 $(A < B \text{ means: } A \leq B, \text{ and not } B \leq A).$

If \leq satisfies (EE1)–(EE5), then $K-$ so defined will satisfy (K-1)–(K-8). Define $A \leq B$ (on the formulas of \mathcal{L}) by $A \leq B :\leftrightarrow A \notin K - (A \wedge B)$ or $\vdash A \wedge B$. If $K-$ satisfies (K-1)–(K-8), then \leq so defined will satisfy (EE1)–(EE5).

As emphasized, any \leq satisfying (EE1)–(EE5) will depend essentially on K . Thus, for iterated revision, as K changes, we need a new order \leq_K for every step. Our Proposition 2.2 will show that, given a preference relation \leq for \mathcal{L} , i.e. a total ordering of the formulas of \mathcal{L} which satisfies some very natural requirements (and which correspond well with the order of the Lindenbaum–Tarski algebra), we can construct from that \leq for all K an order \leq_K satisfying (EE1)–(EE5). Proposition 2.3 will show the inverse: Given an epistemic entrenchment relation, we have a preference relation too, such that the construction of Definition 2.1 will recover the epistemic entrenchment relation again.

A direct definition of a contraction function from a preference relation with an additional property to obtain (K-7), is given in Proposition 2.4. Propositions 2.8 and 2.11 will show how to naturally define an order satisfying the prerequisites of Proposition 2.2.

2 Preference relations

Definition 2.1

- (a) Call a relation \leq on the formulas of \mathcal{L} a *preference relation* for \mathcal{L} iff \leq is a binary relation such that

1. $A \vdash B \rightarrow A \leq B$
 2. \leq is transitive
 3. \leq is total
 4. $\forall B, B \leq A \rightarrow \vdash A$
- (b) Let K be closed under \vdash , and a preference relation \leq for \mathcal{L} be fixed.
 Define $A \triangleleft B$ iff D1. $A \vdash B$ or
 D2. $A \notin K$ or
 D3. $B = A \wedge C$ and $A \leq C$ and $B \in K$.

Furthermore, define \leq as the *transitive closure* of \triangleleft .

Fact 2.1 *There is a standard way of establishing $A \leq B$: let $A, C \in K$, then there is B such that $A \triangleleft A \wedge B$ by D3, and $A \wedge B \triangleleft C$ by D1 iff $A \leq A \rightarrow C$.*

Proof: “ \leftarrow ”: As $A, C \in K, A \wedge (A \rightarrow C) \in K$, so $A \triangleleft_{D3} A \wedge (A \rightarrow C) \triangleleft_{D1} C$.

“ \rightarrow ”: As $\vdash A \wedge B \rightarrow C, \vdash B \rightarrow (A \rightarrow C)$, and by Conditions 1 and 2 above $A \leq B \leq A \rightarrow C$.

Proposition 2.2 *If \leq is a preference relation for \mathcal{L} , and \leq defined as in Definition 2.1, then \leq satisfies (EE1)–(EE5). (Thus, given one global preference relation for \mathcal{L} , we can easily obtain epistemic entrenchment relations for all knowledge sets K of \mathcal{L} .)*

Proof: We first show two claims; the proof will then be trivial.

Claim 1 *For no $A \in K, B \notin K$ we have $A \leq B$.*

Proof: Suppose the contrary. Let $A = A_1 \triangleleft A_2 \triangleleft \dots \triangleleft A_n = B$. We have to “cross the border between K and $\mathcal{L} - K$ ” somewhere: There is $A_i \triangleleft A_{i+1}$ such that $A_i \in K, A_{i+1} \notin K$. Examine the cases of \triangleleft . D1 cannot be, as $A_i \in K, A_i \vdash A_{i+1}$ implies, $A_{i+1} \in K$. D2 cannot be, as $A_i \in K$. D3 cannot be, as $A_{i+1} \notin K$. So we have a contradiction.

Claim 2 $\forall B. B \leq A \rightarrow \vdash A$.

Proof: (Induction on the length of the \triangleleft -chain.) By Condition 1, it suffices to consider $B := \text{True}$. $\text{True} \triangleleft_{D1} A \rightarrow \vdash A$. $\text{True} \triangleleft_{D2} A$ cannot be, as $\text{True} \in K$. $\text{True} \triangleleft_{D3} A = \text{True} \wedge C \rightarrow \text{True} \leq C \rightarrow \vdash C$ by maximality of True and 4 of \leq (Definition 2.1(a)).

We now prove Proposition 2.2:

(EE1) is trivial by definition.

(EE2) by D1.

(EE3) We have to prove $A \leq A \wedge B$ or $B \leq A \wedge B$.

By 3. of \leq , $A \leq B$ or $B \leq A$.

Case 1: $A \wedge B \in K$. If $A \leq B$, then $A \triangleleft A \wedge B$ by D3. $B \leq A$ analogously.

Case 2: $A \wedge B \notin K$. Then $A \notin K$ or $B \notin K$, continue with D2.

(EE4) “ \rightarrow ”: $A \notin K \rightarrow A \triangleleft B$ by D2 for all B .

“ \leftarrow ”: Let $K \neq K_\perp, A \leq B$ for all B . Suppose $A \in K$. By $K \neq K_\perp$, there is $B \notin K$, and by prerequisite $A \leq B$, contradicting Claim 1.

(EE5) Claim 2.

Gärdenfors (in personal communication) has raised the question whether we can, given a relation of epistemic entrenchment, define from this relation a preference relation and recover the epistemic entrenchment relation again as in Definition 2.1. The answer is “yes” (in a simplified version due to David Makinson):

Proposition 2.3 *Let \leq_K be an epistemic entrenchment relation for a knowledge set K . Then, by definition, \leq_K is a preference relation and \leq defined for this \leq_K and K as in Definition 2.1.b is equal to \leq_K .*

Proof: “ $\leq \subseteq \leq_K$ ”: It suffices to prove $\nabla \subseteq \leq_K$. D1 and D2 are trivial. D3: Let $A \nabla A \wedge B$ by $A \leq_K B$. If $B \leq_K A \wedge B$, then $A \leq_K B \leq_K A \wedge B$, and we are finished by (EE3). “ $\leq_K \subseteq \leq$ ”: Let $A \leq_K B$. If $A \notin K$, then $A \nabla B$ by D2. If $A \in K$, then $B \in K$ by (EE4), so $A \wedge B \in K$, thus $A \nabla_{A \leq_K B} A \wedge B \nabla B$.

Next, we show how to define a contraction operation directly and in a simple way from a preference relation, without taking a detour via an epistemic entrenchment relation. To show the property (K-7), we now need an additional assumption, which will, however, be satisfied in our final construction.

Proposition 2.4

(a) *Let \leq be a preference relation for a language \mathcal{L} . Define a function $-$: Knowledge sets in \mathcal{L} and Formulas of $\mathcal{L} \rightarrow$ Sets of Formulas of \mathcal{L} by*

$$K - A := \begin{cases} K & \text{if } \vdash A \text{ or } A \notin K \\ \{B \in K : A < A \vee B\} & \text{otherwise.} \end{cases}$$

The restriction $K -$ to K will satisfy (K-1)–(K-6), (K-8) of Definition 1.2.

(b) *$K -$ satisfies (K-7) for all K too iff the following condition holds:*

(*) *For all formulas A, B, C , $(A \wedge B) \vee C \leq A \wedge B$ implies $A \vee C \leq A$ or $B \vee C \leq B$.*

(c) *Condition (*) need not necessarily be satisfied for preference relations. It holds for epistemic entrenchment relations and the construction of Propositions 2.8 and 2.11.*

Proof: (a) The cases $\vdash A$ or $A \notin K$ are trivial, so assume $\nvdash A, A \in K$ in the sequel. $A \leq A \vee B$ is always true for preference relations, so it suffices to show $A \vee B \nleq A$ to prove $B \in K - A$ (for $B \in K$).

(K-2), (K-3), (K-4), (K-6) are trivial.

(K-1): By compactness of \vdash , it suffices to consider $\vdash B \rightarrow B'$. Let $B \in K - A$, $\vdash B \rightarrow B'$, assume $B' \notin K - A$, i.e. $A \vee B' \leq A$. But then $\vdash A \vee B \rightarrow A \vee B'$, and by Condition 1 of Definition 2.1, $A \vee B \leq A \vee B' \leq A$, contradiction.

(K-5): Let $B \in K$, so $\neg A \vee B \in K$. Assume $A \vee (\neg A \vee B) \leq A$, then $\text{True} \leq A$, and by Condition 4, $\vdash A$, contradiction. So $\neg A \vee B = A \rightarrow B \in K - A$, and $B \in (K - A) + A$.

(K-8): We have to show that $(A \wedge B) \vee A \leq A \wedge B$ implies $\forall C (A \vee C \leq A \rightarrow (A \wedge B) \vee C \leq A \wedge B)$. But then $(A \wedge B) \vee C \leq A \vee C \leq A \leq (A \wedge B) \vee A \leq A \wedge B$.

(b) “ \leftarrow ”: If $\vdash A$, then $A \wedge B \leftrightarrow B$, so by (K-6) $K - B = K - (A \wedge B)$, and (K-7) is satisfied. If $A \notin K$, then $A \wedge B \notin K$, and (K-7) holds, too. Likewise for B . Suppose now $C \in K, C \notin K - (A \wedge B)$. Thus, by Condition 1 for preference relations, $(A \wedge B) \vee C \leq A \wedge B$. By (*), $C \notin K - A$, or $C \notin K - B$.

“ \rightarrow ”: Let A, B, C be such that $(A \wedge B) \vee C \leq A \wedge B$, but neither $A \vee C \leq A$ nor $B \vee C \leq B$. Consider $K := \text{Th}(A, B, C)$, the deductive closure of A, B, C . Then $A, B, A \wedge B \in K$, and $\nVdash A, \nVdash B, \nVdash A \wedge B$ (otherwise, $A \vee C \leq A$ etc. by maximality of True). Consequently, $C \in (K - A) \cap (K - B)$, but not $C \in K - (A \wedge B)$.

(c) See Example 2.1 below for a counterexample. For epistemic entrenchment relations $(A \vee C) \wedge (B \vee C) = (A \wedge B) \vee C \leq A \wedge B \leq A, B$, and (EE3) gives the result. For the construction of Propositions 2.8 and 2.11, by Lemma 2.7.7, $(A \wedge B) \vee C \leq A \wedge B$ implies $\vdash C \rightarrow A \wedge B$, thus $\vdash C \rightarrow A$ and $\vdash C \rightarrow B$, and $A \vee C \leq A, B \vee C \leq B$.

We now turn to the task of defining such a total order on the formulas of \mathcal{L} in a natural way. Let, in the following, \mathbb{D} be the Lindenbaum–Tarski algebra for the language \mathcal{L} and the empty theory. (Thus, elements of \mathbb{D} have the form $[\varphi]$ where φ is a formula of \mathcal{L} , and $[\varphi] = [\psi]$ iff $\vdash \varphi \leftrightarrow \psi$. Moreover, $[\varphi] \wedge [\psi] := [\varphi \wedge \psi]$, $-\varphi := [\neg\psi]$, and $[\varphi] \leq [\psi] : \leftrightarrow [\varphi] \wedge [\psi] = [\varphi]$.) We have a first constructive result:

Lemma 2.5 *Extending the natural ordering on the formulas of \mathcal{L} given by \mathbb{D} to a total order, preserving [True] as the only maximal element, will give a preference relation for \mathcal{L} , and thus, by Proposition 2.2, epistemic entrenchment relations \leq_K for all knowledge sets K of \mathcal{L} .*

Next, we assign probability values to formulas of \mathcal{L} , i.e. each $\varphi \in \mathcal{L}$ will have a real value $\nu(\varphi)$, and the natural order of the real numbers will order the formulas too. Of course, logically equivalent formulas should be given the same probability. We proceed indirectly, assigning first probabilities to models, and defining the probability of a formula as the sum of the probabilities of its models. The above equivalence condition will then be trivially true. It is easily seen (Proposition 2.8), that our construction will give a preference relation \leq for \mathcal{L} as needed to define the epistemic entrenchment relations \leq_K . We can improve our result and the equivalence condition to obtain $(\varphi \leq \psi \text{ and } \psi \leq \varphi)$ iff $\vdash \varphi \leftrightarrow \psi$ (Proposition 2.11). For this end, we use algebraic closure properties of the reals (Fact 2.10). We can thus construct in a natural way a total (and natural) extension of the natural order of the Lindenbaum–Tarski algebra \mathbb{D} , such that $([\varphi] \leq [\psi] \text{ and } [\psi] \leq [\varphi])$ is equivalent to $[\varphi] = [\psi]$. In conclusion, we remark that the whole process can be easily relativized to a fixed theory, by considering only models of that theory (see Definition 2.5). But first, we need some constructions:

Let \mathcal{A} be the σ -algebra (i.e. the \aleph_1 -complete Boolean algebra) of Lebesgue-measurable sets restricted to subsets of the unit interval $[0,1]$. Let μ be the usual Lebesgue measure. (The reader unfamiliar with these notions will find definitions and properties in any book on measure and integration theory.)

Definition 2.2 Let $\langle x_i : i \in \omega \rangle$ be a sequence of reals in the open interval $(0,1)$. Define by induction:

$$a_0 := [0, x_0), b_0 := \{0, x_0, 1\}.$$

Let a_n, b_n be defined ($n \in \beta$). b_n will be a set of $2^{n+1} + 1$ elements, a_n a disjoint union of 2^n nonempty intervals. Let $b_n = \{y_j : j < 2^{n+1} + 2\}$, the y_j in increasing order. Define

$$a_{n+1} := \bigcup \{[y_j, y_j + (y_{j+1} - y_j) * x_{n+1}) : j < 2^{n+1} + 1\}$$

$$b_{n+1} := b_n \cup \{y_j + (y_{j+1} - y_j) * x_{n+1} : j < 2^{n+1} + 1\}.$$

Finally, set $\bar{a}_n := [0, 1) - a_n$. (See Diagram 2.1 below.)

Let \mathbb{B} be the \aleph_1 -complete subalgebra of \mathbb{A} generated by $\{a_i : i \in \omega\}$.

Fact 2.6 For the a_i thus defined we have:

- (1) $\mu(a_n) = x_n$
- (2) $\mu(\bar{a}_n) = 1 - \mu(a_n)$ (trivial)
- (3) $\mu(\bigcap \{c_n : n \in X\}) = \prod \{\mu(c_n) : n \in X\}$ where c_n is either a_n or \bar{a}_n for $X \subseteq \omega$ finite, by the “independence” of the construction. This property is essential to all that follows.

In the rest of the paper, let $\mathcal{L} = \{p_i : i \in \omega\}$ be a countable language of propositional calculus.

Definition 2.3

- (a) Define $f: \mathcal{L} \rightarrow \{a_i : i \in \beta\}$ by $f(p_i) := a_i$, i.e. $\mu(f(p_i)) = x_i$.
- (b) Let M be the set of assignments of truth values to finite subsets of \mathcal{L} , $t \in M$, t defined on $\mathcal{L}' \subseteq \mathcal{L}$. (It suffices to consider finite subsets, as standard propositional calculus admits only finite formulas.) Define $g(t) := \bigcap \{a_i : p_i \in \mathcal{L}', t(p_i) = \text{true}\} \cap \bigcap \{\bar{a}_i : p_i \in \mathcal{L}', t(p_i) = \text{false}\}$.

Thus, $\mu(g(t)) = \mu(\bigcap \{a_i : p_i \in \mathcal{L}', t(p_i) = \text{true}\} \cap \bigcap \{\bar{a}_i : p_i \in \mathcal{L}', t(p_i) = \text{false}\}) = \prod \{x_i : p_i \in \mathcal{L}', t(p_i) = \text{true}\} * \prod \{1 - x_i : p_i \in \mathcal{L}', t(p_i) = \text{false}\}$, and we have defined for every assignment $t \in M$ a real value $\mu(g(t))$. There is a natural way to extend this function to formulas:

Definition 2.4 Let φ be a formula with propositional variables $p_i \in \mathcal{L}_\varphi \subseteq \mathcal{L}$ finite.

- (a) Let $\text{Val}(\varphi) := \{t \in M : \text{dom}(t) = \mathcal{L}_\varphi, t(\varphi) = \text{true}, \text{ i.e. } \varphi \text{ is true under } t\}$.
- (b) So we can define $\nu(\varphi) := \sum \{\mu(g(t)) : t \in \text{Val}(\varphi)\}$.

(See Diagram 2.1.)

Let $\mathcal{L} = \{p, q\}$, $t(p) = \text{true}$, $t(q) = \text{false}$, $t'(p) = \text{false}$, $t'(q) = \text{true}$, $\varphi = p \leftrightarrow \neg q$

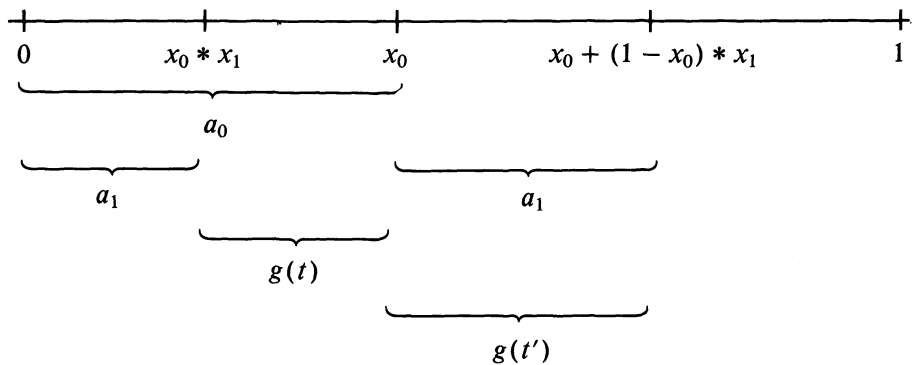


Diagram 2.1.

Thus, $\mu(a_0) = x_0$, $\mu(a_1) = x_1$, $\nu(\varphi) = \mu(g(t)) + \mu(g(t')) = x_0 * (1 - x_1) + (1 - x_0) * x_1$. Our construction has the following properties:

Lemma 2.7

- (1) $\nu(\varphi)$ is independent of $\text{dom}(t)$ in the following sense
Let $\mathcal{L}_\varphi \subseteq \mathcal{L}' \subseteq \mathcal{L}$ finite. Then $\nu(\varphi) := \sum \{\mu(g(t)) : t \in \text{Val}(\varphi)\} = \sum \{\mu(g(t)) : t \in \mathbb{M}, \text{dom}(t) = \mathcal{L}', t(\varphi) = \text{true}\}$.
- (2) By definition of Val and ν , logically equivalent formulas will have the same real value $\nu(\varphi)$.
- (3) $\vdash \varphi \rightarrow \psi$ implies $\nu(\varphi) \leq \nu(\psi)$.
(To see this, consider $\mathcal{L}' = \mathcal{L}_\varphi \cup \mathcal{L}_\psi$, use (1) and the fact that every assignment which makes φ true will make ψ true too.)
- (4) $\nu(\neg\varphi) = 1 - \nu(\varphi)$.
(Use $\nu(\text{true}) = \sum \{\mu(g(t)) : t \in \mathbb{M}, \text{dom}(t) = \mathcal{L}' \text{ finite}\} = 1$, $t(\varphi) = \text{true} \leftrightarrow t(\neg\varphi) = \text{false}$, and for $t, t' \in \mathbb{M}$ with the same domain $t \neq t' \rightarrow g(t) \cap g(t') = \emptyset$.)
- (5) Exactly the valid formulas will have real value $\nu(\varphi) = 1$.
($g: \mathbb{D} \rightarrow \mathbb{B}$ (extended suitably to formulas) is an injective homomorphism of Boolean algebras, and use the above arguments.)
- (6) $\nu(\varphi) \leq \nu(\psi) \leftrightarrow \nu(\neg\psi) \leq \nu(\neg\varphi)$ (by (4))
- (7) $\nu(\varphi \vee \psi) \leq \nu(\varphi) \leftrightarrow \vdash \psi \rightarrow \varphi$.
*“ \leftarrow ”: by (3)
 “ \rightarrow ”: Suppose $\vdash \psi \rightarrow \varphi$. Thus $M' := \{t : t \models \varphi\} \subset M := \{t : t \models \varphi \vee \psi\}$, let $t \in M - M'$. As $x_i \in (0, 1)$, $\mu g(t) \neq 0$, thus $\nu(\varphi) := \sum \{\mu g(t) : t \in M'\} < \sum \{\mu g(t) : t \in M\} =: \nu(\varphi \vee \psi)$.*
- (8) We cannot expect $\nu(\varphi \wedge \psi) = \nu(\varphi) * \nu(\psi)$ or $\nu(\varphi \vee \psi) = \nu(\varphi) + \nu(\psi)$. Just think of $\varphi = \psi$. These equations can only be valid if φ and ψ are independent. For this reason, we gave first a value to models, which are independent, and then to formulas.

We have thus proved our main constructive result:

Proposition 2.8 *Let $p_i: i \in \omega$ be given a probability $x_i \in (0, 1)$, then this gives rise naturally to probabilities $\nu(\varphi)$ for any formula in \mathcal{L} , such that (1)–(6) of Lemma 2.7 are valid, and thus to a preference relation \leq for \mathcal{L} , i.e. satisfying 1–4 of \leq in Definition 2.1, and thus the prerequisites of Propositions 2.2 and 2.4.*

Fact 2.9 *Let $0 \leq a \leq b < 1$. Augment the natural order of the reals by setting $x \leq^+ y$ for all $a \leq x, y \leq b$, i.e. “identify” all elements of the interval $[a, b]$. Let ν be defined as in the construction leading to Proposition 2.8 and set $\varphi \leq \psi$ iff $\nu(\varphi) \leq \nu(\psi)$ or $\nu(\psi) \leq^+ \nu(\varphi)$. Then \leq is still a preference relation on \mathcal{L} .*

Proof: In Definition 2.1, 1 and 3 are trivial, 4 holds by $b < 1$. But 2 is simple too: consider, e.g., $x \leq y \leq^+ z$. If $x > z$, then $a \leq z \leq x \leq y \leq b$, and $x \leq^+ z$.

Example 2.1 Consider now $\mathcal{L} := \{A, B, C\}$, and set $\mu f(A) := 1/2$, $\mu f(B) := 1/3$, $\mu f(C) := 1/5$, $a := 5/30$, $b := 10/30$, and identify in the interval $[a, b]$ as described in Fact 2.9. Then $\nu(A) = 15/30$, $\nu(B) = 10/30$, $\nu(A \wedge B) = 5/30$, $\nu(A \vee C) = 18/30$, $\nu(B \vee C) = 14/30$, $\nu((A \wedge B) \vee C) = 10/30$. By identification,

$(A \wedge B) \vee C \leq A \wedge B$, but neither $A \vee C \leq A$ nor $B \vee C \leq B$. Thus, this order is a counterexample as promised in Proposition 2.4.c.

So far, it is quite possible that $\nu(\varphi) = \nu(\psi)$, but $\not\models \varphi \leftrightarrow \psi$. We now make ν injective (modulo \leftrightarrow). Thus, we improve our result such that $(\varphi \leq \psi$ and $\psi \leq \varphi)$ iff $\vdash \varphi \leftrightarrow \psi$. Choosing the x_i of Definition 2.2 above according to the following fact on the reals will do the trick:

Fact 2.10 *Let $X := \{x_i : i \in \omega\} \subset I \subseteq \mathbb{R}$ (the reals), I uncountable, be given. Then there is $x' \in I$ such that x' is not equal to any real that can be obtained by finite addition, subtraction, multiplication, division from elements of $\mathbb{Q} \cup X$ (\mathbb{Q} the rationals).*

$\text{Card}(I) > \text{card}(\mathbb{Q} \cup X) = \aleph_0$ suffices for the proof.

We choose the x_i for the above construction of the a_i in Definition 2.2 according to this fact. Suppose that φ, ψ are not equivalent, but $\nu(\varphi) = \nu(\psi)$. Thus, there is an assignment t such that $t(\varphi) \neq t(\psi)$. So $\bigcup \{g(t) : t \in \text{Val}(\varphi)\} \neq \bigcup \{g(t) : t \in \text{Val}(\psi)\}$ (w.l.o.g. all t with the same domain $p_0 \dots p_n$ and n chosen least such that the assumption is valid), but $\nu(\varphi) = \nu(\psi)$. Thus,

$$\nu(\varphi) = \sum_{i=0}^m \prod_{j=0}^n y_{i,j}, \quad \nu(\psi) = \sum_{i=0, m'} \prod_{j=0, n} y'_{i,j}$$

where the $y_{i,j}, y'_{i,j}$ are either x_j or $1 - x_j$. After multiplication, the equation looks like this: $s_1 + \dots + s_k = t_1 + \dots + t_l$, the s_u and t_u are of the form: 1 or $\pm x_{r_1} * \dots * x_{r_n}$ and each x_j occurs at most once in each summand. After cancelling summands of the same form that occur on both sides of the equation, x_n will still occur in at least one of the summands, as n was chosen least. So, we can solve the equation (linear in x_n) for x_n and have $x_n = f(x_0 \dots x_{n-1})$ where f is composed of addition, subtraction, multiplication, division—contradicting Fact 2.10. As the x_i can be chosen within any distance > 0 from a desired value, choosing x_i according to this fact is no real restriction. We have thus obtained our injectivity result and shown the following proposition.

Proposition 2.11 *Let $p_i : i \in \omega$ be given a probability $x_i \in (0, 1)$, chosen according to Fact 2.10, then this gives rise naturally to probabilities $\nu(\varphi)$ for any formula in \mathcal{L} , such that (1)–(6) of Remark 4 are valid, and $(\varphi \leq \psi$ and $\psi \leq \varphi)$ iff $\vdash \varphi \leftrightarrow \psi$. In other words, this defines a total (and natural) extension of the natural order of the Lindenbaum–Tarski algebra \mathbb{D} , and, in addition, $([\varphi] \leq [\psi]$ and $[\psi] \leq [\varphi])$ iff $[\varphi] = [\psi]$.*

Remark 2.12 So far, we have worked over the empty theory and its Lindenbaum–Tarski algebra. It is easy to extend our results to nonempty theories, by considering only models of that theory in our Definition 2.4. Thus, we can define, e.g.,

Definition 2.5 Let T be a theory in $\mathcal{L}_T \subseteq \mathcal{L}$ finite, and $\mathcal{L}' := \mathcal{L}_T \cup \mathcal{L}_\phi$. Set

$$\nu_T(\varphi) := \frac{\sum \{\mu(g(t)) : \text{dom}(t) = \mathcal{L}', t(\varphi) = \text{true}, t(\psi) = \text{true for all } \psi \in T\}}{\sum \{\mu(g(t)) : \text{dom}(t) = \mathcal{L}', t(\psi) = \text{true for all } \psi \in T\}}.$$

So $\nu_T(\varphi)$ and $\nu_T(\varphi')$ will be equal, iff the models that make T true treat φ and φ' in the same way.

Remark 2.13 We can work backwards in the following sense too: Suppose we are given a set of formulas $\{\varphi_i : i \in I\}$ and preferences (probabilities) $\pi(\varphi_i)$ for all $i \in I$. Can we find a sequence $x_i : i \in \omega$ such that, constructing as above,

$$\pi(\varphi_i) = \nu(\varphi_i) := \sum \{\mu(g(t)) : \text{dom}(t) = \mathcal{L}_{\varphi_i}, t \models \varphi_i\}?$$

The answer is trivial and canonical. We have a number of equations: $\pi(\varphi_i) = \sum \{\prod \{x_j : t(p_j) = \text{true}\} * \prod \{(1 - x_j) : t(p_j) = \text{false}\} : t \in \text{Val}(\varphi_i)\}$ and any solution $\{x_j : j \in \omega, x_j \in (0, 1)\}$ (if there is one) of this system of equations, and $\{a_j : j \in \omega\}$ chosen as above will do what we need.

3 Measuring theories, and an outlook for a different treatment of theory revision

In this section, we discuss three somewhat different approaches to theory contraction. In the first two, we extend our measure from formulas to theories, and use it to do contraction. The first attempt is very naive, and mentioned only for illustration. The second approach is again (i.e. as in Section 2) “maxichoice contraction” in the sense of [1], [4], [6], and as such plagued by the well-known completeness result: $K - A \cup \{\neg A\}$ is a complete theory (see below for a proof).

In the third case, we take a totally different approach and consider pairs $\langle K, X \rangle$, where K is a theory, and X an axiom set for K . This approach suffers from another defect: it is highly dependent on the syntactic structure of the axiom set: The “coarser” the axiom set is (in the one extreme the conjunction of K), the more drastic and coarse a contraction will be; the more fine-grained it is (in the other extreme K itself), the more we approach the above completeness result. This is made precise in Proposition 3.1, which, basically, shows that splitting an axiom ψ into $\{\psi \vee \varphi, \psi \vee \neg \varphi\}$ will decide φ , i.e. give completeness with respect to φ .

Consider now theories $T, T' \dots$ in some finite $\mathcal{L}' \subseteq \mathcal{L}$. It is natural to define $\nu(T) := \sum \{\mu(g(t)) : \text{dom}(t) = \mathcal{L}', t \models T\}$. ($t \models T$ means, of course, $t(\varphi) = \text{true}$ for all $\varphi \in T$, see Remark 2.12.) In other words, $\nu(T)$ is the sum of the probabilities of all \mathcal{L}' -models t that make all $\varphi \in T$ true. The more specific a theory is, the less likely it is, too: $T \subseteq T' \rightarrow \nu(T') \leq \nu(T)$, and the empty theory has probability 1. On the other hand, we are interested in “good choices”, i.e. we prefer φ to $\neg \varphi$ if $\nu(\varphi) > \nu(\neg \varphi)$. So ν will be a good measure only for theories of the same level of specificity. In other words, $K - A$ (here, $K - A$ means some contraction of K with respect to A) cannot sensibly be the ν -maximal $K' \subseteq K$ s. $K' \not\models A$, as this is always the empty theory.

A better choice might be a ν -maximal one (if it exists) among $K_A := \{K' \subseteq K \text{ maximal: } K' \not\models A, K' \text{ is } \vdash\text{-closed}\}$, this is again “maxichoice contraction”. But there is a problem to maxichoice contraction, pointed out in [2]: For any $K' \in K_A, A \in K$, $\text{Th}(K' \cup \{\neg A\})$ will be a complete theory. (The proof is very simple: Let B be given. As $A \in K$, $A \vee B$ and $A \vee \neg B$ are in K . Suppose $A \vee B \notin K', A \vee \neg B \notin K'$. As $A \vee B \notin K'$, by maximality there is $C_0 \in K'$ such that $C_0 \wedge (A \vee B) \vdash A$, and as $A \vee \neg B \notin K'$, there is $C_1 \in K'$ with $C_1 \wedge (A \vee \neg B) \vdash A$. Thus, for $C := C_0 \wedge C_1 \in K', C \wedge (A \vee B) \vdash A, C \wedge (A \vee \neg B) \vdash A$, consequently $C \wedge B \vdash A, C \wedge \neg B \vdash A$, and $C \vdash A$, contradicting $C \in K' \in K_A$. Thus, $A \vee B \in K'$ or $A \vee \neg B \in K'$, and $K' \cup \{\neg A\} \vdash B$, or $K' \cup \{\neg A\} \vdash \neg B$.)

We now show that this problem essentially carries over to theory revision based on axiom sets too. So far, we have examined theories without any specified axiom system generating the theory. In the following, we consider pairs $\langle K, X \rangle$, where X is an axiom set for K . Define $\overline{K_{A,X}} := \{\langle K', X' \rangle : X' \subseteq X \text{ maximal, } X' \not\models A, K' = \text{Th}(X')\}$ and choose $K - A$ as a ν -maximal $\langle K', X' \rangle$ from $\overline{K_{A,X}}$ (if possible). Consider now $\langle K, X_1 \rangle, \langle K, X_2 \rangle$, where $X_1 := \{\varphi, \varphi \rightarrow \psi, \psi\}$, and $X_2 := \{\varphi, \varphi \rightarrow \psi\}$. In both cases, we can infer ψ , and the resulting theories are the same. Suppose we now retract $\varphi \rightarrow \psi$. In case 1, it is very sensible to uphold ψ , whereas in case 2, it will not be a good choice. (This example can be found analogously in [3] and [5].) So we are highly dependant on the syntactic form of the axioms, and this is certainly not very desirable. As another example, consider revising a theory which is given by the axiom sets $\{a_1, a_2\}$ or $\{a_1 \wedge a_2\}$. So revision may give different results ($\{a_1\}$ or $\{a_2\}$ vs. the empty theory), which is a doubtful outcome. To avoid this influence of the syntactic form, we might split the axioms as far as possible to obtain optimal results. This procedure, however, approaches completeness, as the following proposition will show:

Let $Y := \{y_1 \dots y_m\}$ be minimal with $Y \vdash A$. Let $Y' := \{y_2 \dots y_m\}$. Split y_1 into $\{y_1 \vee \varphi, y_1 \vee \neg\varphi\}$. Both $Y' \cup \{y_1 \vee \varphi\} \vdash A$ and $Y' \cup \{y_1 \vee \neg\varphi\} \vdash A$ cannot be, since otherwise $Y' \vdash A$, contradicting minimality of Y . So $Y_0 := Y' \cup \{y_1 \vee \varphi\}$ or $Y_1 := Y' \cup \{y_1 \vee \neg\varphi\}$ is a good candidate for $Y - A$, i.e. for contracting Y with respect to A . So, let $Y - A$ be Y_0 if $Y_0 \not\models A$, and Y_1 otherwise.

Proposition 3.1 *If $Y - A$ is as just defined, then $\bar{Y} := Y - A \cup \{\neg A\}$ decides φ .*

Proof: We have $y_1 \wedge \dots \wedge y_m \vdash A$, thus $\neg A \wedge y_2 \wedge \dots \wedge y_m \vdash \neg y_1$. Consequently, $\bar{Y} \vdash \neg y_1$. If $Y - A$ is Y_0 , then (by $y_1 \vee \varphi \in \bar{Y}$) $\bar{Y} \vdash \varphi$, if $Y - A$ is Y_1 , then $\bar{Y} \vdash \neg\varphi$.

This is further illustrated by the following point of view: We may consider “maxichoice contraction” as theory revision with axioms—taking the full theory as axiom set, and choosing a maximal subset from which A does not follow—resulting in full completeness.

4 Conclusion We first have shown how to construct, from a simple order on all formulas of a given language, orderings suitable for theory revision in the style of Gärdenfors/Makinson. Next, we have shown how to define naturally such a simple order with some additional very nice properties. In the end, we have reconsidered theory revision more generally. Roughly speaking, it turns out that, for theory revision, every axiom set is in a sense a compromise between completeness (too fine-grained) and over-reaction to revision (too coarse). In other words, we should consider to make the choice of the axiom set part of the revision process itself.

REFERENCES

- [1] Alchourron, C., P. Gärdenfors, and D. Makinson, “On the logic of theory change,” *The Journal of Symbolic Logic*, vol. 50 (1985), pp. 510–530.

- [2] Alchourron, C., and D. Makinson, "On the logic of theory change : Contraction functions and their associated revision functions," *Theoria*, vol. 48 (1982), pp.14–37.
- [3] Fagin, R., J. D. Ullmann, and M. Y. Vardi, "On the semantics of updates in databases," *Proceedings of the Second ACM SIGACT-SIGMOD*, 1983, pp. 352–365.
- [4] Gärdenfors, P., *Knowledge in Flux*, MIT Press, Cambridge, Massachusetts, 1988.
- [5] Gärdenfors, P. and D. Makinson, "Revisions of knowledge systems using epistemic entrenchment," in *Theoretical Aspects of Reasoning about Knowledge*, edited by M. Y. Vardi, Morgan Kaufmann, Los Altos, California, 1988, pp.83–95.
- [6] Makinson, D., "How to give it up: A survey of some formal aspects of the logic of theory change," *Synthese*, vol. 62 (1985), pp. 347–363.
- [7] Spohn, W., "Ordinal conditional functions: A dynamic theory of epistemic states," in *Causation in Decision, Belief Change, and Statistics*, edited by W. L. Harper and B. Skyrms, Dordrecht 1988, Reidel, vol. 2, pp. 105–134.

IBM Germany, IWBS W&S3
PO Box 80 08 80
D-7000 Stuttgart 80, Germany