Minimal Belief Change and Pareto-Optimality

Oliver Schulte

Department of Computing Science
University of Alberta
Edmonton, AB T6G 2H1, Canada
oliver.schulte@ualberta.ca

Abstract. This paper analyzes the notion of a minimal belief change that incorporates new information. I apply the fundamental decision-theoretic principle of *Pareto-optimality* to derive a notion of minimal belief change, for two different representations of belief: First, for beliefs represented by a *theory*—a deductively closed set of sentences or propositions—and second for beliefs represented by an axiomatic *base* for a theory. Three postulates exactly characterize Pareto-minimal revisions of theories, yielding a weaker set of constraints than the standard AGM postulates. The Levi identity characterizes Pareto-minimal revisions of belief bases: a change of belief base is Pareto-minimal if and only if the change satisfies the Levi identity (for "maxichoice" contraction operators). Thus for belief bases, Pareto-minimality imposes constraints that the AGM postulates do not.
Keywords: belief revision, decision theory

1 Minimal Theory Change

New information changes our beliefs continually. How should we incorporate new assertions into a body of existing ones? This question arises in many situations of practical interest. For example, if the new assertion describes new data, incorporating the evidence into current beliefs is an essential part of learning systems. If the new assertion is a datum presented to a database system, we face the question of how to update a database, and the same goes for knowledge bases.

In the last two decades or so, the following principle has attracted much interest among computer scientists and logicians [3, 8, 6, 10, 2]: Revise your beliefs so as to *minimize the extent of change* from the original beliefs. The aim of this paper is to analyze the notion of minimal belief change. I derive axioms for minimal belief change from basic principles of *decision theory*. The same decision-theoretic principles lead to different results for different ways of formally representing beliefs. Specifically, I consider two such representations: Belief modeled as a deductively closed set of sentences (or propositions), and belief modeled by an axiomatic "belief base". For each of these representations of belief, I consider the consequences of using the fundamental decision-theoretic principle of *Pareto-Optimality* to define minimal belief changes.

Roughly, Pareto-minimal belief revisions are those that cannot be improved by adding fewer beliefs without giving up more, or by giving up fewer beliefs without adding more. As it turns out, there is a purely set-theoretic definition of Pareto-minimal belief revisions in terms of the symmetric set differences between the current theory and alternative revisions. The main theorem of this paper establishes that certain axioms for belief revision characterize Pareto-minimal theory changes, in the sense that a theory change is Pareto-minimal if and only if the change satisfies these axioms. The chief difference between Pareto-minimality and the standard AGM postulates [3] arises in the case in which the current theory neither entails the new information nor its negation. In that case, the AGM revision is the result of adding the new information to the current theory. Pareto-minimal revisions, however, may be logically *weaker* than the AGM revision.[1]

Pareto-optimality leads to different results for minimal revisions of *belief bases*, sets of sentences that need not contain all of their logical consequences. The well-known *Levi identity* characterizes Pareto-minimal changes of belief bases: I prove that they are *exactly* those that result from, first, retracting just enough basic beliefs to make the agent's basic beliefs consistent with the new information (technically, a "maxichoice contraction" [3, Ch. 4.2]), and second, adding the new information to the basic beliefs contracted in this manner. Since AGM revisions may give up more beliefs than maxichoice contraction permits, this characterization shows that Pareto-minimality yields some constraints on the revision of belief bases that the AGM axioms do not require (cf. [1]).

2 Theories

Following much of the belief revision literature, I employ a syntactic representation of an agent's beliefs. However, all the developments to follow are valid for a semantic approach based on propositions (sets of models) as well. I assume that some language L has been fixed, and take a theory to be a deductively closed set of formulas from L. In Section 5 I considers belief sets that are not deductively closed.

As is usual in belief revision theory, my assumptions about the structure of the language in which an agent formulates her beliefs are sparse; essentially, all I assume is that the language features the usual propositional connectives. I take as given a suitable consequence relation between sets of formulas in the language, obeying the standard Tarskian properties. The formal presuppositions are as follows.

A **language** L is a set of formulas satisfying the following conditions. (1) L contains a **negation operator** $\neg$ such that if p is a formula in L, so is $\neg p$. (2) L contains a **conjunction connective** $\wedge$ such that if p and q are formulas in L, so is $p \wedge q$. (3) L contains an **implication connective** $\rightarrow$ such that if p and q are formulas in L, so is $p \rightarrow q$.

A **consequence operation** $Cn : 2^L \rightarrow 2^L$ represents a notion of entailment between sets of formulas from a language L. A set of formulas Γ **entails** another

[1] In this respect, Pareto-minimal revisions agree with Katsuno and Mendelzon's approach to "belief update" [6]; see Section 4.

set of formulas Γ', written $\Gamma \vdash \Gamma'$, iff $Cn(\Gamma) \supseteq \Gamma'$. A set of formulas Γ entails a formula p, written $\Gamma \vdash p$, iff $p \in Cn(\Gamma)$. I assume that Cn satisfies the following properties, for all sets of formulas Γ, Γ': **Inclusion**: $\Gamma \subseteq Cn(\Gamma)$; **Monotonicity**: $Cn(\Gamma) \subseteq Cn(\Gamma')$ whenever $\Gamma \subseteq \Gamma'$; and **Iteration**: $Cn(Cn(\Gamma)) = Cn(\Gamma)$.

A **theory** is a deductively closed set of formulas. That is, a set of formulas $T \subseteq L$ is a theory iff $Cn(T) = T$. The entailment relation $\vdash$ is related to the propositional connectives as follows.

Modus Ponens If $\Gamma \vdash p$, $(p \to q)$, then $\Gamma \vdash q$.
Implication If $\Gamma \vdash q$, then $\Gamma \vdash (p \to q)$.
Deduction $\Gamma \cup \{p\} \vdash q$ iff $\Gamma \vdash (p \to q)$.
Conjunction $\Gamma \vdash (p \wedge q)$ iff both $\Gamma \vdash p$ and $\Gamma \vdash q$.
Consistency Suppose that $\Gamma \not\vdash p$. Then $\Gamma \cup \{\neg p\} \not\vdash p$.
Inconsistency $\{p \wedge \neg p\} \vdash L$.
Double Negation $\Gamma \vdash p$ iff $\Gamma \vdash \neg\neg p$.

Classical propositional logic satisfies these assumptions. Belief revision theorists usually assume that the consequence relation Cn is compact; none of the results in this paper require compactness.[2] For the remainder of this paper, assume that a language L and a consequence relation Cn (and hence an entailment relation $\vdash$) have been fixed that satisfy the conditions laid down above.

3 Theory Change: Additions and Retractions

My approach to defining minimal belief change is to seek a *partial order* $\prec_T$ where we read $T_1 \prec_T T_2$ as "T_1 is a smaller change from T than T_2 is". Since this ordering is partial, there may be possible changes that are incomparable. As far as a given partial order among theory changes goes, if two changes are incomparable, we should view neither as a smaller change than the other. However, a theory change T_2 from an old theory T is *not* minimal if there is another, comparable, new theory T_1 such that $T_1 \prec_T T_2$. Thus I shall take minimal changes from a current theory T to be the minimal elements in the given partial order $\prec_T$.

I make use of decision-theoretic principles to define partial orders among theory changes. Let's begin by distinguishing two kinds of change: A *retraction* in which the old theory entails a formula that the new theory does not entail, and an *addition*, in which the new theory entails a formula that the old theory does not entail. Thus T' **retracts** the formula p from T iff $T \vdash p$ and $T' \not\vdash p$, and T' **adds** the formula p to T iff $T \not\vdash p$ and $T' \vdash p$.

Next, I define two partial orders among theory changes. The first partial order defines a notion of a new theory T_1 "retracting more" from a previous theory T than another new theory T_2, namely if T_1 retracts all the formulas from T that T_2 retracts from T, and T_1 retracts at least one formula from T that T_2 does not

[2] A consequence relation Cn is compact iff for all formulas p and sets of formulas Γ, we have that $p \in Cn(\Gamma)$ only if $p \in Cn(\Gamma')$ for some *finite* subset Γ' of Γ.

retract. The second partial order defines a notion of a new theory T_1 "adding more" to a previous theory T than another new theory T_2, namely if T_1 adds all the formulas from T that T_2 adds to T, and T_1 adds at least one formula to T that T_2 does not add to T. It is not difficult to see that these notions can be expressed in terms of set inclusions as follows ($\subset$ denotes proper set inclusion).

Definition 1. *Let T, T_1, T_2 be three theories.*

1. *T_1 **retracts more** formulas from T than T_2 does $\iff T - T_2 \subset T - T_1$.*
2. *T_1 **adds more** formulas to T than T_2 does $\iff T_2 - T \subset T_1 - T$.*

We may think of the addition partial order and the retraction partial order as defining two distinct dimensions of "cost" in theory revision. If additions and retractions were linked such that minimizing one minimizes the other, this distinction would have no interesting consequences for the question of how to minimize theory change: we would just minimize both additions and retractions at once. What makes the distinction important is the fact that in general, additions and retractions *trade off* against each other. Typically, avoiding retractions entails adding more sentences than necessary, and avoiding additions entails retracting more sentences than necessary. An example will clarify this point.

Example. Imagine a cognitive scientist who believes that a certain AI system, say SOAR, is the only candidate for machine intelligence. This scientist believes that "if SOAR is not intelligent ($\neg s$), there is no intelligent machine ($\neg m$)". Thus the scientist believes the sentence $p = \neg s \rightarrow \neg m$. Suppose that the scientist believes only the consequences of p, that is, her current theory is $T = Cn(\{p\})$. In particular, the scientist neither believes that there is an intelligent machine (m), nor does she believe that there is no intelligent machine ($\neg m$). Now the scientist receives new information to the effect that SOAR is not intelligent. She has to revise her theory T on evidence $\neg s$. Let us consider two possible revisions, T_1 and T_2. Revision T_1 adds the new information $\neg s$ to T and accepts the deductive consequences of this addition; thus $T_1 = Cn(\{p\} \cup \{\neg s\})$. This revision T_1 is logically stronger than T and hence retracts nothing from T. However, the revision adds the sentence $\neg m$ ("there is no intelligent machine"), since p and $\neg s$ entail $\neg m$.

Contrast this with a different revision T_2 that retracts the scientist's initial belief that SOAR is the only road to machine intelligence, and adds the new information that SOAR is not intelligent. That is, $T_2 = Cn(\{\neg s\})$. This revision T_2 retracts more from T than T_1 does. On the other hand, T_2 adds less to T than T_1 does, since T_2 is strictly weaker than T_1. In particular, T_2 continues to reserve judgment about whether machine intelligence is possible or not, whereas T_1 concludes that it is impossible ($\neg m$).

As the results below show, this example illustrates a general tension between avoiding additions and avoiding retractions; essentially, additions and retractions trade off against each other unless the current theory already entails the new information. When additions and retractions stand in conflict, how shall we make trade-offs between them? This is the topic of the next section.

4 Pareto-Minimal Theory Change

When a conflict arises between avoiding additions and avoiding retractions in belief revision, an agent may strike a subjective balance between them, as in any case of conflicting aims. She may assign one kind of change more subjective weight than the other, or favour some beliefs as more "entrenched" than others.[3] But before we resort to subjective factors, we can look to decision theory for an objective constraint that applies to all agents seeking to minimize theory change. If avoiding changes is our aim, then we should avoid revisions that make more additions than necessary without avoiding retractions, and we should avoid revisions that make more retractions than necessary without avoiding additions. This is an instance of the following uncontroversial principle for rational choice under certainty between objects with multiple relevant attributes: If A is at least as desirable as B with respect to all relevant attributes, and A is strictly better than B with respect to at least one attribute, choose A over B. The decision-theoretic term for this principle is *Pareto-optimality*.[4] For minimal theory change, we can render it as follows.

Definition 2. *Let T, T_1, T_2 be three theories. T_1 is a* ***greater change*** *from T than T_2 is $\iff$*

1. *T_1 retracts more formulas from T than T_2 does, and for all formulas p, if T_2 adds p to T, then T_1 adds p to T;* ***or***
2. *T_1 adds more formulas to T than T_2 does, and for all formulas p, if T_2 retracts p from T, then T_1 retracts p from T.*

An equivalent purely set-theoretic definition is: T_1 is a greater change from T than T_2 is iff $T_2 \triangle T \subset T_1 \triangle T$, where $\subset$ denotes proper inclusion and $\triangle$ is symmetric difference ($A \triangle B = A - B \cup B - A$).[5] (I owe this definition to an anonymous referee.)

Thus the principle of Pareto-Optimality defines a partial relation $\prec_T$ between theories: $T_2 \prec_T T_1$ iff T_1 is a greater change from T than T_2 is. It seems that we can now take a minimal change from T to be a minimal theory in the

[3] Many investigators assume that a relation of "epistemic entrenchment" guides belief revision (e.g., Gärdenfors and Nayak [3, Ch.4], [8]). They typically take epistemic entrenchment to be subjective in the sense that different rational agents may view the same belief as entrenched to different degrees.

[4] Social choice theorists often use Pareto-optimality as a principle for comparing social states. The Pareto principle applies both to social choice and to choice between objects with multiple attributes because these two choice situations are formally equivalent (identify the set of "attributes" with the set of individual members of society).

[5] Chou and Winslett too define a partial order among (first-order) models of the form "N is closer to M than N' is" in terms of symmetric difference [2]. From the perspective of this paper, their definition is a special case of Definition 2, namely Pareto-minimality applied to models rather than theories.

$\prec_T$-ordering. But on that definition, the only minimal change from T is T itself! Of course, it is generally true that the smallest change is no change, on any acceptable notion of "small change". What we want is a minimal change that satisfies *additional constraints.* In the case of belief update, the additional constraint is that the minimal theory change should incorporate the new information. Accordingly, I define a Pareto-minimal theory change from T, given new information p, as a theory that is minimal in the $\prec_T$-ordering among the theories that entail p.

Definition 3. *Let T, T_1 be two theories, and let p be a formula. Then T_1 is a* ***Pareto-minimal change*** *from T that incorporates $p \iff$*

1. *$T_1 \vdash p$, and*
2. *there is no other theory T_2 such that $T_2 \vdash p$ and T_1 is a greater change from T than T_2 is.*

Now we are ready for the main result of this paper: Necessary and sufficient conditions for a theory revision to be a Pareto-minimal change.

Theorem 1. *Let T be a theory and let p be a formula. A theory revision $T * p$ is a Pareto-minimal change from T that incorporates $p \iff$*

1. *$T * p \vdash p$, and*
2. *$T \cup \{p\} \vdash T * p$, and*
3. *if $T \vdash p$, then $T * p = T$.*

The theorem shows that the tension between additions and retractions arises whenever the agent's current theory does not already entail the new information. When this is the case, the revisions that make Pareto-acceptable trade-offs run in strength from adding the evidence to the current theory ($T \cup \{p\}$) to entailing nothing but the evidence and its consequences ($\{p\}$). This account of minimal change distinguishes sharply between the case in which the current theory already entails the new information and the case in which it does not. The standard AGM axioms [3, Ch.3.3] also make a sharp distinction, but along a different line: They distinguish between the case in which the evidence is consistent with the current theory (but not necessarily already part of it) and the case in which the evidence is inconsistent with the current theory. Specifically, the AGM axiom K*3 requires that $T \cup p \vdash T * p$, which is the characteristic axiom of Pareto-minimal theory change. The postulate K*4 posits that if $T \cup p$ is consistent, then $T * p \vdash T \cup p$. Thus the AGM axioms require the revised theory to be $Cn(T \cup \{p\})$ whenever p is consistent with T. In that case, the revision $Cn(T \cup \{p\})$ is a Pareto-minimal theory change, but it is just one of many possible Pareto-minimal revisions, namely the logically strongest one.

Another theory of belief change that endorses K*3 but not K*4 is the "updating" approach [6]. Intuitively, the connection between Pareto-minimality and the Update operator is this: Katsuno and Mendelzon postulate that "an update method should give each of the old possible worlds [in which the previous theory is true] equal consideration" [5, p.4]. Translating from possible worlds to sets of sentences, this means that Update treats adding new beliefs (removing possible worlds) as a "cost" in belief change, which can justify retracting previous beliefs (adding new possible worlds), even when the new information is consistent with the agent's current theory (for an example, see [5, p.7]).

Katsuno and Mendelzon argue that giving equal consideration to each of the old possible worlds is appropriate when an agent learns how the world has changed (update) rather than new facts about a static world (revision). This suggests that an agent's attitude towards the relative importance of additions and retractions may depend on the context and content of her beliefs. Pareto-minimality weights additions and retractions equally; in other contexts we may wish to give priority to minimizing retractions.[6] In the limiting case, we give absolute priority to minimizing retractions first, and only then consider avoiding additions. It can be shown that an agent's theory revision satisfies K*4 if and only if the agent makes the trade-off between additions and retractions in this way.

5 Pareto-Minimal Revision of Belief Bases

So far I have treated all of an agent's beliefs as equally important. A more refined representation of the agent's epistemic state may distinguish between a "basic" set of beliefs B, and the consequences of B that the agent might be said to hold because he believes B.[7] Hansson endorses the distinction between a basic set of beliefs and their consequences as a "small step toward capturing the justificatory structure" of an agent's beliefs [4]. I shall take a **base** for a theory T to be a set of formulas B, which may or may not be deductively closed, such that $B \vdash T$. (For more on belief bases, see [9, 10] and the references therein).

To define Pareto-minimal revision of belief bases, I begin again with two ways of making a change to a belief base. If B, B' are two bases, I say that B' **retracts** the formula p from B iff $p \in B$ and $p \notin B'$, and that B' **adds** the formula p to B iff $p \notin B$ and $p \in B'$. The definition of "adding more" and "retracting more" from a base is just like that for theories (cf. Definition 1). Thus B_1 **retracts more** formulas from B than B_2 iff $B - B_2 \subset B - B_1$, and B_1 **adds more** formulas to B than B_2 iff $B_2 - B \subset B_1 - B$.

As with Definition 3, we can apply the principle of Pareto-optimality to define a partial comparison of base revisions with respect to the extent of change that they induce.

Definition 4. *Let B, B_1, B_2 be three bases. Then B_1 is a* ***greater change*** *from B than B_2 is* $\iff$

[6] Levi presents a theory of how an agent may minimize the loss of "damped informational value" [7, Ch.2.1]. In my terms, this is advice for how to retract some beliefs to avoid adding too many.

[7] A paradigm example is a database, where we may distinguish between the records that are explicitly stored in the database and what follows from the explicitly stored information.

1. *B_1 retracts more formulas from B than B_2 does, and for all formulas p, if B_2 adds p to B, then B_1 adds p to B;* ***or***
2. *B_1 adds more formulas to B than B_2 does, and for all formulas p, if B_2 retracts p from B, then B_1 retracts p from B.*

As with Definition 2, an equivalent set-theoretic definition is that B_1 is a greater change from B than B_2 is iff $B_2 \triangle B \subset B_1 \triangle B$.

When we consider the extent of change of a belief base, it is natural to take into account only changes in basic beliefs, not changes in the logical consequences of the basic beliefs that "just follow" from them. My definition of retracting and adding to a belief base expresses this view of minimal belief change by considering only which sentences are added to or retracted from the set of basic beliefs. For example, there may be a sentence q such that $B * p \vdash q$ and $B \nvdash q$ but $q \notin B * p$. In that case the revision $B * p$ adds q to the logical consequences of the agent's beliefs, but does not add q to her basic beliefs. In effect, Definition 4 does not count such additions to the consequences of the agent's basic beliefs as an addition, unless they are also additions to the agent's basic beliefs themselves. Discounting changes in the logical consequences of basic beliefs in this way gives rise to a fundamental difference between the Pareto-minimal revision of basic beliefs and Pareto-minimal theory change: Pareto-minimal base revisions never add basic beliefs to the previous ones other than the new information. For suppose that a revision $B * p$ adds a belief q to a base B; then $B * p - \{q\}$ adds less to B and retracts no more. Hence $B * p$ is not a Pareto-minimal change of B. In contrast, a theory revision $T * p$ will typically add many beliefs to T, namely logical consequences of previous beliefs conjoined with the new information p. Another way to put the point is that for bases a conflict between additions and retractions does not arise: it is possible to minimize both additions and retractions at the same time. In the case in which the new information contradicts the current basic beliefs, this will lead an agent to hold inconsistent beliefs. Since many researchers accept as a general norm of epistemic rationality that an agent ought to avoid inconsistent beliefs, I shall restrict Pareto-minimal revisions to consistent bases.

Definition 5. *Let B, B_1 be two bases, and let p be a formula. Then B_1 is a* ***Pareto-minimal consistent change*** *from B that incorporates $p \iff$*

1. *$p \in B_1$, and*
2. *B_1 is consistent, and*
3. *there is no other consistent base B_2 such that $p \in B_2$ and B_1 is a greater change from B than B_2 is.*

What are the characteristic properties of Pareto-minimal base revisions? It turns out that a version of a proposal originally due to Levi amounts to necessary and sufficient conditions for a base revision to be Pareto-minimal and consistent. The proposal is to think of a Pareto-minimal revision of a belief base B on new information p as proceeding in two steps: First, remove just enough beliefs from B to obtain a belief base B' that is consistent with p; then add p to B'. Formally, we require that B' be a belief base that is consistent with p—thus $B' \nvdash \neg p$—and removes as few beliefs from B as possible. Hence I define a retraction-minimal *contraction* of a belief base as follows.

Definition 6. *Let B, B_1 be two bases, and let p be a formula. Then B_1 is a* ***retraction-minimal contraction*** *from B on $p \iff$*

1. *$B_1 \subseteq B$, and*
2. *$B_1 \not\vdash p$, and*
3. *there is no other base B_2 such that $B_2 \not\vdash p$ and B_1 retracts more from B than B_2 does.*

Retraction-minimal contractions of a base B on new information p have a simple characterization: They are exactly those subsets of B that cannot be expanded without entailing p. (The proof is left to the reader.)

Lemma 1. *Let B, B_1 be two bases such that $B_1 \subseteq B$, and let p be a formula. Then B_1 is a* ***retraction-minimal contraction*** *from B on $p \iff$ for all formulas q, if B_1 retracts q from B, it is the case that $B_1 \cup \{q\} \vdash p$.*

Thus retraction-minimal contractions are those that belief revision theorists refer to as "maxichoice contractions" [3, Ch.4.2]. The **Levi identity** says that minimal revisions of a belief set K given new information p are the result of adding p after contracting K on $\neg p$ (see [3, Ch.3.6]). The next proposition shows that the Levi identity for retraction-minimal (maxichoice) contractions characterizes Pareto-minimal revisions of belief bases that lead to consistent belief bases.

Theorem 2 (The Levi Identity for Belief Bases). *Let B be a base and let p be a formula. Suppose that a revision $B * p$ contains p. Then $B * p$ is a Pareto-minimal consistent change from B that incorporates $p \iff$ there is a retraction-minimal contraction B' from B on $\neg p$ such that $B * p = B' \cup \{p\}$.*

I omit the proof for space reasons. In view of Theorem 2, it is not difficult to see that Pareto-minimal consistent revisions of belief bases satisfy the AGM axioms K*1–K*5 (interpreted for base revisions with $\supseteq$ in place of $\vdash$; see also [1, Part II]).[8] The converse is not true, however: Pareto-minimality places more constraints on the revision of belief bases than K*1–K*5, since AGM revisions need not be the result of maxichoice contractions and hence may give up more beliefs than Pareto-minimal revisions.

[8] For K*2 I require that $p \in B * p$. For K*5 we must assume that the underlying consequence relation $\vdash$ is consistent in the sense that $\emptyset \nvdash L$; otherwise there is no consistent base. When the new information p is inconsistent, there is no consistent revision on p; in that case I require that $B * p$ is an inconsistent base in accordance with K*5.

Alchourrón and Makinson conjectured that "when applied to bases that are irredundant, choice contraction and revision functions serve as good formal representations of the corresponding intuitive processes" [1, p.21]. Theorem 2 establishes a formal version of this conjecture, in which Pareto-minimality takes the place of "intuition".[9]

6 Conclusion

The principle of minimal belief change is an important and influential idea in several areas of computer science such as artificial intelligence and database theory. This paper showed a method for rigorously deriving axioms for minimal belief change from fundamental decision-theoretic principles. This approach clarifies the foundations of belief revision postulates; and it allows us to distinguish universally valid postulates from those whose applicability depend to a larger extent on the details of how we represent beliefs and the relative weight we assign to retractions and additions in a given application domain.

Specifically, with regard to beliefs represented by deductively closed theories, Theorem 1 shows that the AGM axiom K*3 is universally valid for Pareto-minimal belief change, whereas the axiom K*4 is not.

With regard to beliefs represented by belief bases—which need not be deductively closed—Pareto-minimality validates K*4 and other staples of belief revision theory such as the Levi identity. In fact, Pareto-minimal base revision obeys constraints that go beyond the AGM postulates. Thus the results of my analysis of base revision largely agree with previous work; however, my method is different: I do not appeal to intuition, or even representation theorems, for justifying belief revision maxims, but instead *derive* them from fundamental decision-theoretic principles.

Altogether, the results in this paper show that Pareto-minimality provides a fruitful and principled decision-theoretic foundation for postulates guiding minimal belief revision.

7 Proof of Theorem 1

Theorem 1. *Let T be a theory and let p be a formula. A theory revision $T * p$ is a Pareto-minimal change from T that incorporates $p \iff$*

1. *$T * p \vdash p$, and*
2. *$T \cup \{p\} \vdash T * p$, and*
3. *if $T \vdash p$, then $T * p = T$.*

Proof. ($\Rightarrow$) Part 1: Immediate from Definition 3. Part 2: I show the contrapositive. Suppose that $T \cup \{p\} \nvdash T * p$. Then there is a formula q in $T * p$ such that

[9] Nebel also argues for constructing minimal base revisions from maxichoice contractions followed by adding the new information [9, Secs. 7, 8].

$T \cup \{p\} \nvdash q$. So (a) $T \nvdash q$ by Monotonicity. Now consider $T' = (T * p) \cap Cn(T \cup \{p\} \cup \{\neg q\})$. First I note that T' is closed under deductive consequence. For let $r \in Cn((T * p) \cap Cn(T \cup \{p\} \cup \{\neg q\}))$. Then by Monotonicity, $r \in Cn(T * p)$ and $r \in Cn(Cn(T \cup \{p\} \cup \{\neg q\}))$. We assumed that $T * p$ is closed under consequence, and Iteration implies that $Cn(Cn(T \cup \{p\} \cup \{\neg q\})) = Cn(T \cup \{p\} \cup \{\neg q\})$; thus $r \in T * p \ \cap \ Cn(T \cup \{p\} \cup \{\neg q\})$. This shows that $Cn(T') = T'$.

Next, note that (b) $T' \nvdash q$ because $Cn(T \cup \{p\} \cup \{\neg q\}) \nvdash q$ by Consistency (applied to $T \cup \{p\}$) and Iteration; thus from Monotonicity and the fact that $T' \subseteq Cn(T \cup \{p\} \cup \{\neg q\})$, it follows that $T' \nvdash q$. Moreover, we have from Monotonicity and the fact that $T' \subseteq T * p$ as well that (c) if T' adds a formula to T, so does $T * p$. From (a), (b) and (c) it follows that (d) $T * p$ adds more formulas to T than T'.

Now I show that (e) T' retracts from T exactly the formulas that $T * p$ retracts from T. Monotonicity implies immediately that if $T * p$ retracts a formula from T, so does T'. For the converse, suppose that T' retracts a formula r from T. Since $Cn(T \cup \{p\} \cup \{\neg q\}) \vdash T$, this implies that $r \notin (T * p)$. And that means that $T * p$ retracts r from T as well.

Finally, we have that (f) $T' \vdash p$, since $T * p \vdash p$ by Part 1 and clearly $Cn(T \cup \{p\} \cup \{\neg q\}) \vdash p$. Together, (a)–(f) establish that T' incorporates p and $T * p$ is a greater change from T than T' is. Hence $T * p$ is not a Pareto-minimal change.

Part 3: Immediate, since every theory other than T retracts or adds more formulas to T than T itself does.

($\Leftarrow$) Suppose that $T * p$ satisfies conditions 1, 2 and 3. Then the claim is immediate if $T \vdash p$ and $T * p = T$; suppose that $T \nvdash p$. I show that $T * p$ is not a greater change from T than any other change T' that incorporates p.

First, suppose that $T * p$ retracts a formula q from T but T' does not, such that $T' \vdash q$. Then $T' \vdash (p \wedge q)$ by Conjunction, whereas $T * p \nvdash (p \wedge q)$ by Conjunction as well. Since we supposed that $T \nvdash p$, it follows that $T \nvdash (p \wedge q)$ by Conjunction once more. So T' adds a formula to T—namely $p \wedge q$—that $T * p$ does not add to T, and hence $T * p$ is not a greater change from T than T' is.

Second, suppose that $T * p$ adds a formula q to T, but $T' \nvdash q$. Condition 2 asserts that $T \cup \{p\} \vdash T * p$ and hence $Cn(T \cup \{p\}) \vdash q$. By Deduction, we have that (a) $T \vdash p \rightarrow q$. Moreover, Implication implies that (b) $T * p \vdash p \rightarrow q$. Also, (c) $T' \nvdash p \rightarrow q$. For suppose that on the contrary, $T' \vdash p \rightarrow q$. Then since $T' \vdash p$, it follows from Modus Ponens that $T' \vdash q$, contrary to assumption. From (a), (b) and (c) we have that T' retracts a formula from T—namely $p \rightarrow q$—that $T * p$ does not retract from T. Thus $T * p$ is not a greater change from T than T' is.

These arguments establish that if $T * p$ satisfies conditions 2 and 3, then there is no theory T' incorporating p such that $T * p$ is a greater change from T than T' is. From Condition 1 it follows that $T * p$ is a Pareto-minimal change from T that incorporates p.$\square$

References

1. Alchourrón, C.E. and Makinson, D.: 1982, 'The logic of theory change: Contraction functions and their associated revision functions', *Theoria* 48:14–37.
2. Chou, T. and Winslett, M.: 1994, 'A Model-Based Belief Revision System', in *Journal of Automated Reasoning* 12:157–208.
3. Gärdenfors, P.: 1988, *Knowledge In Flux: modeling the dynamics of epistemic states.* MIT Press, Cambridge, Mass.
4. Hansson, S.O.: 1998, 'Editorial: Belief Revision Theory Today', *Journal of Logic, Language and Information*, Vol.7(**2**):123–126.
5. Katsuno, H. and Mendelzon, A.O. 1990: *On the difference between updating a knowledge base and revising it*, Technical Report on Knowledge Representation and Reasoning, KRR-TR-90-6, University of Toronto, Department of Computer Science.
6. Katsuno, H. and Mendelzon, A.O. 1991: 'On the difference between updating a knowledge base and revising it', in *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, Mass., pp.387–394, Morgan Kaufmann.
7. Levi, I.: 1996, *For the sake of the argument: Ramsey test conditionals, Inductive Inference, and Nonmonotonic Reasoning*, Cambridge University Press, Cambridge.
8. Nayak, A.: 1994, 'Iterated Belief Change Based on Epistemic Entrenchment', *Erkenntnis* 41**:** 353-390.
9. Nebel, B.: 1989, 'A Knowledge Level Analysis of Belief Revision', in: R. J. Brachman, H. J. Levesque, and R. Reiter (eds.), *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning* (KR'89), Toronto, Canada, pp. 301–311, Morgan Kaufmann.
10. Nebel, B.: 1994 'Base Revision Operations and Schemes: Representation, Semantics and Complexity', in: *Proceedings of the 11th European Conference on Artificial Intelligence* (ECAI'94), Amsterdam, Netherlands, pp. 341–345, Springer Verlag.