Reset reproduction of article published in *Proc. Second International Conf. on Principles of Knowledge Representation and Reasoning (KR'91)*, pp. 163–174. Reprinted July 1994.

Rational Belief Revision* (Preliminary Report)

Jon Doyle
Laboratory for Computer Science
Massachusetts Institute of Technology
545 Technology Square
Cambridge, Massachusetts 02139, USA

Abstract

Theories of rational belief revision recently proposed by Alchourrón, Gärdenfors, Makinson, and Nebel illuminate many important issues but impose unnecessarily strong standards for correct revisions and make strong assumptions about what information is available to guide revisions. We reconstruct these theories according to an economic standard of rationality in which preferences are used to select among alternative possible revisions. By permitting multiple partial specifications of preferences in ways closely related to preference-based nonmonotonic logics, the reconstructed theory employs information closer to that available in practice and offers more flexible ways of selecting revisions. We formally compare this new conception of rational belief revision with the original theories, adapt results about universal default theories to prove that there is unlikely to be any universal method of rational belief revision, and examine formally how different limitations on rationality affect belief revision.

1 INTRODUCTION

One of the best developed formal theories of belief revision is the so-called AGM theory of Alchourrón, Gärdenfors, and Makinson [1985], which has been brought together with related results by the same authors in a book by Gärdenfors [1988]. (For convenience we will refer to results in the book rather than the original articles.) The AGM approach models belief states with sets of propositions and develops, motivates, and studies a small set of axioms that characterize how a rational agent should change its belief states when new beliefs are added, subtracted, or changed. The central result of this theory is that these axioms are equivalent to the existence of a complete preordering of all propositions according to their degree of *epistemic entrenchment* such that belief revisions always retain more entrenched propositions in preference to less entrenched ones.

The AGM theory of belief revision is not directly applicable to formalizing belief revision as practiced in artificial intelligence because it requires that revisions be deductively closed, infinite sets of propositions.[1] To better model AI practice, Nebel [1989] adapted the AGM theory so that finite sets of representing propositions mediate revisions. He then shows that the resulting revisions satisfy most of the AGM rationality axioms. In fact, Nebel shows how to define certain orders over propositions and sets of propositions, representing notions of *epistemic relevance*, so that selecting revisions to be maximal in these orders yields finite revisions which satisfy *all* the rationality axioms. He also proves the very satisfying result that revision by picking maximal consistent subsets can be just as rational as more elaborate forms of revision in which the reasoner retains only those beliefs following from all maximal consistent subsets. This makes mechanization of belief revision systems more practical since picking one consistent subset is much simpler than intersecting the consequences of all consistent subsets.

Unfortunately, even Nebel's theory does not completely succeed at formalizing practical notions of belief revision, even though it makes improvements in the AGM theory. Both theories suffer from unnecessarily strong standards for correct revisions, and from very strong assumptions about what information is available to guide revisions. Specifically, the sense of "rationality" postulated by these theories requires that rational revisions be unique (they do not permit equally acceptable alternative revisions). They also require total orderings of all propositions, even though most domains of knowledge formalized in AI systems are too

*A preliminary version of the full paper was presented at the Third International Workshop on Nonmonotonic Reasoning, South Lake Tahoe, California, June 1990.

[1] Makinson [1987] discusses this and related problems with the "recovery" postulate of the AGM theory.

incomplete and ambiguous to supply such complete orderings.

We show that these two problems are related by reconstructing the theory of belief revision according to an economic standard of rationality. Where the AGM axioms refer only to logical properties of revised sets of beliefs and are motivated in terms of logical coherence and conservation of beliefs, we make the underlying motivation more explicit and precise by identifying various preferences guiding revisions and viewing rational revision as choosing the new belief state to be of maximal preferability according to these revision preferences. This change has two major effects. First, since there may be several alternatives of maximal preferability, our theory does not make the strong assumption of unique revisions. Second, the enlarged framework is closer to AI practice than its predecessors since it demands only partial information about revision preferences and permits these partial preferences to be combined and used in more flexible ways.

The plan of the paper is as follows. Sections 2 and 3 summarize in turn the formal theories and central results of Alchourrón, Gärdenfors, Makinson, and Nebel. Section 4 introduces the formal theory of rational revision in the economic sense and studies its impact on the AGM axioms. Determinism aside, rational revision turns out to be quite different from the AGM notion, which makes the generally plausible but occasionally dubious assumption that believing more is always better. Section 5 examines the informational requirements of rational revision and the AGM and Nebel theories. We argue that in practice the preferences available to guide belief revision consist of numerous partial and sometimes conflicting preferences, and show how the orderings assumed in the AGM and Nebel theories are too inflexible for many applications. We then present the formal theory of rational revision guided by multiple partial preferences, which is based on several principles for rationally aggregating partial preferences. This theory is formally similar to Doyle and Wellman's [1991] theory of rational default reasoning and to the economic theory of social choice [Arrow, 1963]. We prove that no method for belief revision based on partial preferences satisfies all the rationality conditions on preference aggregation. Finally, we examine how irrationalities in preference aggregation lead to violations of the AGM rationality axioms.

2 REVISING BELIEF STATES

The AGM formalization of belief revision may be summarized as follows, using an adaptation of the notations of [Alchourrón *et al.*, 1985] and [Nebel, 1989]. In the following we suppose that $\mathcal{L}$ is a propositional language over the standard sentential connectives ($\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$), denote individual propositions by x, y, and z, and denote sets of propositions by A, B, and C. We write $\vdash$ to mean classical propositional derivability, and write Cn to mean the corresponding closure operator

$$Cn(A) \stackrel{\text{def}}{=} \{x \in \mathcal{L} \mid A \vdash x\}.$$

The AGM approach covers states of belief modeled in two ways: as deductively closed (but not necessarily consistent) sets of propositions, that is, propositional *theories* $A \subseteq \mathcal{L}$ such that $A = Cn(A)$, and also as *belief bases* that represent the beliefs contained in their deductive closure. Formally, we say that B is a base for A whenever $A = Cn(B)$. Naturally, a given theory can be represented by many different belief bases. The case of greatest practical interest is when the belief base B is finite (and small), but not every theory has a finite basis. We also use the same terminology if B is infinite, and even if $A = B$.

Each of these models of belief states gives rise to a different theory of belief revision. Most of the theoretical results, however, concern only the closed belief states. We will treat both approaches.

2.1 EXPANSIONS, CONTRACTIONS, AND REVISIONS

The AGM theory considers three types of operations on belief states. For each belief state A and proposition x we have:

Expansion: Expanding A with x, written $A + x$, means adding x to A and requiring that the result be a (possibly inconsistent) belief state.

Contraction: Contracting A with respect to x, written $A \dot{-} x$, means removing x from A in such a way to result in a belief state.

Revision: Revising A with x, written $A \dot{+} x$, means adding x to A in such a way that the result is a consistent belief state.

Expansion is naturally defined in terms of the union of the set of beliefs and the new proposition. In the belief base model, we may take the expansion of A by x as this union itself. In this paper, however, we will focus on the case of closed belief states, and define the expansion to be the closure of this union

$$A + x \stackrel{\text{def}}{=} Cn(A \cup \{x\}).$$

Contraction and revision, on the other hand, have no single natural definitions, only the standard requirement that the change made be as small as possible so as to minimize unnecessary loss of knowledge.

Alchourrón, Gärdenfors, and Makinson [1985] formulate and motivate sets of rationality postulates that these operations should satisfy. The axioms for rational contractions are as follows. For each belief state A and propositions x and y:

($\dot{-}1$) $A \dot{-} x$ is a belief state, and a theory whenever A is; *(closure)*

($\dot{-}2$) $A \dot{-} x \subseteq A$; *(inclusion)*

($\dot{-}3$) If $x \notin Cn(A)$, then $A \dot{-} x = A$; *(vacuity)*

($\dot{-}4$) If $\nvdash x$, then $x \notin Cn(A \dot{-} x)$; *(success)*

($\dot{-}5$) If $\vdash x \leftrightarrow y$, then $A \dot{-} x = A \dot{-} y$; *(equivalence)*

($\dot{-}6$) $A \subseteq Cn((A \dot{-} x) + x)$ whenever A is a theory; *(recovery)*

($\dot{-}7$) $(A \dot{-} x) \cap (A \dot{-} y) \subseteq A \dot{-} (x \wedge y)$ whenever A is a theory;

($\dot{-}8$) If $x \notin A \dot{-} (x \wedge y)$, then $A \dot{-} (x \wedge y) \subseteq A \dot{-} x$ whenever A is a theory.

The *closure* axiom ($\dot{-}1$) says that contracting a theory yields a theory. The *inclusion*, *vacuity*, and *success* postulates ($\dot{-}2$)-($\dot{-}4$) state that removing a proposition does not introduce any new propositions, that no change occurs when one tries to remove a proposition that is not a consequence of the belief state, and that removed propositions are indeed removed, unless they are tautologies (in which case they are present in every theory, and so cannot be removed). The *equivalence* axiom ($\dot{-}5$) (which is usually called the "preservation" axiom) states that the results of contraction do not depend on the syntactic form of the proposition removed; removing any logically equivalent proposition has the same effect. The *recovery* postulate ($\dot{-}6$) states a conservation principle by requiring that contraction of a theory with respect to a proposition removes nothing that cannot be recovered by adding the proposition back in. Axioms ($\dot{-}7$) and ($\dot{-}8$) relate the contraction of a theory with respect to a conjunction to the contractions with respect to the individual conjuncts; they imply that retracting a conjunction preserves more information than retracting both of its conjuncts simultaneously.

Alchourrón, Gärdenfors, and Makinson also develop a parallel set of axioms for revisions, which we will not repeat here. The first main result of their theory is that these postulates for revisions are logically equivalent to the contraction postulates if the revision $A \dot{+} x$ is defined by means of the *Levi identity* (after [Levi, 1977])

$$A \dot{+} x \stackrel{\text{def}}{=} (A \dot{-} \neg x) + x, \tag{1}$$

so that revision by x is equivalent to contracting by $\neg x$ to remove any inconsistent beliefs and then expanding with x. One can also define contractions in terms of revisions by means of the *Harper identity* (after [Harper, 1976])

$$A \dot{-} x \stackrel{\text{def}}{=} (A \dot{+} \neg x) \cap A, \tag{2}$$

so that the contraction by x is equivalent to taking those beliefs that would be preserved if $\neg x$ were now believed.

2.2 EPISTEMIC ENTRENCHMENT

Though the AGM axioms characterize rational revisions, Gärdenfors [1988] views the behaviors these axioms describe as arising from a more fundamental notion, that of *epistemic entrenchment*. Epistemic entrenchment is characterized by a complete preorder (a reflexive and transitive relation) over propositions which indicates which propositions are more valuable than others. This ordering, which may vary from belief state to belief state, influences revisions by the requirement that revisions retain more entrenched beliefs in preference to less entrenched ones.

If x and y are propositions, we write $x \leq y$ to mean that y is at least as epistemically entrenched as x. We define the strict part of this order, $x < y$, which means that y is more entrenched than x, by the conjunction of $x \leq y$ and $y \not\leq x$. The following axioms characterize the qualitative structure of this order.

(≤ 1) If $x \leq y$ and $y \leq z$, then $x \leq z$; *(transitivity)*

(≤ 2) If $x \vdash y$, then $x \leq y$; *(dominance)*

(≤ 3) Either $x \leq x \wedge y$ or $y \leq x \wedge y$; *(conjunctiveness)*

(≤ 4) If A is a consistent theory, then $x \leq y$ for all y iff $x \notin A$; *(minimality)*

(≤ 5) If $x \leq y$ for all x, then $\vdash y$. *(maximality)*

Axiom (≤ 1) just says that $\leq$ is an ordering relation, while the other axioms all concern how the logic of propositions interacts with the ordering. Postulate (≤ 2) says that x entails y, then retracting x is a smaller change than retracting y, since the closure requirement on belief states means that y cannot be retracted without giving up x as well. Axiom (≤ 3) reflects the fact that a conjunction cannot be retracted without giving up at least one of its conjuncts. Taken together axioms (≤ 1)-(≤ 3) imply that $\leq$ is a complete ordering, that is, that either $x \leq y$ or $y \leq x$. Propositions not in a belief state are minimally entrenched in that state, according to (≤ 4), and according to (≤ 5), the only way a proposition can be maximally entrenched is if it is logically valid.

The influence of epistemic entrenchment on belief revisions is characterized by two conditions relating entrenchment orderings and contraction functions over theories. The first condition,

$$x \leq y \text{ iff either } x \notin A \dot{-} (x \wedge y) \text{ or } \vdash x \wedge y, \tag{3}$$

says that in contracting a theory A with respect to a conjunction, we must give up the conjunct of lesser epistemic entrenchment, or both conjuncts if they are equally entrenched. It says, in essence, that $x < y$ is the same as $y \in A \dot{-} (x \wedge y)$. The second condition,

$$y \in A \dot{-} x \text{ iff } y \in A \text{ and either } x < x \vee y \text{ or } \vdash x, \tag{4}$$

explicitly characterizes contraction functions in terms of epistemic entrenchment orderings.

The main result of the theory of epistemic entrenchment is that this notion is essentially equivalent to the previously axiomatized notion of rational revision of theories. Gärdenfors and Makinson [1988] prove that rational contraction functions may be constructed from orderings of epistemic entrenchment, and that entrenchment orderings may be constructed from rational contraction functions.

2.3 SOME IMPORTANT CONTRACTION FUNCTIONS

The central construct in many studies of contractions and revisions is the set of all maximal subsets of a belief state A consistent with a proposition x, which we write as $A \downarrow x$ and read as "A less x." Formally, we have

$$A \downarrow x \stackrel{\text{def}}{=} \{B \subseteq A \mid B \nvdash x \wedge (B \subset C \subseteq A) \rightarrow C \vdash x\}.$$

It is easy to see that if A is a theory, so are the elements of $A \downarrow x$.

Using this construct, one can define the contraction $A \dot{-} x$ as the set of beliefs in either one, some, or all states in $A \downarrow x$. These definitions correspond to the notions of maxichoice, partial meet, and full meet contractions. We will consider only the first two of these here.

2.3.1 Maxichoice contraction

Maxichoice contraction is contraction to one maximal consistent subset. If there are no alternatives, then x must be logically valid and impossible to retract, so we may take A itself as the contraction. Formally, we assume the existence of a choice function $\mathcal{C}$ which selects one of the elements of $A \downarrow x$ and define the maxichoice contraction operation $\stackrel{\text{m}}{-}$ by

$$A \stackrel{\text{m}}{-} x \stackrel{\text{def}}{=} \begin{cases} \mathcal{C}(A \downarrow x) & \text{if } \nvdash x \\ A & \text{otherwise.} \end{cases}$$

Applied to theories, this operation satisfies ($\dot{-}$1)-($\dot{-}$6), but not necessarily ($\dot{-}$7) and ($\dot{-}$8) [Gärdenfors, 1988, Lemma 4.1]. Maxichoice contraction does satisfies ($\dot{-}$7) and ($\dot{-}$8) in the case that the choice function $\mathcal{C}$ is *orderly*, that is, if there is some partial ordering $\sqsubseteq$ of all subsets of A such that $\mathcal{C}$ always chooses an element of $A \downarrow x$ that is maximal with respect to $\sqsubseteq$ [Gärdenfors, 1988, Lemma 4.3]. However, failure to satisfy ($\dot{-}$7) and ($\dot{-}$8) is the least of the problems with maxichoice contraction, for one may prove that using maxichoice contraction to effect the revision $A \dot{+} x$ via (1) makes $A \dot{+} x$ a *complete* theory as long as $\neg x \in A$ [Gärdenfors, 1988, Corollary 4.6]. That is, $A \dot{+} x$ in this case contains either y or $\neg y$ for each proposition y. Since A need not have contained either y or $\neg y$, these new beliefs are clearly gratuitous.

2.3.2 Partial meet contraction

The operation of *partial meet* contraction, $A \stackrel{\text{p}}{-} x$, assumes a selection function $\mathcal{S}$ which selects subsets of $A \downarrow x$, and defines contraction by

$$A \stackrel{\text{p}}{-} x \stackrel{\text{def}}{=} \begin{cases} \bigcap \mathcal{S}(A \downarrow x) & \text{if } \nvdash x \\ A & \text{otherwise.} \end{cases} \quad (5)$$

Partial meet contraction satisfies the basic rationality postulates ($\dot{-}$1)-($\dot{-}$6), and in fact is equivalent to them in the sense that any operation satisfying these axioms is a partial meet contraction operation [Gärdenfors, 1988, Theorem 4.13].

Some partial meet contractions over theories, namely those derived from orderings of theories, are fully rational, satisfying ($\dot{-}$1)-($\dot{-}$8). We say that the contraction operation is *relational* if there is a binary relation $\sqsubseteq$ over 2^A such that the selected subsets are exactly the $\sqsubseteq$-"maximal" subsets, that is, if

$$\mathcal{S}(A \downarrow x) = \{B \in A \downarrow x \mid \forall C \in (A \downarrow x)\ C \sqsubseteq B\}. \quad (6)$$

This gives us a way of constructing a selection function, and hence a partial meet contraction relation, from every relation $\sqsubseteq$ over $A \downarrow x$. The most important case is that when $\sqsubseteq$ is a transitive relation. Gärdenfors [1988, Theorem 4.16] proves that if $\sqsubseteq$ is transitive, the contraction function defined by (5) and (6) satisfies ($\dot{-}$1)-($\dot{-}$8). He also proves that for each transitive relation $\sqsubseteq$ there is a transitive total order $\sqsubseteq'$ which yields the same contraction function as $\sqsubseteq$ [Gärdenfors, 1988, Theorem 4.17]. That is, if a contraction function satisfies ($\dot{-}$1)-($\dot{-}$8), there is a complete preordering of belief states such that contraction is partial meet contraction with respect to this total ordering.

3 REVISING BELIEF BASES

The AGM theory is not directly applicable to artificial intelligence since it requires that revisions of even finite belief states be infinite belief states. Nebel [1989] modified the AGM theory to yield finite revisions of belief bases. He defines belief base contraction, which we write as $\ominus$, by

$$B \ominus x \stackrel{\text{def}}{=} \begin{cases} (B \vee \neg x) \wedge \bigvee_{C \in (B \downarrow x)} C & \text{if } \nvdash x \\ B & \text{otherwise,} \end{cases}$$

and belief base revision ($\oplus$) by

$$B \oplus x \stackrel{\text{def}}{=} (B \ominus \neg x) \wedge x.$$

We may view $\ominus$ as an "implementation" of $\dot{-}$ by lifting base revision to the belief state level. That is, we may use the contraction postulates to judge base contraction by identifying A with $Cn(B)$ and identifying $A \dot{-} x$ with $Cn(B \ominus x)$. With these identifications, Nebel [1989, Lemma 11] proves that belief base contraction satisfies ($\dot{-}$1)-($\dot{-}$6).

Nebel proceeds to show that belief base contraction is a form of partial meet contraction when lifted to belief states. Let $\mathcal{S}_B$ be a selection function corresponding to theory B and defined by

$$\begin{aligned}&\mathcal{S}_B(Cn(B) \downarrow x) \stackrel{\text{def}}{=} \\ &\quad \{C \in (Cn(B) \downarrow x) \mid \forall C' \in (Cn(B) \downarrow x) \\ &\quad\quad C' \cap B \not\supset C \cap B\}. \end{aligned} \tag{7}$$

Nebel [1989, Theorem 14] then proves that contraction of finite premise sets B using $\ominus$ is identical (with respect to $\vdash$) to a partial meet contraction $\stackrel{\text{p}}{-}_B$ defined by the selection function $\mathcal{S}_B$, that is,

$$Cn(B \ominus x) = Cn(B) \stackrel{\text{p}}{-}_B x.$$

This guarantees that the lifted version of $\ominus$ satisfies $(\dot{-}1)$-$(\dot{-}6)$. By defining the order $\sqsubseteq_B$ by

$$X \sqsubseteq_B Y \text{ iff } X \cap B \not\supset Y \cap B,$$

Nebel proves that partial meet contraction using $\mathcal{S}_B$ satisfies $(\dot{-}7)$ as well [Nebel, 1989, Theorem 15]. Since the order $\sqsubseteq_B$ is not transitive in general, this definition of contraction need not satisfy $(\dot{-}8)$.

These operations depend very strongly on the form of the belief base. In particular, iterated contraction of belief bases does not always make much sense, since $\ominus$ and $\oplus$ sometimes replace all previous base propositions with a single new proposition that was not an element of of the previous set. Subsequent contraction may then discard this singular residue of the original belief base.

While one may apply the notion of epistemic entrenchment in the foundational view of belief revision, this approach is not always practical. The dominance axiom (≤ 2), which requires that the logical consequences of a belief be at least as epistemically entrenched as the belief itself, means that to determine whether one proposition is more entrenched than another may require determining whether it entails the other, and since entailment is not decidable, this may not be possible. Nebel [1990, pp. 162-166] introduces the notion of *epistemic relevance* as an analogue of epistemic entrenchment more suited to the needs of computational belief revision. The basic idea is to view the syntactical form of the belief base as indicating or determining which propositions are *relevant* to the agent's purposes, and to view the propositions not in the belief base as irrelevant. The guideline for belief revision is then to minimize the loss of epistemically relevant propositions. Nebel proposes the selection function $\mathcal{S}_B$ defined in (7) as a formalization of this notion.

Nebel goes on to enlarge the conception of epistemic relevance from a simple binary distinction to a total preordering $\leq_\rho$ of the propositions in B, where $x \leq_\rho y$ means that y is at least as relevant as x. Unlike epistemic entrenchment orderings, epistemic relevance orderings may be arbitrary orderings of propositions, regardless of any logical dependencies existing among the propositions. Nebel shows how such orderings of epistemic relevance can be used to embed $\sqsubseteq_B$ in a transitive total ordering $\sqsubseteq_\rho$ over subsets of B, defined by

$$X \sqsubseteq_\rho Y \text{ iff } \forall x \in (X - Y)\ \exists y \in (Y - X) \quad x \leq_\rho y.$$

This order ranks X less relevant than Y just in case X's most relevant elements are less relevant than Y's most relevant elements, ignoring the elements X and Y have in common. Since the propositions in B are totally ordered by $\leq_\rho$, $\sqsubseteq_\rho$ totally orders all subsets of B. Furthermore, since over subsets of B it is clear that $X \sqsubset_B Y$ just means $X \subset Y$, we see that $X \sqsubset_\rho Y$ holds trivially if $X \sqsubset_B Y$, so the new order extends the old one. Nebel then defines a new selection function $\mathcal{S}_{B,\leq_\rho}$ in terms of $\sqsubseteq_\rho$ by

$$\begin{aligned}&\mathcal{S}_{B,\leq_\rho}(A \downarrow x) \stackrel{\text{def}}{=} \\ &\quad \{C \in (A \downarrow x) \mid \forall C' \in (A \downarrow x) \quad C' \cap B \sqsubseteq_\rho C \cap B\}\end{aligned}$$

and proves [Nebel, 1989, Theorem 16] that partial meet contraction defined by this selection function gives rise to a fully rational contraction function satisfying $(\dot{-}1)$-$(\dot{-}8)$.

If $\leq_\rho$ is a linear ordering (no ties allowed), then every finite set of beliefs has a most relevant element, and we may rewrite the definition of $\sqsubseteq_\rho$ as

$$X \sqsubseteq_\rho Y \text{ iff } \max(X - Y) \leq_\rho \max(Y - X).$$

In this case, $\sqsubseteq_\rho$ always singles out the greatest element of $(B \downarrow x)$, and the partial meet contraction function defined using $\mathcal{S}_{B,\leq_\rho}$ resembles a maxichoice contraction on the belief base (see [Nebel, 1989, Lemma 17]). Nebel concludes from this that maxichoice contraction on belief bases does not have the undesirable completion behavior exhibited by maxichoice contraction of belief states. Moreover, maxichoice base contraction is just as rational as the more complicated base contraction. Since it can be iterated easily and does not introduce complicated disjunctions, it is to be preferred in practice.

4 CHOOSING REVISIONS RATIONALLY

The AGM principles of rational belief revision are intended to capture logical constraints on revisions and contractions, as opposed to pragmatic influences. Accordingly, the axioms for contraction functions concern mainly logical relationships among beliefs. But it is not clear that there are any constraints free of pragmatic motivation. In particular, the whole motivation for conserving as many beliefs as possible seems entirely nonlogical. Conservatism has nothing to do with the consistency or completeness of beliefs or the soundness of inferences, which are the *only* characteristics of concern to logic. As Gärdenfors [1989] acknowledges, the motivations for conservatism are instead economic: beliefs are valuable (useful in acting, costly to infer or acquire), so getting rid of beliefs unnecessarily is irrational. This economic motivation for conservatism may not be pragmatic in the sense of being specific to some particular reasoning situation, but it is certainly not logical.

If we really desire a theory of rational belief revision, we must expand the notion of rationality from the purely logical sense, in which one is rational if one's beliefs are consistent and one's inferences are sound, to the economic sense, in which one is rational if one has a consistent set of preferences and makes choices that are optimal with respect to these preferences (cf. [Doyle, 1990]). This means the theory should use preferences about revisions in guiding revisions, and draw on the standard theory of economic rationality to formalize the notions of rational contraction and revision. In particular, the theory should take seriously the variability of the costs and benefits of beliefs. If the theory is to be general and cover all reasoning, it should allow preferences to vary with the reasoning situation and task, and should not presuppose any special measures of costs and benefits, but instead encompass all dimensions of value for guiding revisions.

According to the normative theory of economic rationality, rational agents choose maximally preferred alternatives. Preferences may stem from many different motivations, such as computational costs or moral principles, but all the formal theory requires is that a preference ordering $\precsim$ is a complete, reflexive, and transitive relation that represents the agent's judgments of relative overall preferability (or *utility*) of possible alternatives. That is, $\precsim$ satisfies, for all alternatives X, Y, and Z, the axioms

($\precsim$1) Either $X \precsim Y$ or $Y \precsim X$, and *(completeness)*

($\precsim$2) If $X \precsim Y$ and $Y \precsim Z$, then $X \precsim Z$. *(transitivity)*

It follows from axiom ($\precsim$1) that the order $\precsim$ is also reflexive, that is, $X \precsim X$. We use $\prec$ to denote *strict preference*, the asymmetric part of the preference order, and $\sim$ to indicate *indifference*, the reflexive part of the order. Thus $X \prec Y$ iff $X \precsim Y$ and $Y \not\precsim X$, and $X \sim Y$ iff $X \precsim Y$ and $Y \precsim X$. The combined order $\precsim$ is sometimes called *weak preference*. The agent's preference ordering may change with its state, but we will consider only instantaneous preferences.

Belief revision involves moving from one belief state to another, so the choice in question to which preferences must apply is the choice of one belief state over alternative belief states. In practice, preferences among complete belief states will be constructed from preferences among classes among belief states. The simplest such preferences are those corresponding to preferences among individual beliefs. For example, we may construct preferences among belief states from the ordering of propositions according to epistemic entrenchment. Intuitively, $x \leq y$ means that $X \sqsubseteq Y$ for any X and Y such that $x \in X$ and $y \in Y$. However, some important preference orders over belief states, such as the preferential interpretation of default rules proposed by Doyle and Wellman [1991], are more complex than comparisons of individual beliefs.

4.1 RATIONALITY AND SKEPTICISM

If we wish contraction based on epistemic entrenchment or epistemic relevance to be rational in the logical sense, then the resulting belief states must be maximal with respect to these orderings of belief states. That is not usually the case, however. First consider epistemic entrenchment. The maximally preferred states in $A \downarrow x$ are just the states in $\mathcal{S}(A \downarrow x)$. But partial meet contraction makes the new belief state be the intersection of these. In the terminology of inheritance and default reasoning, partial meet contraction is *skeptical* [Horty *et al.*, 1990]. The difficulty is that while skepticism is sometimes rational, the skepticism induced by partial meet contraction is rarely rational, since the intersection of preferred belief states is generally not itself a preferred belief state, that is, generally we have $\bigcap \mathcal{S}(A \downarrow x) \notin \mathcal{S}(A \downarrow x)$. For example, if we judge the preferability of belief states along numerous dimensions, each of the maximal belief states will be better along some dimensions and worse along others, and the intersection will be nonoptimal along as many dimensions as there are alternatives. The problem is clearest when examined in terms of propositional entrenchment. Consider the contraction $A \dot{-} (x \wedge y)$, and suppose that $x \leq y$, $\nvdash x$, $\nvdash y$, and $x, y \in A$. If $x < y$, then x is given up, while if $y \leq x$, both x and y are given up. There is no way to take the entrenchment-equivalence of x and y as a license to give up just one, chosen indifferently. Rather than increasing rationality, the skepticism resulting from choosing intersections of preferred belief states usually results in choosing suboptimal belief states. We conclude that if belief revision is to be rational, the belief states obtained in contraction and revision must be rational choices themselves rather than the intersection of rational choices.

4.2 RATIONALITY AND DETERMINISM

The first consequence of abandoning the skeptical approach to belief revision is that there is no longer any epistemological reason from logic or economics to suppose that contraction or revision are deterministic operations, as is presupposed in the AGM theory by viewing contraction and revision as functions taking belief states into belief states. Rationality does not prevent belief revision from being functional in deterministic agents. Even if the deterministic construction of the agent ensures that only one of the logically possible revisions is constitutionally possible, every revision performed may still be optimal with respect to the logically possible revisions. But determinism is in no sense a logical constraint on belief revision. This point is also made by Lindström and Rabinowicz [1989], who develop a nondeterministic extension of the AGM theory.

Accordingly, we expand the formalization of contraction and revision to cover the cases in which contraction and revision are correspondences, that is, set-valued functions taking propositions and belief states into sets of belief states. We write $\hat{-}$ to denote rational contraction, and $\hat{+}$ to denote rational revision. For example, if $\nvdash x$ we might define the rational contraction $A \mathbin{\hat{-}} x$ to be $\mathcal{S}(A \downarrow x)$ if the selection function $\mathcal{S}$ is based on a preference order. Of course, we can always construct skeptical deterministic contraction and revision functions from nondeterministic ones by intersection, that is, by defining $A \mathbin{\dot{-}} x$ to be $\bigcap(A \mathbin{\hat{-}} x)$.

4.3 RATIONAL CONTRACTION AND REVISION

Even if $\sqsubseteq$ represents preferences over belief states and $\bigcap \mathcal{S}(A \downarrow x) \in \mathcal{S}(A \downarrow x)$, so that the intersection of the preferred alternatives is itself a preferred alternative, partial meet contraction still need not be rational because there is no reason *a priori* why maximally preferred belief states should be sought only among the sets in $A \downarrow x$. Accordingly, we allow contraction to select among all belief states generated by subsets lacking x. The appropriate definitions vary depending on whether belief states are viewed as theories or as belief bases.

We define $A \Downarrow x$, read "A without x," by

$$A \Downarrow x \stackrel{\text{def}}{=} \{X \subseteq A \mid X \nvdash x\}.$$

Clearly, $A \downarrow x \subseteq A \Downarrow x$ for all A and x. We indicate the closures of these consistent subsets by $A \Downarrow^* x$, defined by

$$A \Downarrow^* x \stackrel{\text{def}}{=} \{Cn(X) \mid X \in A \Downarrow x\}.$$

We then define rational contraction of theories by

$$A \mathbin{\hat{-}} x \stackrel{\text{def}}{=} \begin{cases} \{A\} & \text{if } \vdash x \\ \{A\} & \text{if } A \mathbin{\hat{-}'} x = \emptyset \\ A \mathbin{\hat{-}'} x & \text{otherwise} \end{cases} \tag{8}$$

where

$$A \mathbin{\hat{-}'} x \stackrel{\text{def}}{=} \{X \in (A \Downarrow^* x) \mid \forall X' \in (A \Downarrow^* x) \quad X' \precsim X\}.$$

Similarly, we define rational revision of theories by

$$A \hat{+} x \stackrel{\text{def}}{=} \begin{cases} \{\mathcal{L}\} & \text{if } \vdash \neg x \\ \{\mathcal{L}\} & \text{if } A \hat{+}' x = \emptyset \\ A \hat{+}' x & \text{otherwise} \end{cases} \tag{9}$$

where

$$A \hat{+}' x \stackrel{\text{def}}{=} \{X + x \mid \quad X \in (A \Downarrow^* \neg x) \wedge \forall X' \in (A \Downarrow^* \neg x) \quad X' + x \precsim X + x\}.$$

Although we will study the implications of these definitions in the following, they should only be taken as initial attempts to define the notions of rational contraction and rational revision. The reader will note, for example, that the cases defining contraction and revision in the cases $\vdash x$ and $\vdash \neg x$, respectively, are each subsumed by the immediately following conditions in (8) and (9). The former cases correspond to there being no possible effective contraction or revision, while the latter cases correspond to there being possible effective contractions or revision, but no maximally preferable ones. The definitions make sense in spite of the subsumption because the same result is indicated for both of these cases. But it is not at all clear that the two deserve to be treated the same way.

There are even more interesting possibilities for defining rational contraction and revision of belief bases. For example, we may choose contractions and revisions either by ranking belief bases themselves (comparing them directly), or by ranking their closures (judging them by their "effects" as it were). Moreover, the contraction and revision operations may be permitted to perform some amount of deductive inference, that is, to augment the belief bases with parts of their closures, or may be restricted to sets of beliefs deemed to be "legal" states. Space limitations preclude formal elaboration of any of these possibilities.

The generality exhibited in the definitions of rational contraction and revision goes against the usual presupposition of epistemologists that knowing more is always better. Indeed, one can even prove that knowing more is better in some standard theories (see [Good, 1983]). But this presupposition (or theorem) is not always justified when holding or revising beliefs have non-epistemological effects, such as incurring computational costs that must be borne by the agent. For example, if it takes too long too compute enough of a revision that satisfies the AGM postulates, a practical belief revision method might instead simply seek to remove the easily identified inconsistencies between the belief state and the new information, rather than work to remove all inconsistencies between these. In other cases, knowing some specific beliefs may make the agent worse off than if it did not hold them. For example, *1984*'s Winston Smith is much better off not knowing some of the facts about his crazy world, since he will be severely tortured if he reveals this knowledge, and has no hope of keeping up a perfect pretense of ignorance. More mundanely, there are many facts that people find very painful to know (infidelity of a spouse, criminality of an employer) unless they also know enough extra facts to permit effective action (adequate legal grounds for divorce, sufficient evidence for conviction). The general pattern here is that for certain questions believing something may be worse than not believing it, even though not believing it is in turn worse than believing it for good reason.

4.4 ECONOMIC VS. AGM RATIONALITY

Do economically rational contractions satisfy the AGM postulates? The obvious answer is no: rational contraction is of a different type than AGM contraction, a nondeterministic relation rather than a functional relation. Nevertheless, we can compare rational contractions with AGM contractions in a more interesting way by seeing if deterministic contraction operations that are rational in the economic sense need also satisfy the AGM axioms. Formally, we say that $\dot{-}$ is a *ratichoice* contraction if it is both rational and functional, that is, if $A \dot{-} x \in A \hat{-} x$ for every A and x. We then have the following result.

Theorem 1 *Ratichoice contraction of theories satisfies the AGM axioms* ($\dot{-}$1), ($\dot{-}$2), *and* ($\dot{-}$4), *and can violate* ($\dot{-}$3), ($\dot{-}$5)-($\dot{-}$8), *and the Levi and Harper identities* (1) *and* (2).

Proof: Suppose that $\dot{-}$ is a functional rational contraction operation over closed belief states and that x, y, and z are logically independent and collectively consistent propositions.

To begin with, $A \dot{-} x$ is clearly a theory, so ($\dot{-}$1) holds. Similarly, $(A \dot{-} x) \subseteq A$, so ($\dot{-}$2) holds, as does ($\dot{-}$4).

We show that the remaining axioms need not hold by exhibiting preference orders that demonstrate this. First, though, we introduce some notation to make describing the preference orders more convenient. If $\hat{X}$ and $\hat{Y}$ are disjoint sets of subtheories of A, we write $\hat{X} \lhd \hat{Y}$ to mean that $X \prec Y$, $X \sim X'$ and $Y \sim Y'$ for every $X, X' \in \hat{X}$ and $Y, Y' \in \hat{Y}$. We also write $\mathrm{H}z$ (read "holds z") as an abbreviation for the subtheories of A in which the proposition z is held, that is, $\mathrm{H}z \stackrel{\text{def}}{=} \{Cn(X) \subseteq A \mid z \in Cn(X)\}$. Similarly, $\neg\mathrm{H}z$ means the complement of $\mathrm{H}z$, and $\mathrm{H}z \wedge \mathrm{H}w$ (resp. $\vee$) means the intersection (union) of these subsets.

Now suppose that $A = Cn(\{y\})$ and $A \nvdash x$, so $A \Downarrow^* x = \{Cn(\emptyset), Cn(\{y\}), Cn(\{x \vee y\}), \ldots\}$. If $\mathrm{H}y \lhd \neg\mathrm{H}y$, then $A \dot{-} x \neq A$, so ($\dot{-}$3) need not be satisfied. Similarly, if $A = Cn(\{x, y\})$ and $\mathrm{H}y \lhd \neg\mathrm{H}y$, it may be that $A \dot{-} x = Cn(\emptyset)$, in which case $(X \dot{-} x) + x = Cn(\{x\})$ and $A \not\subseteq Cn(\{x\})$. Thus ($\dot{-}$6) need not hold.

Suppose that $\vdash x \leftrightarrow y$. It is clear that the nondeterministic analogue of the equivalence axiom ($\dot{-}$5) holds for rational contraction, in that $A \hat{-} x = A \hat{-} y$. But since a ratichoice contraction may select a different alternative from $A \hat{-} x$ when contracting by x than when contracting by y, ($\dot{-}$5) need not hold.

Next, neither ($\dot{-}$7) nor ($\dot{-}$8) always hold. Suppose that $A = Cn(\{x, y, z\})$. If

$$\begin{aligned}\mathrm{H}x \wedge \neg\mathrm{H}y \wedge \neg\mathrm{H}z &\lhd \neg\mathrm{H}x \wedge \mathrm{H}y \wedge \neg\mathrm{H}z \\ &\lhd \neg\mathrm{H}x \wedge \neg\mathrm{H}y \wedge \mathrm{H}z \lhd \neg\mathrm{H}x \wedge \neg\mathrm{H}y \wedge \neg\mathrm{H}z \\ &\lhd \mathrm{H}x \wedge \neg\mathrm{H}y \wedge \mathrm{H}z \lhd \neg\mathrm{H}x \wedge \mathrm{H}y \wedge \mathrm{H}z,\end{aligned}$$

then $Cn(\{y, z\})$ is maximal in $A \Downarrow^* x$, $Cn(\{x, z\})$ is maximal in $A \Downarrow^* y$, and $Cn(\emptyset)$ is maximal in $A \Downarrow^* (x \wedge y)$, in which case ($\dot{-}$7) may be violated because

$$Cn(\{y, z\}) \cap Cn(\{x, z\}) = Cn(\{z\}) \neq Cn(\emptyset).$$

On the other hand, if

$$\begin{aligned}\neg\mathrm{H}x \wedge \neg\mathrm{H}y \wedge \neg\mathrm{H}z &\lhd \mathrm{H}x \wedge \neg\mathrm{H}y \wedge \mathrm{H}z \\ &\lhd \neg\mathrm{H}x \wedge \mathrm{H}y \wedge \mathrm{H}z \lhd \neg\mathrm{H}x \wedge \neg\mathrm{H}y \wedge \mathrm{H}z \\ &\lhd \mathrm{H}x \wedge \neg\mathrm{H}y \wedge \neg\mathrm{H}z \lhd \neg\mathrm{H}x \wedge \mathrm{H}y \wedge \neg\mathrm{H}z,\end{aligned}$$

then $Cn(\{y\})$ is maximal in $A \Downarrow^* x$ and $Cn(\{z\})$ is maximal in $A \Downarrow^* (x \wedge y)$, in which case ($\dot{-}$8) may be violated because $Cn(\{z\}) \not\subseteq Cn(\{y\})$.

Finally, neither the Levi nor Harper identities always hold. Suppose $A = Cn(\{\neg x, y\})$. If

$$\neg\mathrm{H}x \wedge \mathrm{H}y \lhd \neg\mathrm{H}x \wedge \neg\mathrm{H}y \lhd \mathrm{H}x \wedge \neg\mathrm{H}y \lhd \mathrm{H}x \wedge \mathrm{H}y,$$

it may be that $A \dot{-} \neg x = Cn(\emptyset)$ and $A \dot{+} x = \{Cn(\{x, y\})\}$, in which case

$$(A \dot{-} \neg x) + x = \{Cn(\{x\})\} \neq A \dot{+} x,$$

which contradicts the Levi identity (1). Similarly, we have

$$(A \dot{+} x) \cap A = Cn(\{y\}) \neq= A \dot{-} \neg x,$$

which contradicts (with x and $\neg x$ interchanged) the Harper identity (2). □

These complications may be avoided if we require that contractions be chosen from among $A \downarrow x$ rather than $A \Downarrow^* x$, as shown by the following result.

Theorem 2 *If* $(A \hat{-} x) \subseteq (A \downarrow x)$ *whenever* $\nvdash x$, *then ratichoice contraction satisfies the AGM axioms* ($\dot{-}$1)-($\dot{-}$8).

Proof: Suppose $(A \hat{-} x) \subseteq (A \downarrow x)$ whenever $\nvdash x$. Then there is some maxichoice function $\mathcal{C}$ such that $\mathcal{C}(A \downarrow x) \in (A \hat{-} x)$ for all x such that $\nvdash x$. Since each particular contraction given by $\hat{-}$ can be viewed as the result of a maxichoice contraction, ratichoice contraction in this case satisfies all the axioms that maxichoice contraction satisfies, namely (1), (2), and ($\dot{-}$1)-($\dot{-}$6). In fact, ratichoice contraction satisfies ($\dot{-}$7) and ($\dot{-}$8) as well because any ratichoice contraction operation is *orderly*, that is, any element of $A \hat{-} x$ is an element of $A \downarrow x$ that is maximal with respect to a partial ordering of subsets of A. As mentioned earlier, any orderly maxichoice contraction operation satisfies all the contraction axioms, and to apply this result, we need only consider the partial order $\sqsubseteq$ defined so that $X \sqsubseteq Y$ iff either $X = Y$ or $X \prec Y$. □

Of course, the conditions of Theorem 2 leads to ratichoice contractions that share the undesirable properties of maxichoice contraction of theories noted earlier: if $\neg x \in A$, then every proposition is either believed or denied when one first contracts with respect to $\neg x$ and then extends by x. It appears, however that the ratichoice revision by x in the same circumstances need not be complete, since it need not satisfy the Levi identity, but I have no proof of this.

Another way of avoiding the complications of ratichoice contraction is to require that believing more is better than believing less. Formally, we say that the preference relation $\precsim$ is *(positively) informationally monotone* if $X \precsim Y$ whenever X and Y are both consistent and $Y \vdash X$. We then have the following result.

Theorem 3 *If $\precsim$ is informationally monotone, then ratichoice contraction satisfies the Levi and Harper identities and the AGM axioms* ($\dot{-}1$)-($\dot{-}8$).

Proof: Obviously, if $\precsim$ is informationally monotone, then $(A \hat{-} x) \subseteq (A \downarrow x)$ if $\nvdash x$, so Theorem 2 applies. □

Many questions remain to be answered. For example, are the AGM axioms equivalent to specific conditions on preferences over belief states? Do rational state preferences induce rational propositional preferences?

5 PARTIAL EPISTEMIC PREFERENCES

Nebel suggests two reasons why epistemic relevance is a more practical basis for belief revision than epistemic entrenchment. In the first place, epistemic relevance orderings need not respect logical dependencies among propositions. That is, we are free to order $x \leq_\rho y$ without regard to whether $x \vdash y$ or $y \vdash x$, in contrast with the dominance condition ($\leq$2) on epistemic entrenchment orderings, which can be quite costly (or impossible) to ensure. In the second place, linear epistemic relevance orderings of the propositions in a belief base make belief base contraction easy to implement by simply dropping the lowest ranked propositions in any conflicting set (as in RUP [McAllester, 1982]). Moreover, this form of belief base contraction is fully rational, and corresponds to maxichoice contraction on belief bases [Nebel, 1989, Lemma 17ff]. Representing epistemic entrenchment, however, can be more costly. Gärdenfors and Makinson [1988] show that epistemic entrenchment orders can be represented by information linear in the size of a certain algebraic construct from belief states (the dual atoms of the lattice of equivalence classes of beliefs). Unfortunately, this may still be exponential in the number of atomic propositions in the belief base. But pure representational size is only half the problem, and they leave the problem of logical dependencies unaddressed. Orderings of propositions and belief bases will only be useful in practice if they can be both represented and computed quickly.

While Nebel's epistemic relevance orderings make rational belief revision more practical, this approach is not without significant costs. The main problem is the inflexibility of this means for effecting contractions. Specifically, linear epistemic relevance orderings rank each possible contraction by the most valuable proposition retained, irrespective of what other propositions are retained. This might seem reasonable since $\sqsubseteq_\rho$ is applied to the elements of $B \downarrow x$, in which case choosing the subset with the maximal element is the same as choosing the subset which abandons the least valuable propositions. But suppose $X = \{a_1, \ldots, a_{1000}\}$ and $Y = \{a_{1001}\}$, where $\leq_\rho$ ranks these propositions by the natural indicial order, so that $X \sqsubseteq_\rho Y$. If all of these propositions are of roughly the same value (but each differing slightly from the rest), then the revision

$$\{a_1, \ldots, a_{1001}\} \ominus (a_{1001} \wedge (a_1 \vee \ldots \vee a_{1000})) = Y$$

chooses among the two alternatives X and Y and discards a thousand good propositions in favor of a single proposition that is little better than those discarded. This seems unreasonable compared with, for example, using a weighted comparison or voting scheme in which equally valuable propositions get equal say in the selection. Unfortunately, there is no way to express schemes like majority voting with linear orders over propositions. More generally, achieving any dependence of ordering on the global composition of the alternatives means revising the linear propositional order to fit each set of alternatives.

It would be valuable to have some more flexible way of specifying preferences for guiding contraction and revision. If we look to the usual explanations of why one revision is selected over another , we see that many different properties of propositions influence whether one proposition is preferred to another. For example, one belief might be preferred to another because it is more specific, or was adopted more recently, or has longer standing (was adopted less recently), or has higher probability of being true, or comes from a source of higher authority. These criteria, however, are often partial, that is, each may be viewed as a preorder $\precsim$ such that both $X \not\precsim Y$ and $Y \not\precsim X$ for some X and Y. For example, there are many different dimensions of specificity, and two beliefs may be such that neither is more specific than the other. Similarly, probabilities need not be known for all propositions, and authorities need not address all questions. Moreover, none of these are comprehensive criteria that take all possible considerations into account. If we want contraction and revision to be truly flexible, we need some way of combining different partial, noncomprehensive orderings of propositions into complete global orderings of belief states.

But combining partial orderings into a global ordering can be difficult because the partial criteria may conflict in some cases. To borrow an example from nonmonotonic logic, we might reasonably prefer to believe that Quakers are pacifist, and that Republicans are not pacifists. These preferences can conflict on cases like that of Nixon, and a preference for more specific rules does not help since "Quaker" and "Republican" are incomparable categories. Indeed, as argued elsewhere [Doyle and Wellman, 1991], other preference criteria can conflict as well, including very specific criteria corresponding to individual default rules. Constructing a global ordering thus means resolving the conflicts among the preference criteria being combined.

In addition to flexibility, we seek a revision method which is potentially mechanizable. This means that whatever method is employed for resolving conflicts must also be mechanizable because placing responsibility for resolving potential conflicts on the theorist is infeasible. For large sets of criteria it is difficult to anticipate all of the potential conflicts and all of the varying circumstances that may influence how the conflicts should be resolved. It also seems difficult to anticipate discovery of new criteria. Thus we seek conflict resolution mechanisms based on general, modular rules of combination that apply even as the criteria used evolve.

5.1 CONSTRUCTING GLOBAL PREFERENCE ORDERS

To analyze the problem of modular construction of orderings, we follow the formal approach elaborated by Doyle and Wellman [1991] for analyzing the related problem for preference-based nonmonotonic logics. We say that an *aggregation policy* is a function that specifies the global order corresponding to any given set of partial preference orders. Let the set I index the set of partial preference orders that are to be combined, so that if $i \in I$, $\precsim_i$ denotes the preference order corresponding to the ith criterion for belief revision to be included in the overall revision conception. The problem is then to aggregate the set of orders $\{\precsim_i | \; i \in I\}$ into a global preference order $\precsim$.

The principled design of an aggregation policy for partial preference criteria begins with a consideration of properties we think a reasonable policy should exhibit. The properties we propose are analogs of Arrow's [1963] desiderata for social choice. (See [Doyle and Wellman, 1991] for further explanation and justification of these desiderata as principles for reasoning.)

1. *Collective rationality.* The global order $\precsim$ is a function of the individual orders $\precsim_i$, which are unrestricted, possibly partial, preorders.
2. *Pareto principle (unanimity).* If $X \prec_i Y$ for some $i \in I$ and for no $j \in I$ does $Y \prec_j X$, then $X \prec Y$. In other words, the global order agrees with unanimous strict preferences.
3. *Independence of irrelevant alternatives (IIA).* The relation of X and Y according to the global order depends only on how the individual orders rank those two candidates. That is, considering new alternatives does not alter rankings among the originals.
4. *Nondictatorship (noncomprehensive criteria).* There is no $i \in I$ such that for every X and Y, $X \precsim Y$ whenever $X \precsim_i Y$, regardless of the $\precsim_j$ for $j \neq i$. That is, there is no "dictator" whose preferences automatically determine the group's, no matter how the other individual orderings are varied.
5. *Conflict resolution.* If $X \precsim_i Y$ for some i, then $X \precsim Y$ or $Y \precsim X$. That is, if two candidates are comparable in an individual order, then they are comparable in the global order.

Leaving aside the conflict resolution condition for now, the following theorem states that the desirable and apparently reasonable properties enumerated above are not simultaneously satisfiable by any aggregation policy for preferences expressed by total orders.

Theorem 4 (Arrow) *If the domain includes more than two alternatives, no aggregation policy mapping sets of total preorders to total global preorders satisfies the collective rationality, Pareto, IIA, and nondictatorship conditions.*

We omit the proof because, with the restriction to total orders, this is exactly Arrow's theorem [Arrow, 1963].

There is no problem finding good aggregation policies for choices among only two alternatives: majority rule works fine, for example. But for the case of belief revision, there are typically several possible alternatives to choose from. This means that the following theorem applies to the typical case of belief revision. We omit the proof, which may be found, together with further discussion, in [Doyle and Wellman, 1991, Theorem 3].

Theorem 5 (Doyle-Wellman) *If the domain includes more than two alternatives, no aggregation policy for partial preference preorders satisfies the collective rationality, Pareto, IIA, nondictatorship, and conflict resolution conditions.*

Thus if rational belief revision requires a preorder that completely orders the alternative contractions or revisions, we may expect that the only general way of obtaining the order is manual construction, that is, to supply a dictatorial policy. One way to do this is to impose a linear ordering over all the criteria, so that the first criterion always gets its way regardless of what the rest of the criteria say, unless it expresses no preference, in which case the second criterion gets its way, and so on. Choice rules of this form are called *lexicographic* because they resemble the method for ordering words alphabetically: compare the first letters; if tied compare the second, etc. But like Nebel's orderings, these linear orderings are not very flexible, and can be expected to require ongoing manual revision to achieve satisfactory performance.

There are a number of possible ways around Theorem 5, some of which are discussed in [Doyle and Wellman, 1991] in the context of nonmonotonic logic. We mention only one here: the approach of restricting the domain over which preference aggregation occurs. The collective rationality condition stipulates that the aggregation policy must work no matter what preference orders are presented for combination. But if one can show that all preference orderings of interest take one of several particular forms, these limitations may permit construction of an aggregation policy satisfying all of the conditions over this restricted domain of preferences. Does Theorem 5 still hold if each preference order satisfies properties like informational monotonicity or $Z \precsim X \cap Y$ whenever $Z \precsim X$ and $Z \precsim Y$? Are such restrictions on the allowable preferences reasonable outside the special domain of epistemology?

5.2 REVISION RATIONALITY AND AGGREGATION RATIONALITY

Some irrationalities in preference aggregation do not affect the formal rationality of rational contraction and revision, which requires only that choices be maximally preferable according to a total preference ordering. For example, aggregation policies which yield total preference orderings and so provide the necessary basis for applying (8) and (9) may nevertheless violate the Pareto, IIA, or nondictatorship principles.

Violations of the Pareto or nondictatorship principles just mean that the resulting belief revisions may not be rational with respect to the ignored preference criteria, even though they are maximally preferred with respect to the aggregate order. An aggregation policy that violates IIA, on the other hand, can exhibit erratic behavior. Specifically, it can lead to different results depending on the presence or absence of irrelevant information. Suppose, for example, that $A = Cn(\{x, y, z\})$, and that P_1 and P_2 are two preference criteria to be combined with P_1 satisfying

$$\neg \mathrm{H}x \wedge \neg \mathrm{H}y \wedge \neg \mathrm{H}z \lhd \mathrm{H}x \wedge \neg \mathrm{H}y \lhd \neg \mathrm{H}x \wedge \mathrm{H}y$$

and P_2 satisfying

$$\neg \mathrm{H}x \wedge \neg \mathrm{H}y \wedge \neg \mathrm{H}z \lhd \neg \mathrm{H}x \wedge \mathrm{H}y \lhd \mathrm{H}x \wedge \neg \mathrm{H}y$$

If the aggregation policy violates IIA, we may have a ratichoice contraction $\dot{-}$ such that $A \dot{-} (x \wedge y) = \{Cn(\{y\})\}$ but $A \dot{-} (x \wedge y) \vee z = \{Cn(\{x\})\}$, with the overall choice between $Cn(\{x\})$ and $Cn(\{y\})$ depending on what other alternatives are available.

In contrast to these formally rational but substantively irrational revision orderings, violations of collective rationality and the incompleteness addressed by conflict resolution can lead to global preference "orderings" that are intransitive or incomplete. This raises difficulties when such comparisons used to select among alternative contractions and revisions. We examine these difficulties in turn.

An aggregation policy may violate collective rationality in several ways. First, it may not aggregate every possible partial preference criterion. As with violations of the Pareto or nondictatorship conditions, this need not affect the formal rationality of belief revision. Second, the result of aggregation may not be transitive. The effects of this depend on just what we mean by intransitivity. Intransitivity of $\precsim$ means that for some X, Y, Z we have $X \precsim Y$ and $Y \precsim Z$ but $X \not\precsim Z$. That is, if $A \Downarrow^* x = \{X, Z\}$, then both alternatives are maximally preferred, so either is a "rational" choice. On the other hand, intransitivity of $\prec$ means that for some X, Y, Z we have $X \prec Y$, $Y \prec Z$, and $Z \prec X$. In this case, if $A \Downarrow^* x = \{X, Y, Z\}$, then no alternative is maximally preferred, so $A \hat{-} x = A$.

The definitions of rational contraction and revision presumed a complete preference ordering of belief states. But there is no objection in principle to weakening the definitions by allowing $\precsim$ to be partial. (In this case, of course, we can no longer view $X \prec Y$ and $Y \not\precsim X$ as equivalent.) The effect of incompleteness is just to allow more alternatives to be maximally preferred than might be the case were the order to be total. Incompleteness when no relevant preferences are available does not seem particularly objectionable. But incompleteness when preferences do exist is a different matter, and such incompletenesses can arise either through intransitivity (as noted above), through aggregation policies that do not aggregate some preference orders, or through violations of the conflict resolution condition. In these cases, the incompleteness arises through an explicit violation of a reasonable rationality condition.

6 CONCLUSION

We examined the AGM theory of rational belief revision and its extension by Nebel. The heart of the AGM theory is the ordering of beliefs according to epistemic entrenchment and its equivalence with his notion of rational belief revision. Nebel adapted this theory to cover revision of finite bases of belief, and his results show that finite revision can be rational in AGM sense, and that in the finite case maxichoice revision is as rational as skeptical revision.

We argued that these theories are inadequate because they demand more information about which revisions are preferable than is typically available in practice, and because their notion of "rationality" has little to do with the economic notion of rational choice among alternatives. We presented a modification of these theories which makes weaker, more realistic assumptions, namely that belief revisions must be guided by partial preferences which may conflict with each other. This theory is closely related to our theory of rational default inference, and supports similar results. We showed how the notion of rationality proposed by Alchourrón, Gärdenfors, and Makinson is violated by revisions which are rational in the economic sense, and proved that fully rational belief revision is sometimes impossible when rationality is judged by partial preferences.

It is an open problem to find rationality postulates to describe revision that is rational in the economic sense. There may not be any, though a complete axiomatization seems more likely if one presupposes a global expected utility preordering of beliefs or of belief states. There certainly will not be axioms that refer only to beliefs, as in the AGM axioms. In addition, Theorem 5 suggests that there may not be a single ideal theory of rational revision.

Acknowledgments

This work draws substantially on previous work conducted in collaboration with Michael Wellman, and I thank him for many valuable suggestions and for reading drafts of the full version of this paper. The paper has also been improved significantly through discussions with of Peter Gärdenfors, David Makinson, and Bernhard Nebel, who each supplied many useful comments and pointed out errors in the preliminary version presented at the Third International Workshop on Nonmonotonic Reasoning. I also thank Sten Lindström, Wlodzimierz Rabinowicz, Joseph Schatz, and Richmond Thomason for valuable discussions. This work was supported by the National Library of Medicine through National Institutes of Health Grant No. R01 LM04493 and by the International Business Machines Corporation through a grant to MIT's Laboratory for Computer Science.

References

[Alchourrón *et al.*, 1985] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction functions and their associated revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[Arrow, 1963] K. J. Arrow. *Social Choice and Individual Values*. Yale University Press, 2nd edition, 1963.

[Doyle and Wellman, 1991] J. Doyle and M. P. Wellman. Impediments to universal preference-based default theories. *Artificial Intelligence*, 1991. To appear.

[Doyle, 1990] J. Doyle. Rationality and its roles in reasoning (extended abstract). In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pp. 1093–1100, Menlo Park, CA, 1990. AAAI Press.

[Gärdenfors and Makinson, 1988] P. Gärdenfors and D. Makinson. Revisions of knowledge systems using epistemic entrenchment. In M. Y. Vardi, ed., *Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 83–95, Los Altos, CA, March 1988. Morgan Kaufmann.

[Gärdenfors, 1988] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA, 1988.

[Gärdenfors, 1989] P. Gärdenfors. The dynamics of belief systems: Foundations vs. coherence theories. *Revue Internationale de Philosophie*, 1989.

[Good, 1983] I. J. Good. On the principle of total evidence. In *Good Thinking: The Foundations of Probability and Its Applications*. University of Minnesota Press, 1983. Originally appeared in *British Journal of Philosophy of Science*, 17:319-321, 1967.

[Harper, 1976] W. L. Harper. Rational conceptual change. In *PSA 1976*, volume 2, pages 462–494, East Lansing, MI, 1976. Philosophy of Science Association.

[Horty *et al.*, 1990] J. F. Horty, R. H. Thomason, and D. S. Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial Intelligence*, 42(2-3):311–348, March 1990.

[Levi, 1977] I. Levi. Subjunctives, dispositions and chances. *Synthese*, 34:423–455, 1977.

[Lindström and Rabinowicz, 1989] S. Lindström and W. Rabinowicz. Belief revision, epistemic conditionals and the Ramsey test. Department of Philosophy, Uppsala University, Sweden, 1989.

[Makinson, 1987] D. Makinson. On the status of the postulate of recovery in the logic of theory change. *Journal of Philosophical Logic*, 16:383–394, 1987.

[McAllester, 1982] D. A. McAllester. Reasoning Utility Package user's manual. AIM 667, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Cambridge, MA, 1982.

[Nebel, 1989] B. Nebel. A knowledge level analysis of belief revision. In R. J. Brachman, H. J. Levesque, and R. Reiter, editors, *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, pages 301–311, San Mateo, CA, 1989. Morgan Kaufmann.

[Nebel, 1990] B. Nebel. *Representation and Reasoning in Hybrid Representation Systems*. Springer-Verlag, Berlin, 1990.